

---

# Two- and three-point clustering statistics

Holger Alois Schlagenhauser

---



München 2012



---

# Two- and three-point clustering statistics

Holger Alois Schlagenhauser

---

Dissertation  
an der Fakultät für Physik  
der Ludwig-Maximilians-Universität  
München

vorgelegt von  
Holger Alois Schlagenhauser  
aus Nabburg

München, den 27.09.2012

Erstgutachter: Prof. Dr. Ralf Bender  
Zweitgutachter: Prof. Dr. Jochen Weller  
Tag der mündlichen Prüfung: 05.12.2012

# Contents

<b>List of figures</b>	<b>vii</b>
<b>List of tables</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Theoretical framework for the two- and three-point statistics</b>	<b>7</b>
2.1 Cosmological distances . . . . .	8
2.2 Baryonic Acoustic Oscillations (BAOs) . . . . .	10
2.2.1 Gravitation driven oscillations . . . . .	11
2.2.2 Baryon loading . . . . .	11
2.2.3 Influence on the clustering statistics . . . . .	13
2.3 L-BASICC simulations . . . . .	14
2.4 Power spectrum and Bispectrum: Definition . . . . .	15
2.5 Power spectrum and Bispectrum: Measurement . . . . .	15
2.5.1 Power spectrum and Bispectrum: Gaussian density field . . . . .	18
2.6 Power Spectrum and Bispectrum: Model . . . . .	20
2.6.1 Real space . . . . .	20
2.6.2 Redshift space . . . . .	35
2.6.3 Reduced bispectrum . . . . .	41
2.6.4 Extracting the bias parameters . . . . .	42
2.7 Correlation functions . . . . .	42
2.7.1 Definition of the correlation functions . . . . .	43
2.7.2 Theory of the correlation functions . . . . .	44
2.7.3 Redshift errors for the three-point statistic? . . . . .	51
<b>3 Investigation of <math>\xi(r_p, \pi)</math> and <math>B(k_1, k_2, k_3)</math></b>	<b>53</b>
3.1 Two-point statistics: Determination of $w_{DE}$ and $b$ . . . . .	53
3.1.1 Constraining $w_{DE}$ and $b$ . . . . .	54
3.1.2 Anisotropic two-point correlation function . . . . .	55
3.1.3 Projected two-point correlation function . . . . .	66
3.1.4 Other work in the literature . . . . .	70
3.2 Bispectrum . . . . .	72

3.2.1	Equilateral configurations: dark matter . . . . .	72
3.2.2	Equilateral configurations: dark matter halos . . . . .	76
3.2.3	All configurations . . . . .	82
3.2.4	Application to galaxy mocks: Lyman Alpha Emitters . . . . .	102
3.2.5	Other work in the literature . . . . .	109
<b>4</b>	<b>Summary and Conclusions</b>	<b>113</b>
4.1	Summary . . . . .	113
4.1.1	Anisotropic two-point correlation function: Estimation of $w_{DE}$ and $b$	113
4.1.2	Bispectrum: Bias parameter estimation . . . . .	115
4.2	Improvements of the model for the bispectrum . . . . .	117
4.3	Application to data sets . . . . .	118
4.3.1	Anisotropic two-point correlation function . . . . .	118
4.3.2	Bispectrum . . . . .	119
<b>A</b>	<b>Appendix</b>	<b>121</b>
A.1	Analytic solution for the bispectrum in redshift space . . . . .	121
A.2	Derivation of the non-linear redshift space distortions for $\xi(r_p, \pi)$ . . . . .	123
A.3	Legendre polynomials . . . . .	125
A.4	Spherical Bessel functions . . . . .	125
A.5	Bias parameter estimation with collapsed triangles . . . . .	126
	<b>Acknowledgements</b>	<b>137</b>
	<b>Curriculum vitae</b>	<b>139</b>

# List of Figures

1.1	SDSS-III galaxy map . . . . .	3
2.1	WMAP 5-year temperature power spectrum . . . . .	13
2.2	Power spectrum and bispectrum for a Gaussian density field . . . . .	20
2.3	Comparison: Linear power sepctrum vs. non-linear power spectrum . . . . .	26
2.4	Bispectra comparison: usage of the linear PS vs. non-linear PS . . . . .	28
2.5	F-kernel at $z = 1.0$ for EdS and $\Lambda$ CDM universe in real space . . . . .	29
2.6	F-kernel at $z = 0.0$ for EdS and $\Lambda$ CDM universe in real space . . . . .	30
2.7	Bias parameters: Power spectrum vs. bispectrum . . . . .	32
2.8	Effect of the $b_2 b_1^2$ -term . . . . .	34
2.9	F-kernel at $z = 1.0$ for EdS and $\Lambda$ CDM universe in redshift space . . . . .	41
2.10	Dark matter halo $\xi_{rs}(r_p, \pi)$ at $z = 0.5$ . . . . .	46
2.11	Dark matter halo $\xi_{zs}(r_p, \pi)$ at $z = 0.5$ . . . . .	48
2.12	Dark matter halo $\xi_{zerr}(r_p, \pi)$ for 3% redshift errors at $z = 0.5$ . . . . .	50
2.13	Slice through L-BASICC simulation number 48 at $z = 0.5$ . . . . .	52
3.1	Cuts through $\xi_{rs}(r_p, \pi)$ at $z = 0.5$ . . . . .	57
3.2	Contours of $\xi_{rs}(r_p, \pi)$ and $\xi(r_p, \pi)$ at $z = 0.5$ . . . . .	58
3.3	Cuts through $\xi_{zs}(r_p, \pi)$ at $z = 0.5$ . . . . .	59
3.4	Contours of $\xi_{zerr}(r_p, \pi)$ for $\sigma_z = 0.015, 0.03, 0.06$ and $0.12$ at $z = 0.5$ . . . . .	61
3.5	Cuts through $\xi_{zerr}(r_p, \pi)$ for $\sigma_z = 0.015$ at $z = 0.5$ . . . . .	62
3.6	Cuts through $\xi_{zerr}(r_p, \pi)$ for $\sigma_z = 0.03$ at $z = 0.5$ . . . . .	63
3.7	Cuts through $\xi_{zerr}(r_p, \pi)$ for $\sigma_z = 0.06$ at $z = 0.5$ . . . . .	64
3.8	Cuts through $\xi_{zerr}(r_p, \pi)$ for $\sigma_z = 0.12$ at $z = 0.5$ . . . . .	65
3.9	Results of $w_{DE}$ and $b$ summarized for $\xi(r_p, \pi)$ . . . . .	65
3.10	$w(r_p)$ for real, redshift and redshift error space with different $\sigma_z$ s . . . . .	66
3.11	$\Lambda$ CDM, best-fitting and L-BASICC $w(r_p)$ for $\pi_{max} = 298.5 h^{-1}$ Mpc . . . . .	68
3.12	$\Lambda$ CDM, best-fitting and L-BASICC $w(r_p)$ for $\pi_{max} = 163.5 h^{-1}$ Mpc . . . . .	69
3.13	Results of $w_{DE}$ and $b$ summarized for $w(r_p)$ . . . . .	69
3.14	Equilateral bispectra: dark matter at $z = 0.5$ and $z = 1.0$ . . . . .	73
3.15	Dark matter bispectrum: Ratio between theory and measurement . . . . .	73
3.16	Kaiser boost-factor for the L-BASICC dark matter $B_{eq}(k)$ . . . . .	75
3.17	Equilateral bispectra: dark matter halos at $z = 0.5$ and $z = 1.0$ . . . . .	78

3.18	Dark matter halo bispectrum: Ratio between theory and measurement . . .	78
3.19	Kaiser boost-factor for the L-BASICC dark matter halo $B_{eq}(k)$ . . . . .	79
3.20	$Q_{eq}(k)$ : Dark matter at $z = 0.5$ and $z = 1.0$ . . . . .	81
3.21	$B_{dm}(k_1, k_2, k_3)/B_{dm,theo}(k_1, k_2, k_3)$ versus the <i>configuration id</i> for $z = 0.0$ . .	84
3.22	$B_{dm}(k_1, k_2, k_3)/B_{dm,theo}(k_1, k_2, k_3)$ versus the <i>configuration id</i> for $z = 0.5$ . .	84
3.23	$B_{dm}(k_1, k_2, k_3)/B_{dm,theo}(k_1, k_2, k_3)$ versus the <i>configuration id</i> for $z = 1.0$ . .	85
3.24	$b_1/b_2$ against $k_{max}$ for $B_{dm}(k_1, k_2, k_3)$ at $z = 0.0$ . . . . .	88
3.25	$b_1/b_2$ against $k_{max}$ for $B_{dm}(k_1, k_2, k_3)$ at $z = 0.5$ . . . . .	88
3.26	$b_1/b_2$ against $k_{max}$ for $B_{dm}(k_1, k_2, k_3)$ at $z = 1.0$ . . . . .	89
3.27	$\sigma_v$ against $k_{max}$ for $B_{dm}(k_1, k_2, k_3)$ at $z = 0.0, 0.5$ and $1.0$ . . . . .	89
3.28	$b_1/b_2$ against $k_{max}$ for $B_{dm,lin}(k_1, k_2, k_3)$ at $z = 0.0$ . . . . .	92
3.29	$b_1/b_2$ against $k_{max}$ for $B_{dm,lin}(k_1, k_2, k_3)$ at $z = 0.5$ . . . . .	93
3.30	$b_1/b_2$ against $k_{max}$ for $B_{dm,lin}(k_1, k_2, k_3)$ at $z = 1.0$ . . . . .	93
3.31	$Q(k_1, k_2, k_3)/Q_{theo}(k_1, k_2, k_3)$ versus the <i>configuration id</i> for $z = 0.0$ . . . .	95
3.32	$Q(k_1, k_2, k_3)/Q_{theo}(k_1, k_2, k_3)$ versus the <i>configuration id</i> for $z = 0.5$ . . . .	95
3.33	$Q(k_1, k_2, k_3)/Q_{theo}(k_1, k_2, k_3)$ versus the <i>configuration id</i> for $z = 1.0$ . . . .	96
3.34	$B_{halo}(k_1, k_2, k_3)/B_{halo,theo}(k_1, k_2, k_3)$ versus the <i>configuration id</i> for $z = 0.0$ .	97
3.35	$B_{halo}(k_1, k_2, k_3)/B_{halo,theo}(k_1, k_2, k_3)$ versus the <i>configuration id</i> for $z = 0.5$ .	97
3.36	$B_{halo}(k_1, k_2, k_3)/B_{halo,theo}(k_1, k_2, k_3)$ versus the <i>configuration id</i> for $z = 1.0$ .	98
3.37	$b_1/b_2$ against $k_{max}$ for $B_{halo}(k_1, k_2, k_3)$ at $z = 0.0$ . . . . .	98
3.38	$b_1/b_2$ against $k_{max}$ for $B_{halo}(k_1, k_2, k_3)$ at $z = 0.5$ . . . . .	99
3.39	$b_1/b_2$ against $k_{max}$ for $B_{halo}(k_1, k_2, k_3)$ at $z = 1.0$ . . . . .	99
3.40	Kaiser boost-factor for $B_{halo}(k_1, k_2, k_3)$ for $z = 0.0, 0.5$ and $1.0$ . . . . .	102
3.41	Theoretical estimation of the variance . . . . .	104
3.42	$B_{LAEs}(k_1, k_2, k_3)/B_{LAEs,theo}(k_1, k_2, k_3)$ versus the <i>configuration id</i> for $z = 2.2$	105
3.43	$b_1/b_2$ against $k_{max}$ for $B_{LAEs}(k_1, k_2, k_3)$ at $z = 2.2$ . . . . .	105
3.44	$B_{LAEs}(k_1, k_2, k_3)/B_{LAEs,theo}(k_1, k_2, k_3)$ versus the <i>configuration id</i> for $z = 3.0$	106
3.45	$b_1/b_2$ against $k_{max}$ for $B_{LAEs}(k_1, k_2, k_3)$ at $z = 3.0$ . . . . .	107
3.46	$\sigma_v$ against $k_{max}$ for $B_{LAEs}(k_1, k_2, k_3)$ at $z = 2.2$ and $z = 3.0$ . . . . .	108
3.47	Kaiser boost-factor for $B_{LAEs}(k_1, k_2, k_3)$ for $z = 2.2$ and $z = 3.0$ . . . . .	109
A.1	Bias parameter estimation with collapsed configurations at $z = 0.0$ . . . .	127
A.2	Bias parameter estimation with collapsed configurations at $z = 0.5$ . . . .	127
A.3	Bias parameter estimation with collapsed configurations at $z = 1.0$ . . . .	127

# List of Tables

2.1	Parameters and specifications of the L-BASICC II simulation . . . . .	14
3.1	Results for $w_{DE}$ and $b$ from $\xi_{zs}(r_p, \pi)$ at $z = 0.5$ . . . . .	58
3.2	Results for $w_{DE}$ and $b$ from $\xi_{zerr}(r_p, \pi)$ at $z = 0.5$ . . . . .	60
3.3	Results for $w_{DE}$ and $b$ from $w(r_p)$ at $z = 0.5$ . . . . .	67
3.4	Bias parameters $b_1$ and $b_2$ from $Q_{red,eq}(k)$ at $z = 0.5$ and $1.0$ . . . . .	76
3.5	Bias parameters $b_1$ and $b_2$ from $Q_{red,eq}^h(k)$ at $z = 0.5$ and $1.0$ . . . . .	77
3.6	Bias parameters $b_1$ and $b_2$ from $B_{dm}(k_1, k_2, k_3)$ at $z = 0.0, 0.5$ and $1.0$ . . . . .	87
3.7	Bias parameters $b_1$ and $b_2$ from $B_{halo}(k_1, k_2, k_3)$ at $z = 0.0, 0.5$ and $1.0$ . . . . .	99
3.8	Bias parameters $b_1$ and $b_2$ from $B_{LAEs}(k_1, k_2, k_3)$ at $z = 2.2$ and $3.0$ . . . . .	107
3.9	Bias parameters $b_1$ and $b_2$ from $P_{LAEs}(k)$ at $z = 2.2$ and $z = 3.0$ . . . . .	107



In dieser Arbeit wird das Clustering von Materie auf großen Skalen ( $\gtrsim 64 h^{-1}\text{Mpc}$ ) mit Hilfe der Zweipunkt- und Dreipunktstatistik untersucht. Dabei wird im ersten Teil der Arbeit ein Hauptaugenmerk auf die Zweipunktstatistik gelegt und im zweiten Teil auf die Dreipunktstatistik, um den Zustandsgleichungsparameter der dunklen Energie sowie die Biasparameter bestimmen zu können.

Die anisotrope Zweipunkt-Korrelationsfunktion  $\xi(r_p, \pi)$ , bei der der Abstand zwischen zwei Objekten in eine Komponente senkrecht ( $r_p$ ) sowie parallel ( $\pi$ ) zur Sichtlinie aufgespalten wird, wird dabei verwendet um den Zustandsgleichungsparameter der dunklen Energie  $w_{DE}$  und den linearen Bias  $b$  zu bestimmen. Zum ersten Mal werden nichtlineares Strukturwachstum als auch nichtlinearer kohärenter Einfall für  $\xi(r_p, \pi)$  modelliert, um Skalen größer als  $r \gtrsim 64 h^{-1}\text{Mpc}$  möglichst realistisch beschreiben zu können. Zusätzlich ermöglicht dieses neue Modell die Rotverschiebungsfehler zu berücksichtigen, welche zu einer Verzerrung der Zweipunktstatistik führen: Mit der Einführung sehr großer, weitwinkliger photometrischer Galaxiendurchmusterungen wie Pan-STARRS, DES oder PAU ist es äußerst wichtig geworden, den Einfluss von Ungenauigkeiten bei der Messung der Rotverschiebung zu verstehen. Die Gültigkeit des Modells wird mit Hilfe von 50 groß-volumigen numerischen  $N$ -Körper-Simulationenskuben mit mittlerer Auflösung getestet. Dabei wird eine Analyse nicht nur im realen und Rotverschiebungs-Raum durchgeführt, sondern auch im Rotverschiebungsfehlerraum, unter der Annahme von gaußverteilten Rotverschiebungsfehlern. Das Modell wird bezüglich vier verschiedener absoluter rms-Werte, nämlich  $\sigma_z = 0.015, 0.03, 0.06$  sowie  $0.12$ , untersucht. Bei dem gegebenen Volumen ( $V_{box} = 2.4 h^{-3}\text{Gpc}^3$ ) und der gegebenen Anzahldichte ( $\bar{n} \approx 1.25 \times 10^{-4} (h/\text{Mpc})^3$ ) der Objekte kann geschlussfolgert werden, dass die Form von  $\xi(r_p, \pi)$  gut genug bestimmt werden kann, um unverfälschte Werte für  $w_{DE}$  sowie  $b$  zu erhalten, selbst bei Rotverschiebungsfehlern von  $\sigma_z = 0.06$ .

Der zweite Teil der Arbeit befasst sich mit der Dreipunktstatistik im Fourierraum, dem Bisppektrum  $B(k_1, k_2, k_3)$ . Für die Bestimmung kosmologischer Parameter wie  $w_{DE}$  benötigt man die Kenntnis des Galaxienclustering verglichen mit dem des darunterliegenden Materiedichtefeldes. Diese sogenannten Biasparameter beeinflussen das Bisppektrum sehr stark; es ist jedoch frei von der Entartung zwischen  $b_1$  and  $\sigma_8$ , welche in der Zweipunktstatistik vorliegt. Daher ist das Bisppektrum bestens geeignet um die Biasparameter zu bestimmen. Allerdings ist es im Bereich der Kosmologie noch nicht so gründlich untersucht worden wie die Zweipunktstatistik. Der Zweck dieser Arbeit ist es daher, die Auswirkungen des Strukturwachstums sowie der Eigengeschwindigkeiten auf das Bisppektrum zu verstehen, im Hinblick auf eine spätere Anwendung auf beobachtete Daten. Daher wurden für eine bessere Beschreibung des Strukturwachstums die linearen Leistungsspektren, die in der ursprünglichen Modellierung gebraucht wurden, durch die Leistungsspektren der Störungstheorie dritter Ordnung ersetzt. Diese Modifizierung wird dann mit Hilfe der selben numerischen  $N$ -Körper-Simulationen getestet wie im Falle von  $\xi(r_p, \pi)$ . Dabei werden die Rotverschiebungen  $z = 0.0, 0.5$  und  $1.0$  betrachtet, um mögliche Abhängigkeiten dieser Effekte von der Rotverschiebung untersuchen zu können. Abschließend kann geschlussfolgert werden, dass das Bisppektrum auf großen Skalen ( $k_{max} \lesssim 0.10 h \text{Mpc}^{-1}$ ) sowie bei hohen Rotverschiebungen ( $z \gtrsim 0.5$ ) zuverlässig modelliert werden kann, wohingegen das Modell für niedrigere Rotverschiebungen und größere Moden verbessert werden muss.



In this thesis, the clustering of matter on large scales ( $\gtrsim 64 h^{-1}\text{Mpc}$ ) by means of the two- and three-point statistics will be investigated. The goal is to use these as a means to extract the equation of state parameter of dark energy and bias parameters.

In the first part of this thesis, the anisotropic two-point correlation function  $\xi(r_p, \pi)$ , where the distance between two objects is split into a component perpendicular to line-of-sight ( $r_p$ ) and parallel to the line-of-sight ( $\pi$ ), is examined. This analysis is used to extract the equation of state parameter of dark energy  $w_{DE}$  and the linear bias  $b$ . Non-linear structure growth and non-linear coherent infall is for the first time incorporated in the model of  $\xi(r_p, \pi)$  with a realistic description at scales larger than  $r \gtrsim 64 h^{-1}\text{Mpc}$ . The new model introduced in this thesis is also able to account for redshift errors which distort the two-point clustering statistic: The advent of very large, wide-angle photometric redshift surveys like e.g. Pan-STARRS, DES, or PAU, made it extremely important to understand the influence of redshift inaccuracies. In order to investigate the validity of the model it is tested against a suite of 50 large-volume, medium resolution numerical  $N$ -body simulation boxes where an analysis can be performed in real and redshift space as well as in redshift error space. The redshift error distribution can safely be assumed to be Gaussian and the model is tested for four different absolute rms values with  $\sigma_z = 0.015, 0.03, 0.06, \text{ and } 0.12$ . Such an investigation has never before been performed on  $\xi(r_p, \pi)$ . It is concluded that for the given volume ( $V_{box} = 2.4 h^{-3}\text{Gpc}^3$ ) and number density ( $\bar{n} \approx 1.25 \times 10^{-4} (h/\text{Mpc})^3$ ) of objects, the full shape of  $\xi(r_p, \pi)$  is modeled accurately enough that it can be used to derive unbiased constraints on the equation of state parameter of dark energy  $w_{DE}$  and the linear bias  $b$ , even in the presence of redshift errors of the order of  $\sigma_z = 0.06^1$ .

The second part of this thesis focuses on three-point statistics, mainly the Fourier space the bispectrum  $B(k_1, k_2, k_3)$ . The estimation of the cosmological parameters such as  $w_{DE}$  requires knowledge of the clustering of galaxies compared to the underlying matter density field. The bispectrum is sensitive to these bias parameters and does not suffer from the same the degeneracy between  $\sigma_8$  and the linear bias as the two-point statistics. So it is perfectly suited to measure bias parameters. However, until now the bispectrum has not been as thoroughly examined as two-point statistics in the field of cosmology. Hence, the purpose of this part of the thesis is to understand the effects of structure growth and peculiar velocities on the bispectrum for dark matter, with regard to a later application to real data sets. For a better description of structure growth, the linear power spectra from the original model of the bispectrum are replaced by the corresponding  $3^{rd}$  order perturbation theory power spectra to account for the effects of non-linearities. This modified model is tested against the same numerical  $N$ -body simulation boxes used to test  $\xi(r_p, \pi)$  for the snapshots at  $z = 0.0, 0.5 \text{ and } 1.0$ , in order to understand the evolution with redshift of these effects. It is concluded that the large scale bispectrum ( $k_{max} \lesssim 0.10 h \text{ Mpc}^{-1}$ ) can be modeled reliably at least at high redshifts ( $z \gtrsim 0.5$ ), while at lower redshifts and larger  $k$ -modes the current model has to be improved.

---

<sup>1</sup>from Schlagenhauer et al. (2012)



# Chapter 1

## Introduction

Our understanding of the Universe underwent a dramatic change over the last one hundred years. In the beginning of the 20<sup>th</sup> century, it was believed that the Universe contains only normal matter, which means mostly baryonic matter. In several observations, it was shown that there is not enough baryonic matter to explain e.g. the rotation curves of nearby galaxies or the Milky Way, where the velocities of the stars do not follow Kepler's law (Bosma, 1978; Rubin et al., 1980), the spatial extent of clusters of galaxies measured with X-ray emission from the hot intergalactic gas (White et al., 1993) and the effect of gravitational lensing, where the visible matter is not enough to explain the lensing effects (Clowe et al., 2006). The missing matter which is required to explain these observations is called dark matter and interacts with its environment only gravitationally.

Another important discovery was made in the late nineties of the last century. For the first time, the late time accelerated expansion of the Universe was inferred from the observed luminosities of Supernovae Type Ia which were too faint for a non-accelerated Universe (Riess et al., 1998; Perlmutter et al., 1999). An additional component of the energy density of the Universe is required for to be responsible for the acceleration. In this thesis, it is assumed that this additional component is the so-called dark energy. Different forms of dark energy can be characterized by the equation of the state parameter

$$w_{DE} \equiv \frac{p_{DE}}{\rho_{DE}c^2} \quad (1.1)$$

where  $p_{DE}$  is the pressure and  $\rho_{DE}$  the energy density of the dark energy. Current measurements of  $w_{DE}$  are still consistent with the simplest solution, the cosmological constant ( $w_{DE} = -1.0$ ), with around ten percent errors (Sánchez et al., 2006; Komatsu et al., 2009, 2011; Montesano et al., 2011). The cosmological constant was first introduced by Einstein to ensure a steady state universe. However, after the discovery of Edwin Hubble that galaxies move away from us (Hubble, 1929) Einstein rejected it. Another possible explanation for dark energy could be the so-called quintessence models, in which it is assumed that a scalar field is present throughout the whole Universe. Furthermore, it is possible to modify gravity itself, e.g. with higher dimensional approaches as in string cosmology, instead of introducing another particle or field to the energy momentum tensor. These theoretical

approaches are reviewed in Copeland et al. (2006).

Understanding the physical origin of the late time accelerated expansionary phase of the Universe is one of the most important challenges in the field of cosmology nowadays. The clustering of galaxies is an important source of information to learn more about the constituents of the Universe (Albrecht et al., 2006; Sánchez et al., 2006; Montesano et al., 2011). In this thesis, this clustering is examined by means of the two- and three-point statistics, where the density field is traced using galaxies in two or three different locations, respectively.

**Baryonic Acoustic Oscillations (BAOs)**, which formed via acoustic waves in the primordial plasma right after the big bang, are a very important feature in the galaxy clustering. They can be used as a standard ruler, comparable to the Supernovae Type Ia, which are considered to be standard candles. For the extraction of e.g.  $w_{DE}$  a relation between redshift and distance is used. This is dependent on the cosmological parameters which therefore can be inferred from a known redshift-distance relation (Blake & Bridle, 2005). The physical origin of BAOs is well understood and described by plasma physics (Hu et al., 1995; Hu & White, 1996). In the very early Universe ( $t \sim 10^{-32} s$  after the big bang), when dark energy was negligible, small perturbations in the density field were present. From each of these perturbations sound waves propagated away. Due to the tight coupling between the photons and electrons via Thomson scattering at that time, the baryonic matter was dragged by the photons. At the time of recombination, when neutral hydrogen was formed (no free electrons were available anymore), these acoustic waves were imprinted in the matter density distribution due to the rapid decrease of the sound speed towards zero. The maximum distance these acoustic waves were able to propagate is called the sound horizon and can be used as a standard ruler. These frozen acoustic waves can be observed in the **Cosmic Microwave Background (CMB)** (Lange et al., 1995; Hinshaw et al., 2003). At the scale of the sound horizon, a slight overdensity in the matter distribution can be found. In general, in overdense regions galaxies are more likely to form than in underdense ones (Kaiser, 1984). Therefore, a higher galaxy density is expected at separations equal to the scale of the sound horizon than at slightly smaller or larger scales. This behavior is assumed to be also reflected in clustering statistics.

In Figure 1.1<sup>2</sup>, a map of luminous red galaxies from SDSS-III is shown. The red circle indicates the imprinted scale of the BAOs. This shows that the BAOs are a large scale feature. Therefore, a large volume with as many galaxies as possible is required to measure a clear BAO signal, which can then be used to extract cosmological parameters. In Eisenstein et al. (2005), the BAOs were observed for the first time in the galaxy two-point correlation function as a single broad peak. The two-point statistics are described by the two-point correlation function or power spectrum, depending on whether they are considered in configuration space or Fourier space, respectively. In Fourier space, this BAO peak is transformed into a series of small wiggles on top of the overall shape of the power spectrum. Analogously, the three-point statistics are described by the three-point correlation function in configuration space and the bispectrum in Fourier space.

---

<sup>2</sup>Press release for SDSS-III: <http://www.astronomy.ohio-state.edu/~dhw/SDSS3/sdss3pr.html>

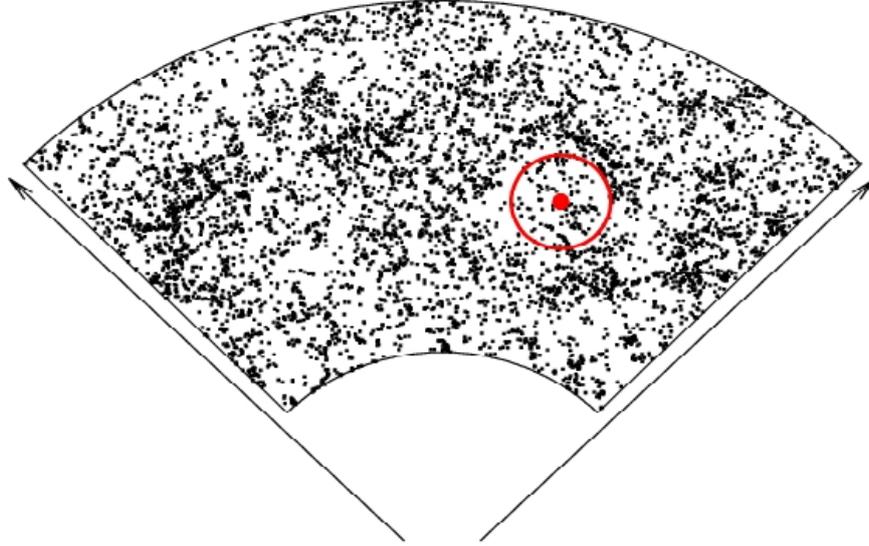


Figure 1.1: A map of the luminous red galaxies from SDSS-III. The red circle indicates the size of the BAOs in the galaxy survey.

The extraction of features imprinted by BAOs from the galaxy distribution, and derivation of cosmological parameters is a main objective of many galaxy surveys, such as the two-degree-field galaxy redshift survey (2dFGRS) (Colless et al., 2001) and the **S**loan **D**igital **S**ky **S**urvey SDSS (Eisenstein et al., 2001). These two surveys are examples of spectroscopic surveys which are able to deliver high precision galaxy distances. Each galaxy must be exposed a certain amount of time which can be very time consuming (dependent on the desired quality of the spectra and the sensitivity of the instrument). From their spectra the distance to the objects can be obtained from the shift of the measured wavelengths of known spectral features compared to the laboratory. Due to the expansion of the Universe these features are shifted towards redder frequencies.

Photometric surveys are a different and much faster method for observing a large number of galaxies covering large volumes in which photometric redshifts are derived from observed fluxes in five or more broad to medium band filters (Baum, 1962). This results in a distorted clustering of the galaxies due to the uncertainties of their distances. The **P**anoramic **S**urvey **T**elescope **A**nd **R**apid **R**esponse **S**ystem Pan-STARRS (Chambers & Pan-STARRS Team, 2004), the upcoming **D**ark **E**nergy **S**urvey DES (Tucker et al., 2010), or PAU (**P**hysics of the **A**ccelerating **U**niverse, see Benítez et al. (2009)) are examples of photometric redshift surveys.

In order to extract unbiased cosmological parameters from these surveys an accurate modeling of clustering statistics is required, which is also important for the modeling on the scales of the BAOs. The first problem which arises is in the description of the growth of structure. The coupling of growth on various scales drives non-linear growth which cannot be solved precisely by theoretical modeling. However, this is required for the formation of structures which can be observed in the Universe. The second problem is related to the

peculiar velocities of galaxies which mainly distort clustering in the line-of-sight direction. This coordinate frame is also called redshift space and the effect of the peculiar velocities is called redshift space distortions. If the peculiar velocities of the galaxies were zero and their measured positions were not distorted (which is only possible in simulations but not for real observations), this would be called real space. In the presence of large redshift errors, as it is the case for photometric surveys, the clustering statistics is further distorted along the line-of-sight. This situation is called redshift error space. On top of all these issues, galaxies do not trace the underlying matter density field perfectly. The ratio between these two density fields is called bias (Kaiser, 1984). In the most simple case, it is assumed to be a scale independent quantity (Fry, 1996). All these problems must be considered in the modeling of clustering statistics in order to obtain unbiased results for the cosmological parameters.

In this thesis, the BAO feature will be used to infer the parameter of the equation of state of dark energy  $w_{DE}$  and the linear bias by means of the anisotropic two-point correlation function  $\xi(r_p, \pi)$ , where the BAO bump becomes a ring. For  $\xi(r_p, \pi)$  the distance between two galaxies is split into two components,  $r_p$  and  $\pi$ , which are perpendicular and parallel to the line-of-sight, respectively. The information content of  $\xi(r_p, \pi)$  is larger than that of the spherically averaged two-point correlation function  $\xi(r)$  (where  $r = \sqrt{r_p^2 + \pi^2}$ ), because the whole information obtained from clustering is not only condensed in a few data points. For the first time, the model of  $\xi(r_p, \pi)$  presented in this thesis takes into account non-linear structure growth, non-linear redshift space distortions as well as redshift errors in order to make use of clustering information which can be extracted from photometric redshift surveys. Then,  $w_{DE}$  can be determined with the help of the distance-redshift relation, making use of the Alcock-Paczynski effect (Alcock & Paczynski, 1979).

In this thesis, only  $w_{DE}$  is determined in that way, while linear bias is extracted from the amplitude of  $\xi(r_p, \pi)$ . If more cosmological parameters are varied, the linear bias will be degenerate with the amplitude of the dark matter density field. This will also introduce degeneracies with other cosmological parameters. In general, this is a problem for all two-point statistics. In order to obtain tight constraints on these parameters this degeneracy must be broken.

The three-point statistics offer a solution to this problem, because the linear bias and the amplitude of the matter density field affect them differently. From an analysis of three-point statistics, not only the linear bias can be measured, but also the quadratic bias, because these two parameters have different effects on the three-point statistics. However, it has not been as thoroughly examined as two-point statistics. Therefore, an intensive investigation is required, before a combined analysis of these two- and three-point clustering statistics can be realized.

In this thesis, the bispectrum  $B(k_1, k_2, k_3)$  will be used for this examination. The investigation will not only be performed for one specific triangle configuration, as it is the case in other publications, but for all configurations up to a predefined maximum  $k$ -mode. The focus will lie on an understanding of non-linear structure growth and redshift space distortions by determining the linear bias as well as the quadratic bias. These effects must be studied first before the measurement errors can be incorporated into the model of

---

$B(k_1, k_2, k_3)$ .

These are the two main topics of this thesis, split into three main chapters and structured as follows: In Chapter 2, the required definitions, derivations and methods are introduced for the later analysis of  $\xi(r_p, \pi)$  and  $B(k_1, k_2, k_3)$ . In this chapter, the clustering statistics will be derived in real space, see Section 2.6.1. Following this, the effect of peculiar velocities of galaxies induced by gravity is investigated theoretically, see Section 2.6.2. In all of these cases the biased nature of galaxies will be taken into account. The discussion in these sections will be focused on Fourier space clustering quantities. In Section 2.7, configuration space is examined with a focus on  $\xi(r_p, \pi)$ , including the incorporation of redshift errors. This issue will be excluded from the analysis of the three-point clustering statistics, because the extracted parameters from such statistics are expected to be poorly determined due to the low signal-to-noise ratio. In any case, before measurement errors can be taken into account, it is important to understand the real and redshift space bispectrum first. To conclude the theoretical part of this thesis, the results obtained from the clustering analysis is presented in Chapter 3. First, the results of the determination of the parameter for the equation of state of dark energy and the linear bias, by means of  $\xi(r_p, \pi)$ , are given in Section 3.1. This analysis is followed by Section 3.2, including a detailed investigation of the bispectrum, and the extraction of bias parameters. This thesis is concluded with Chapter 4. Results obtained from Chapter 3 are summarized in Section 4.1. Possible improvements for modeling of the bispectrum are introduced in Section 4.2, and future application to real data sets are proposed in Section 4.3.



## Chapter 2

# Theoretical framework for the two- and three-point statistics

In this chapter, the theoretical framework will be prepared for the calculation of the two- and three-point clustering statistics. For their derivation the distance between the objects is of interest (e.g. halos or galaxies) as well as their distance to the observer must be known. Section 2.1 is focused on the evaluation of cosmological distances. As mentioned in Chapter 1, the BAOs serve as a standard ruler and can be utilized to measure cosmological parameters e.g.  $w_{DE}$  via the distance-redshift relation. The physical origin of the BAOs will be discussed in Section 2.2. The discussion of the following sections will mainly be performed in Fourier space because the derivations of the clustering statistics can be carried out easier in this framework.

In Section 2.3, the numerical  $N$ -body simulations used for testing the validity of the modeling will be introduced. The required definitions for measuring and modeling the clustering statistics, the power spectrum  $P(k)$  and the bispectrum  $B(k_1, k_2, k_3)$ , will be given in Section 2.4. The measurement of  $P(k)$  and  $B(k_1, k_2, k_3)$  will be discussed in Section 2.5. The extraction of cosmological parameters or bias parameters can only be achieved if a reliable model of the clustering statistics is available. Section 2.6 is mainly interested in this topic for real and redshift space as well as for biased objects like halos or galaxies.

The results derived in the previous sections have a configuration space counterpart, which can be evaluated by a back Fourier transformation of  $P(k)$  and  $B(k_1, k_2, k_3)$ . The quantities which will be obtained by this process are called two-point correlation function  $\xi(r)$  and three-point correlation function  $\zeta(r_{12}, r_{23}, r_{13})$ , respectively. The derivations in Section 2.7 are mainly interested in the anisotropic two-point correlation function  $\xi(r_p, \pi)$  in configuration space.

With the conclusion of this chapter all the needed theoretical frameworks are set up for an application to  $N$ -body simulations or real data sets.

## 2.1 Cosmological distances

For mapping the spatial distribution of the galaxies the distance to these galaxies must be known. This section will be focused on the calculation of cosmological distances. This requires an understanding of the dynamics of cosmology.

In this thesis, it is assumed that the cosmological expansion is driven by dark energy, as introduced in Chapter 1. The cosmological dynamics are given by the field equations of General Relativity

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R} = \frac{8\pi G}{c^4}T_{\mu\nu} + \Lambda g_{\mu\nu} \quad (2.1)$$

where  $G$  is Newton's constant,  $\pi$  ( $\approx 3.14$ ) is the constant ratio between the circumference and the diameter of a circle,  $R_{\mu\nu}$  the Ricci tensor (Einstein, 1915), which calculates the space-time curvature,  $\mathcal{R} \equiv g^{\mu\nu}R_{\mu\nu}$ ,  $g_{\mu\nu}$  the metric tensor in the Einstein choice which describes the space-time distance. For example  $g_{\mu\nu}$  is called  $\eta_{\mu\nu}$  for the Minkowski space-time and is given by

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

$\Lambda$  is the cosmological constant and  $T_{\mu\nu}$  the energy momentum tensor. The left hand side of Equation (2.1) describes the geometry of space-time and the right hand side the energy and matter content of it. This thesis is only interested in the large scales of the Universe. From observations it can be assumed, that the matter is distributed homogeneously and isotropically on scales larger than  $100 h^{-1}$  Mpc. The energy-momentum tensor for such a case can be written as

$$T_{\mu\nu} = \frac{1}{c^2} \begin{pmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & -p & 0 & 0 \\ 0 & 0 & -p & 0 \\ 0 & 0 & 0 & -p \end{pmatrix}$$

where  $c$  is the speed of light in vacuum ( $c = 299\,792\,458\text{ m/s}$ )<sup>3</sup>,  $\rho$  the energy density and  $p$  the pressure of the matter component, which can be dark, luminous, hot (particles with a relativistic speed,  $v \approx c$ ), cold (particles with non-relativistic speed) or warm matter. These components can be quantified by their current fraction of the energy density of the Universe which is defined as

$$\Omega_i \equiv \frac{3H_0^2 \rho_i}{8\pi G} \quad (2.2)$$

where  $\rho_i$  is the energy density of a certain particle species and  $H_0$  the Hubble parameter today. In this thesis, the index  $i$  can be replaced by  $CDM$  for cold dark matter,  $b$  for baryonic matter (where  $\Omega_M = \Omega_{CDM} + \Omega_b$ ) and  $\Lambda$  for dark energy.

The curvature of the Universe is measured to be consistent with a flat geometry (Sánchez et al., 2006; Spergel et al., 2007; Komatsu et al., 2009, 2011; Montesano et al., 2011). The most

<sup>3</sup>[http://physics.nist.gov/cgi-bin/cuu/Value?c|search\\_for=speed+of+light](http://physics.nist.gov/cgi-bin/cuu/Value?c|search_for=speed+of+light)

general metric tensor in this case is given by the Robertson-Walker metric with zero curvature which is the only examined case in this thesis. The metric is given by

$$ds^2 = a^2(\tau) (d\tau^2 - d\chi^2 - \chi^2 d\Omega^2) \quad (2.3)$$

where the conformal time is defined as  $d\tau \equiv c \cdot dt/a(t)$  and the comoving distance  $\chi$  is given by

$$\chi(z) = \int_0^z \frac{c \cdot dz}{H(z)} = \int_a^1 \frac{c \cdot da}{a^2 H(a)} \quad (2.4)$$

where

$$H(z) = H_0 \sqrt{\left( \Omega_M (1+z)^3 + \Omega_\Lambda (1+z)^{3(1+w_{DE})} \right)} = \frac{\dot{a}}{a}. \quad (2.5)$$

The quantity  $z$  is called the redshift and is defined by the fractional shift in wavelength of a photon emitted at a time  $t_{emitted}$  and observed today

$$1+z = \frac{\lambda_{obs} - \lambda_{emitted}}{\lambda_{emitted}} = \frac{a_0}{a(t_{emitted})} \quad (2.6)$$

where  $a$  is the scale factor which is set to unity for today ( $a_0 = 1$ ). The physical distance of an object can be evaluated by multiplying the comoving distance, which can be computed by Equation (2.4), with the scale factor. Equation (2.4) is obtained by inserting  $ds^2 = 0$  in Equation (2.3), because photons travel along a geodesic (Einstein, 1915) like all massless particles.

The evolution of the scale factor is described by the Friedmann equations (Friedmann, 1922), which are obtained by inserting the Robertson-Walker metric into Equation (2.1). The second Friedmann equation was already given in Equation (2.5) and the first one describes the change of the time evolution of the scale factor  $a$  (the acceleration)

$$\ddot{a} = -\frac{4\pi G}{3} \left( \rho + \frac{3p}{c^2} \right) a + \frac{\Lambda a}{3}. \quad (2.7)$$

The comoving transverse size of an object or a structure is given by multiplying of Equation (2.4) with its angular extent on the sky. It is assumed that the object emitted the observed light at the same time. This means that  $d\tau = 0$  and  $d\chi = d\phi = 0$ , which is only a choice of the coordinate system. From the ratio of the physical transverse size of the object and the angular extent, the angular diameter distance can be defined as

$$D_A(z) \equiv \frac{a \cdot r_p}{\theta} = \frac{c}{1+z} \int_0^z \frac{dz}{H(z)} \quad (2.8)$$

where  $a \cdot r_p$  is the physical transverse size of the object. The comoving distance along the line-of-sight  $\pi$  and the comoving distance transverse to the line-of-sight  $r_p$  of an object or structure is then given by

$$\pi = \frac{c \cdot \Delta z}{H(z)} \quad (2.9)$$

and

$$r_p = (1 + z) \cdot D_A(z) \cdot \Delta\theta \quad (2.10)$$

where for Equation (2.9) it was assumed that the scale factor did not change along the line-of-sight extent of the object.

The observed quantities are  $\Delta z$  and  $\Delta\theta$  and with Equation (2.9) and Equation (2.10)  $H(z)$  and  $D_A(z)$  can be calculated. After knowing how distances are computed in cosmology, the physical origin of the peak of the correlation function must be understood in order to use it as a standard ruler. This discussion will take place in the next section.

## 2.2 Baryonic Acoustic Oscillations (BAOs)

A brief summary of the baryonic acoustic oscillations will be given here. In this section the same notation as in Hu et al. (1995) and Hu & White (1996) is used and the speed of light is set to unity ( $c = 1$ ). This discussion is focused on the physics in the early Universe at redshifts  $z > 1000$ .

At that time the Universe was in a very hot and dense stage. So far, no neutral atoms were built. Photons and electrons were able to interact with each other via Thomson scattering. The positive charged atomic nuclei were coupled via Coulomb scattering to electrons. Therefore, photons were able to drag baryonic matter. This period is also called drag epoch. The baryon-photon fluid can be described by plasma physics like a single fluid and it was optically thick. Due to Thomson scattering photons were not able to propagate freely. At that time already small fluctuations were present in the density field. These initial perturbations drove acoustic waves in the baryon-photon fluid. The overdense regions accreted matter and the photon-baryon fluid was heated. The pressure in that overdense region counteracted gravity and created acoustic waves. Photons moving out of such a perturbation lose energy and hence look redder than photons from an underdense region.

In the simplest approach, these acoustic waves can be characterized by a simple harmonic oscillator equation

$$\ddot{\Theta} + c_s^2 k^2 \Theta = 0 \quad (2.11)$$

written in Fourier space for the temperature fluctuations  $\Theta$  with  $c_s$  being the sound speed in the photon gas without any baryons and the dots denote derivatives with respect to the conformal time. In this description, the oscillations are neither driven nor damped. They just represent the heating and cooling of the fluid due to compression and rarefaction. In this zero order approximation the sound speed is given by  $c_s^2 = 1/3$  for the photon-dominated fluid (as a reminder  $c = 1$ ).

In configuration space, Equation (2.11) is solved by standing waves which are oscillating. This situation remained unchanged until the number density of the free electrons fell rapidly. This happened at the time of recombination where neutral hydrogen was formed. The imprint of these fluctuations on the CMB is also called last scattering surface.

After recombination photons and electrons are not tightly coupled anymore. This is also

the first time when the Universe became optically thin and the photons were able to propagate freely. The modes of the oscillations are caught in their extrema given by

$$k_n s_* = n \cdot \pi, \quad n = 1, 2, 3, \dots \quad (2.12)$$

where  $s_*$  is defined as the comoving sound horizon at recombination

$$s_* \equiv \int c_s d\tau. \quad (2.13)$$

The comoving sound horizon is the distance the sound waves were able to travel until recombination and only within the sound horizon the waves were able to oscillate. The compression of the plasma is represented by the odd extrema and the rarefaction by the even extrema.

### 2.2.1 Gravitation driven oscillations

Since the acoustic oscillations are driven by gravity, on the right hand side of Equation (2.11) the time-time and the space-space fluctuations of the metric  $\Psi$  and  $\Phi$  have to be added and then result in

$$F = -\frac{k^2}{3}\Psi - \ddot{\Phi}. \quad (2.14)$$

There is an equilibrium between the pressure gradient  $k\Theta$ , which is responsible for the rarefaction, and the gravity gradient  $k\Psi$ , which is responsible for the compression. If the gravitational potential is assumed to be constant  $\dot{\Phi} = 0$  and  $\ddot{\Phi} = 0$ , which is true as long as the energy density of the Universe is dominated by only one component, the harmonic oscillator can be rewritten as

$$\ddot{\Theta} + \ddot{\Psi} + c_s^2 k^2 (\Theta + \Psi) = 0 \quad (2.15)$$

and the minima of the solution of this equation are displaced compared with the solution of Equation (2.11).

The radiation energy density contribution at recombination is small but not negligible and thus  $\Phi$  varies slowly with time. The temperature fluctuations which are induced by the change of the gravitational potential due to radiation are called early integrated Sachs-Wolfe effect. In the matter dominated epoch of the Universe, the gravitational potential is constant again but when dark energy of whatever form begins to dominate the energy density of the Universe, the gravitational potential changes again. This temperature fluctuations which are induced by the change of the gravitational potential due to dark energy is called late integrated Sachs-Wolfe effect.

### 2.2.2 Baryon loading

The next important impact on the oscillations are the baryons which add extra mass to the momentum density of the plasma. For the relativistic plasma the momentum density

has to be changed from

$$(\rho_\gamma + p_\gamma) v_\gamma \quad \text{into} \quad (\rho_\gamma + p_\gamma) v_\gamma + (\rho_b + p_b) v_b \equiv (1 + R) (\rho_\gamma + p_\gamma) v_{\gamma b} \quad (2.16)$$

where

$$R \equiv \frac{(\rho_b + p_b)}{(\rho_\gamma + p_\gamma)} \quad (2.17)$$

and  $v_{\gamma b} \approx v_b \approx v_\gamma$  in the tight coupling regime where  $v_\gamma$  and  $v_b$  are the fluid velocities of the photons and baryons and  $v_{\gamma b}$  is the bulk velocity of the baryon-photon fluid in the tight coupling regime. The harmonic oscillator equation becomes

$$\left[ (1 + R) \ddot{\Theta} \right] + \frac{1}{3} k^2 \Theta = -\frac{1}{3} k^2 (1 + R) \Psi - \left[ (1 + R) \dot{\Phi} \right] \cdot \quad (2.18)$$

Equation (2.18) can be interpreted as an oscillator equation with a change of mass given by  $R$  and a decreased sound speed given by

$$c_s^2 = \frac{\dot{p}_\gamma + \dot{p}_b}{\dot{\rho}_\gamma + \dot{\rho}_b} = \frac{1}{3(1 + R)} \cdot \quad (2.19)$$

The sound horizon is also decreased because of the change of the sound speed. In the matter dominated regime and assuming that  $R$  changes slowly the baryon loading has the following effects on the oscillations. The amplitude is increased by a factor  $1 + 3R$  and the zero point of the oscillations is shifted so that even and odd peaks have different amplitudes. The odd peaks have a higher amplitude than the even peaks because an additional gravitational compression is induced due to the baryons.

There are many more effects which have to be taken into account in order to obtain a correct description of the BAOs e.g. damping, viscosity, heat conduction etc., but all these physical ingredients are well understood at this regime.

The sound horizon can be calculated within this framework because dark energy does not play any role near recombination. This is the reason why the sound horizon can be used as a standard ruler and finally at the drag epoch is given by

$$s = \int_0^{\tau_{drag}} c_s d\tau = \frac{2}{3k_{eq}} \sqrt{\frac{6}{R_{eq}}} \ln \left( \frac{\sqrt{1 + R_{eq}} + \sqrt{R_d + R_{eq}}}{1 + \sqrt{R_{eq}}} \right) \quad (2.20)$$

where it is only dependent on the wavenumber of the horizon at matter-radiation equality  $k_{eq}$ , the photon-baryon density at the drag epoch  $R_d$  and at matter-radiation equality  $R_{eq}$ . All of these parameters are only dependent on the baryon density, the matter density and the temperature of the CMB today.

At recombination the photons were not scattered anymore and diffuse in all directions. However, the matter density perturbations are stalled. The photons from the CMB contain the information of these fluctuations at that time. Due to the rapid decrease of free electrons the BAOs were frozen out on the last scattering surface which is at the same time the oldest possible information astronomers can get from the early Universe (as long as primordial neutrinos or gravitational waves are not observable).

### 2.2.3 Influence on the clustering statistics

Overdense regions in configuration space can be found more likely at the scale of the sound horizon, compared to a power law like decrease of the density field. The power law like behavior of the probability function of the density field is originated from inflation for which the power spectrum has the form  $\langle |\delta(k)|^2 \rangle \propto k^{n_s}$ , where  $n_s \approx 1$  (Mukhanov, 2005). This probability function is compared to a Poisson process in order to identify perturbations induced by gravity. Overdense regions are supposed to be the seeds of galaxies. The spatial distribution of the galaxies can be investigated by counting galaxies at a given separation. As mentioned in Chapter 1, the BAOs are present as a single broad peak in the two-point correlation function and as small wiggles in the power spectrum. In this thesis, the BAOs in the galaxies density field will be used to constrain the parameter of the equation of state of dark energy  $w_{DE}$ . In the discussion above, it was pointed out that the physical origin of the BAOs is well understood and that they are not influenced by dark energy. This is the reason why the BAOs can be used as a standard ruler.

In Figure 2.1, the WMAP 5-year temperature power spectrum is shown from Nolte et al. (2009). The red solid line is the best-fit theory for a  $\Lambda$ CDM cosmology on the WMAP alone. The extrema of the oscillations are now peaks because maxima and minima contribute to the power spectrum. The compression of the plasma is represented by the odd peaks and the rarefaction by the even peaks. The minima, which can be identified in Figure 2.1, are the velocity maxima of the plasma.

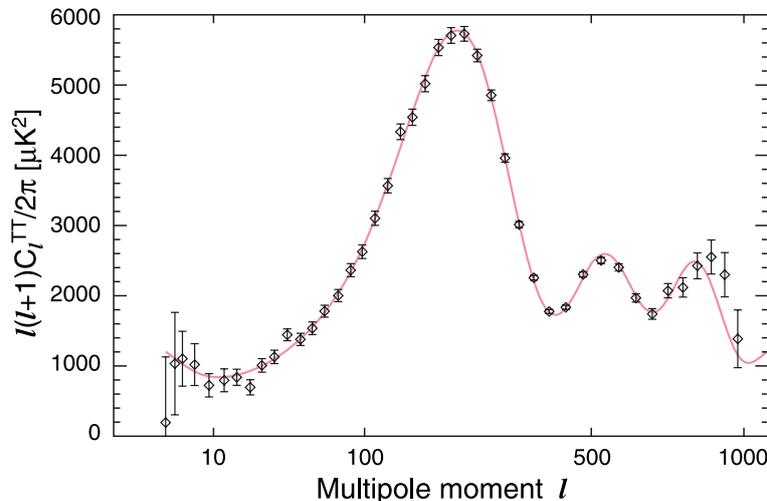


Figure 2.1: The WMAP 5-year temperature power spectrum for the best-fit parameters for a  $\Lambda$ CDM cosmology (red solid line) for WMAP (the measurement: black dots with error bars) alone (Nolta et al., 2009).

In the following sections, the two- and three-point clustering statistics in Fourier space will be defined and derived. In order to visualize some properties of these quantities and to perform already some smaller investigations on them, numerical  $N$ -body simulations

are required. For the calculation of the theoretical quantities also cosmological parameters are needed which will be chosen to be ones of the  $N$ -body simulations. Therefore, the simulations which will be used in the following will be introduced in the next section.

### 2.3 L-BASICC simulations

Numerical  $N$ -body simulations are commonly used to test the validity of models. Throughout this chapter, some of these tests will be carried out and this requires a set of fiducial cosmological parameters which will be introduced here as well.

The **L**ow resolution **B**aryonic **A**coustic oscillation **S**imulation at the **I**nstitute for **C**omputational **C**osmology II (L-BASICC II or just L-BASICC) is a numerical  $N$ -body simulation and is used to validate the theoretical predictions. Each of the 50 realizations of the L-BASICC II simulation done by Angulo et al. (2008) contains  $448^3$  dark matter particles with a mass of  $M_{dm} = 1.75 \cdot 10^{12} h^{-1} M_{\odot}$ . Halos are defined by at least 10 gravitationally bound dark matter particles which were found by a Friends-of-Friends (Davis et al., 1985) algorithm with a linking parameter of  $b_{link} = 0.2$  times the mean separation. A flat  $\Lambda$ CDM cosmology was chosen for the simulations to be consistent with the cosmological parameters derived from the CMB and large scale clustering of galaxies (Sánchez et al., 2006; Spergel et al., 2007). The initial density field, the seeds for the later structure growth, was set to be Gaussian distributed for each realization. The cosmology of this simulations will be used as fiducial cosmology for the following calculations if not stated otherwise. Snapshots at  $z = 0.0, 0.5$  and  $1.0$  were made to investigate the evolution of the density field. In Table 2.1, the parameters of the simulation can be found.

explanation	symbol	value
matter density	$\Omega_M$	0.237
baryonic density	$\Omega_b$	0.041
equation of state parameter for dark energy	$w_{DE}$	-1
scalar spectral index	$n_s$	0.954
Hubble parameter	$H_0$	$73.5 \frac{km}{s \text{ Mpc}}$
amplitude of the density perturbations	$\sigma_8$	0.773
box length	$L_{box}$	$1340 h^{-1} \text{ Mpc}$

Table 2.1: Table with the cosmological parameters and the specifications of the L-BASICC simulations.

## 2.4 Power spectrum and Bispectrum: Definition

In this section, the two- and three-point statistics will be evaluated in Fourier space, the so-called power spectrum and bispectrum, respectively. For the estimation of these clustering statistics not the real density field in configuration space is used but the overdensities  $\delta(\vec{r})$  which are defined as

$$\delta(\vec{r}) \equiv \frac{\rho(\vec{r}) - \bar{\rho}}{\bar{\rho}} \quad (2.21)$$

where  $\rho(r)$  is the local density field and  $\bar{\rho}$  the mean density field, if all the matter would be distributed homogeneously over the whole considered volume. This overdensity field is given in configuration space and must be Fourier transformed in order to extract the power spectrum and bispectrum. In this thesis, the following convention will be used

$$\delta(\vec{k}) = \int d^3\vec{r} \delta(\vec{r}) e^{-i\vec{k}\cdot\vec{r}} \quad (2.22)$$

and the inverse Fourier transform can be written as

$$\delta(\vec{r}) = \int \frac{d^3\vec{k}}{(2\pi)^3} \delta(\vec{k}) e^{i\vec{k}\cdot\vec{r}} \quad (2.23)$$

where  $\delta(\vec{k})$  and  $\delta(\vec{r})$  are the overdensity fields in Fourier and configuration space, respectively. The definitions of the statistical clustering quantities needed in this thesis, the power spectrum  $P(k_1)$  and the bispectrum  $B(k_1, k_2, k_3)$ , are given by

$$\langle \delta(\vec{k}_1) \delta(\vec{k}_2) \rangle \equiv (2\pi)^3 \delta_D(\vec{k}_{12}) P(k_1) \quad (2.24)$$

and

$$\langle \delta(\vec{k}_1) \delta(\vec{k}_2) \delta(\vec{k}_3) \rangle \equiv (2\pi)^3 \delta_D(\vec{k}_{123}) B(\vec{k}_1, \vec{k}_2, \vec{k}_3) \quad (2.25)$$

where  $\vec{k}_n = \vec{k}_1 + \dots + \vec{k}_n$  and  $\delta_D$  is the delta Dirac function. After the definition of the power spectrum and the bispectrum, they have to be extracted out of data sets. The following section will be focused on this topic.

## 2.5 Power spectrum and Bispectrum: Measurement

This thesis follows the description in Sefusatti (2005). First of all, the particles in the catalog must be distributed into a grid for the following **F**ast **F**ourier **T**ransform (FFT) by means of mass assignment functions (Hockney & Eastwood, 1981; Jing, 2005). The mass assignment is required for the FFT but not for the direct Fourier transform, as given in Equation (2.23), where just the actual spatial positions of the particles are needed. However, this is very time consuming and still not feasible for the huge data sets which are available by now. The cloud-in-cell (CIC) scheme (Hockney & Eastwood, 1981) is one

possible mass assignment function and was chosen for this thesis.

The notation for the following discussion of the mass assignment was adopted from Cui et al. (2008). For the extraction of the correct clustering statistics the Fast Fourier transformed density field  $\delta_{FFT}(\vec{k})$  must be corrected for the mass assignment function by

$$\delta(\vec{k}) = \frac{\delta_{FFT}(\vec{k})}{W(\vec{k})} \quad (2.26)$$

where

$$W(\vec{k}) = \prod_{i=1}^3 W(k_i) = \prod_{i=1}^3 \left( \frac{\sin\left(\frac{\pi k_i}{2k_{Ny}}\right)}{\frac{\pi k_i}{2k_{Ny}}} \right)^2 \quad (2.27)$$

and  $k_i$  is the  $i$ th component of  $\vec{k}$ ,  $k_{Ny} = \frac{\pi N_g}{L_{Box}}$  is the Nyquist mode,  $L_{Box}$  the side length of the box and  $N_g$  the number of grid cells into which the volume is divided. The information contained in modes larger than  $k_{Ny}$  cannot be reconstructed correctly anymore (an explanation for this is given by the Nyquist-Shannon sampling theorem, it says that the sampling frequency must be at least half the Nyquist frequency).

The distribution of the particles are performed in configuration space within a given volume, which is defined by the size of the grid cells. All power spectra and bispectra, which are extracted from  $N$ -body simulations, were divided into  $N_g = 512^3$  cells, e.g. for the L-BASICC simulations. This means that the side length for one cell can be evaluated as  $2.62 h^{-1}\text{Mpc}$ . The particles are distributed according to the CIC in configuration space:

$$W(\vec{x}) = \prod_{i=1}^3 W(x_i) = \prod_{i=1}^3 \begin{cases} 1 - |x_i|, & \text{if } |x_i| < 1 \\ 0, & \text{else,} \end{cases} \quad (2.28)$$

where  $x_i$  is the  $i$ th component of  $\vec{x}$  which is the distance of a particle to the center of the grid cell in which it is located. Equation (2.27) and Equation (2.28) are related with each other by the Fourier transformation. After the mass assignment in configuration space, the FFT of the density field is performed by the public available **F**astest **F**ourier **T**ransform in the **W**est (FFTW)<sup>4</sup> routines (Frigo & Johnson, 2005).

The extraction of the power spectrum and the bispectrum from a Fourier density field given by Equation (2.26) can be discussed now. However, the clustering statistics cannot be estimated just by the definition in Equation (2.24), a so-called estimator is required for this task. One possibility to estimate the power spectrum is the FKP (Feldman-Kaiser-Peacock) estimator (Feldman et al., 1994) which is defined as

$$P_{est}(k) = \frac{k_f^3}{V_P} \int_{k_1 - \Delta k/2}^{k_1 + \Delta k/2} d^3 q_1 \int_{k_2 - \Delta k/2}^{k_2 + \Delta k/2} d^3 q_2 \delta_D(\vec{q}_{12}) \delta(q_1) \delta(q_2) \quad (2.29)$$

<sup>4</sup><http://www.fftw.org/>

where  $\Delta k$  is the size of the bin,  $V_{survey}$  is the volume of the survey and  $k_f$  is the fundamental mode of the considered volume and is defined as

$$k_f = \frac{2\pi}{L_{Box}}, \quad (2.30)$$

and

$$V_P(k) = \int_{k_1-\Delta k/2}^{k_1+\Delta k/2} d^3 q_1 \int_{k_2-\Delta k/2}^{k_2+\Delta k/2} d^3 q_2 \delta_D(\vec{q}_{12}) \approx 4\pi k^2 \Delta k \quad (2.31)$$

counts the number of modes. The integrations in Equation (2.29) and Equation (2.31) are performed over the binsize. This means that the density field within the so defined  $k$ -shells with a width of  $\Delta k$  is summed up. The estimator for the bispectrum (Scoccimarro et al., 1998; Sefusatti et al., 2006) is defined in a similar manner as the power spectrum and is given by

$$\begin{aligned} B_{est}(\vec{k}_1, \vec{k}_2, \vec{k}_3) &= \frac{k_f^3}{V_B} \int_{k_1-\Delta k/2}^{k_1+\Delta k/2} d^3 q_1 \int_{k_2-\Delta k/2}^{k_2+\Delta k/2} d^3 q_2 \\ &= \int_{k_3-\Delta k/2}^{k_3+\Delta k/2} d^3 q_3 \delta_D(\vec{q}_{123}) \delta(q_1) \delta(q_2) \delta(q_3) \end{aligned} \quad (2.32)$$

where

$$V_B(k) = \int_{k_1-\Delta k/2}^{k_1+\Delta k/2} d^3 q_1 \int_{k_2-\Delta k/2}^{k_2+\Delta k/2} d^3 q_2 \int_{k_3-\Delta k/2}^{k_3+\Delta k/2} d^3 q_3 \delta_D(\vec{q}_{123}) \approx 8\pi^2 k_1 k_2 k_3 \Delta k^3, \quad (2.33)$$

which counts the modes for the bispectrum. These estimators count the modes directly and perform the integration of the density field in Fourier space. In this thesis, the delta Dirac function is rewritten as follows

$$\delta_D(\vec{k}) = \int_{q-\Delta k/2}^{q+\Delta k/2} d^3 q e^{i\vec{r}\cdot\vec{q}} \quad (2.34)$$

and is inserted in the equations for the power spectrum (see Equation (2.29)), and the bispectrum (see Equation (2.32)). After some rearrangements of these equations, they were transformed into

$$P_{est}(k) = \frac{k_f^3}{V_P} \int d^3 x I_{\vec{k}_1}(x) I_{\vec{k}_2}(x) \quad (2.35)$$

where  $\vec{k}_1 = \vec{k}_2$  (because of the delta Dirac function) and

$$B_{est}(\vec{k}_1, \vec{k}_2, \vec{k}_3) = \frac{k_f^3}{V_B} \int d^3 x I_{\vec{k}_1}(x) I_{\vec{k}_2}(x) I_{\vec{k}_3}(x) \quad (2.36)$$

where

$$I_{\vec{k}}(x) = \int_{k-\Delta k}^{k+\Delta k} \frac{d^3 q}{(2\pi)^3} e^{i\vec{x}\cdot\vec{q}} \delta(\vec{q}). \quad (2.37)$$

For every  $k$ -mode of interest a  $k$ -shell has to be built and an inverse Fourier transform has to be carried out. Afterwards, the  $k$ -shell is a configuration space quantity. Then, all of these  $k$ -shells are added up as shown in the equations above and multiplied by the prefactors in front of the integrals resulting in the statistical measure which was chosen to be extracted from the catalog.

For a correct estimation of the power spectrum or the bispectrum the shotnoise (Peebles, 1980) has to be subtracted from the results above. The shotnoise of the power spectrum is just given by

$$P_{shot} = \frac{1}{\bar{N}_p} = \frac{V_{survey}}{N_p} \quad (2.38)$$

where  $N_p$  is the number of particles within the survey volume  $V_{survey}$ . The shotnoise for the bispectrum is a little bit more complicated (Smith et al., 2008) and can be written as

$$B_{shot} = (P_{est}(k_1) + P_{est}(k_2) + P_{est}(k_3) - 3P_{shot})P_{shot} + P_{shot}^2 . \quad (2.39)$$

It is worth noting that the shotnoise for the bispectrum is dependent on the  $k$ -modes, whereas for the power spectrum it is just an overall constant value. At this point it is also clearer why not only the estimation of bispectrum but also for the power spectrum was discussed in detail here. It is required for the shotnoise correction of the bispectrum. The power spectrum can be estimated during the determination of the bispectrum. This calculation is not very time consuming, as it can easily be recognized by looking at Equation (2.35) and can be carried out on the fly during the estimation of the bispectrum.

### 2.5.1 Power spectrum and Bispectrum: Gaussian density field

Before the theoretical model will be discussed in detail, a few more interesting and important statements about the bispectrum will be given in this section. The  $\delta(\vec{r})$  contains only real numbers because the density field in configuration space  $\rho(\vec{r})$  is given by the spatial positions of the objects. However, by the Fourier transformation  $\delta(\vec{k})$  (see Equation 2.22)) becomes a complex number but with a special symmetry.  $\delta(\vec{k})$  is hermitian which means

$$\delta(-\vec{k}) = \delta^*(\vec{k}) \quad (2.40)$$

the conjugate complex operation is expressed by the superscripted "\*". This is a very important symmetry which is required for a correct extraction of the clustering statistics in Fourier space. The overdensity field  $\delta(\vec{k})$  can be expressed by its real ( $Re(\delta(\vec{k}))$ ) and imaginary ( $Im(\delta(\vec{k}))$ ) part

$$\delta(\vec{k}) = Re(\delta(\vec{k})) + i Im(\delta(\vec{k})) \quad (2.41)$$

where  $i (= \sqrt{-1})$  is the imaginary unit or by

$$\delta(\vec{k}) = A e^{i \phi} \equiv \sqrt{(Re(\delta(\vec{k})))^2 + (Im(\delta(\vec{k})))^2} e^{i \arctan \frac{Im(\delta(\vec{k}))}{Re(\delta(\vec{k}))}} \quad (2.42)$$

$$= A (\cos(\phi) + i \sin(\phi)) \quad (2.43)$$

where the amplitude is given by  $A$  and the phase by  $\phi$ .

By inserting Equation (2.43) into the definition of the power spectrum in Equation (2.24) it can be noticed that the power spectrum is just the amplitude of the overdensity at  $\delta(\vec{k})$  (the phase between the  $k$ -modes is always given by  $\pi$  due to  $\delta_D(\vec{k}_1 + \vec{k}_2)$ ). The hermitian condition (see Equation (2.40)) only requires the calculation of the squared amplitude of  $\delta(\vec{k})$ . The phase  $\phi$  is not needed for measuring the power spectrum but contains all information about higher order clustering statistics. It is possible to conserve the amplitude  $A$  of the overdensity field but change the phase  $\phi$  and of course extract the same power spectrum out of this new overdensity field.

This will be shown in the following discussion. A recipe will be given for populating the Fourier space with a Gaussian density field. Basically, the volume in form of a box is filled up in spherical shells, in a way that in each shell the same power can be found statistically. The random process is chosen to be Gaussian and the most important quantities of a Gaussian are the mean and the variance, which is given by

$$\sigma_{Gauss} = \sqrt{\left(\frac{2\pi}{k_f}\right)^3 \frac{P(k)}{2}} \quad (2.44)$$

where  $P(k)$  is the power spectrum measured from the L-BASICC dark matter  $N$ -body simulations and  $k_f$  is the fundamental mode of the considered volume.

The mean value of the real and imaginary part of the density field which is built up with this process must be zero separately to ensure that the phases are uncorrelated. Therefore, the population process in Fourier space has to be performed independently for the real and imaginary part of the density field in order to avoid an accidental introduction of any higher order correlations. For this discussion fifty boxes with a length of  $L_{Box} = 1.340 h^{-1}$  Mpc divided into  $256^3$  grid cells were chosen and artificially populated with a Gaussian density field in Fourier space in order to avoid discreteness problems.

On the left panel in Figure 2.2, the power spectrum of the Gaussian density field is shown, which is non-zero, as expected from the discussion above. The power spectrum, from which the Gaussian density field was built up, is represented by the red solid line. It was extracted from the dark matter L-BASICC simulations at  $z = 1.0$  where the phase information was produced by the growth of structure. As expected, the bispectrum fluctuates around zero for the Gaussian density field, as it can be seen on the right panel in Figure 2.2. For this discussion, the bispectrum was computed with a binsize of  $2 \times k_f$  for the configuration  $k_1 = 0.094 h \text{ Mpc}^{-1}$  and  $k_2 = 0.187 h \text{ Mpc}^{-1}$  and  $k_3$  was varied to ensure the closure of the triangle. The square symbols show the mean of the 50 realizations of the Gaussian density field with their rms. The dotted line at zero is the expectation value for the bispectrum of a Gaussian density field. The blue squares show the bispectrum from the L-BASICC dark matter simulations. This shows that if any information is contained in the phases, the odd statistics will not vanish.

It can be concluded, that the information contained in the amplitude is conserved and estimated correctly by the two-point statistics by the introduced process above. However, the phases of the density field do not contain any information due to the process of creating

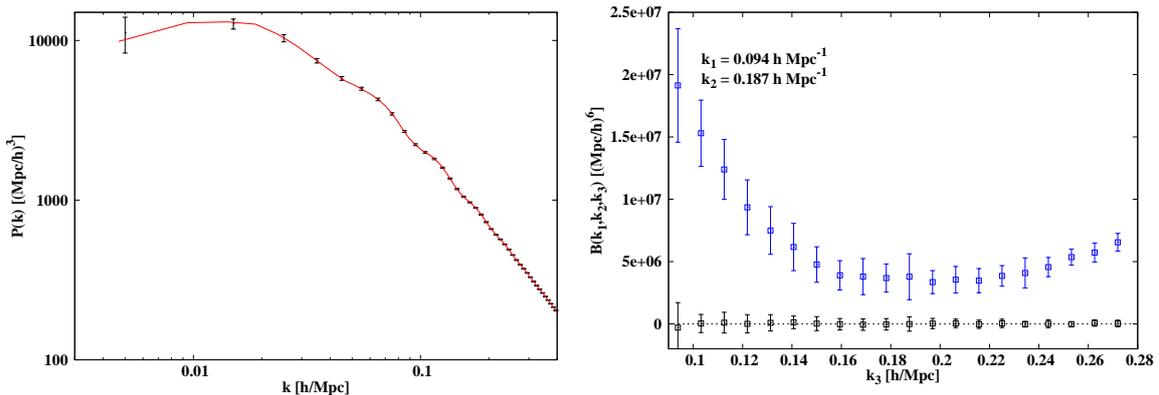


Figure 2.2: On the left panel the mean power spectrum (dots with error bars) from the 50 realizations of a Gaussian density field with their rms and the input power spectrum from the dark matter L-BASICC simulations at  $z = 1.0$  (red solid line) are plotted. The triangle configuration for the calculation of the bispectrum was set to be  $k_1 = 0.094 h \text{ Mpc}^{-1}$ ,  $k_2 = 0.187 h \text{ Mpc}^{-1}$  and varying  $k_3$  and can be seen on the right panel. The mean bispectrum from the 50 realizations of a Gaussian density with their rms is shown as the black squares with error bars. This bispectrum fluctuates around the zero line which is indicated by the black dotted line, as it is expected. The blue squares represent the bispectrum from the simulations from which the input power spectrum was extracted.

the Gaussian density field.

After having defined the clustering quantities one has to think about their theoretical modeling which will be discussed in Section 2.6.

## 2.6 Power Spectrum and Bispectrum: Model

In this section, the theoretical base for the power spectrum and the bispectrum will be derived. It will be started with the dark matter real space case in Section 2.6.1. After that derivation a discussion of biased objects, like halos or galaxies, will be given. This section will be concluded with adding peculiar velocities to obtain the redshift space power spectrum and bispectrum (Section 2.6.2).

### 2.6.1 Real space

In order to understand the meaning of the power spectrum and the bispectrum, these two clustering quantities will be derived in this section. It will be started with the linear case. The power spectrum will also be derived for the next leading order the so-called  $3^{\text{rd}}$  order perturbation theory where the density field is described in a perturbative series.

### Linear evolution of the density field

In order to compute the evolution of the power spectrum and the bispectrum, the evolution of the density fluctuations has to be known. The derivation of the growth of structure will start with the most simple case, the linear perturbation theory following Bernardeau et al. (2002), in which a detailed explanation is given.

After recombination it is assumed that matter is a pressure-less fluid without any vorticity. The scales, which are of interest for this thesis, are well inside the Hubble radius and the peculiar velocities  $\vec{v}_{pec}$  are non-relativistic so that a Newtonian treatment is valid. In the most simple case the continuity, Euler and Poisson equation are linearized and can be solved then. In comoving coordinates  $\vec{x}$  and with the conformal time  $\tau$  these equations are given by

$$\frac{\partial \delta_M}{\partial \tau} + \nabla \cdot \vec{v}_{pec} = 0 \quad (2.45)$$

for the linearized continuity equation

$$\frac{\partial \vec{v}_{pec}}{\partial \tau} + \mathcal{H}(\tau) \vec{v}_{pec} + \nabla \delta \phi = 0 \quad (2.46)$$

for the linearized Euler equation and

$$\nabla^2 \delta \phi = 4\pi G a^2 \bar{\rho} \delta_M \quad (2.47)$$

for the linearized Poisson equation where  $\delta_M \equiv (\rho - \bar{\rho}) / \bar{\rho}$  (as per definition in Equation (2.21)),  $\mathcal{H}(\tau) \equiv d \ln(a) / d\tau = H \cdot a$  and  $\delta \phi$  the gravitational potential induced by  $\delta_M$ . Taking the divergence of Equation (2.46), the conformal time derivative of Equation (2.45) and use Equation (2.47) to express  $\nabla \cdot \vec{v}_{pec}$  and  $\nabla^2 \delta \phi$  in terms of  $\delta_M$ , then one obtains

$$\frac{\partial^2 \delta_M}{\partial \tau^2} + \mathcal{H}(\tau) \frac{\partial \delta_M}{\partial \tau} = 4\pi G a^2 \bar{\rho} \delta_M \quad (2.48)$$

which describes gravitational instabilities in an expanding universe. In Fourier space, the partial differential equations are transformed to a set of independent ordinary differential equations and one gets for Equation (2.48)

$$\frac{\partial^2 \delta_M(\vec{k}, \tau)}{\partial \tau^2} + \mathcal{H}(\tau) \frac{\partial \delta_M(\vec{k}, \tau)}{\partial \tau} = 4\pi G a^2 \bar{\rho} \delta_M(\vec{k}, \tau) . \quad (2.49)$$

It can be seen that every mode is evolving independently. If higher order terms are regarded in the three basic equations, a coupling between different modes will arise. The multivariate Gaussian distribution of the density perturbations is destroyed and at later times, the power spectrum will not contain sufficient information to obtain the probability distribution function of the density field at that time. It can also be said in this way, the growth of structure, driven by gravity, leads to a mixture of modes. This means that higher point statistics are needed (e.g. bispectrum, trispectrum which is the four point statistics

in Fourier space etc.) to obtain the whole information for measuring the probability distribution function of the density field. The general solution of Equation (2.49) in linear perturbation theory is given by

$$\delta_M(\vec{k}, \tau) = D^+(a(\tau)) \tilde{A}(\vec{k}) + D^-(a(\tau)) \tilde{B}(\vec{k}) \quad (2.50)$$

where the initial density field is described by the two arbitrary functions  $\tilde{A}(\vec{k})$  and  $\tilde{B}(\vec{k})$ . Here, the matter density perturbations  $\delta_M$  are discussed for a background cosmology with dark matter and a cosmological constant. The growing mode of Equation (2.49) is called the linear growth factor

$$D^+(a) = \frac{5\Omega_M}{2} \frac{H(a)}{H_0} \int_0^a \frac{da'}{(a'H(a')/H_0)^3} . \quad (2.51)$$

If the power spectrum is given at a certain redshift it will not be any problem with Equation (2.51) to compute it for any other redshift.

The evolution of the divergence of the peculiar velocity field in the linear regime is also of interest because this information is required for a correct treatment of the redshift space distortions on large scales, which will be discussed in Section 2.6.2. The calculation is straight forward and the main part of the work was already carried out. The solution is contained in Equation (2.45) and can be derived by solving Equation (2.45) to the divergence of the peculiar velocity field and in configuration space is given by

$$\nabla \cdot \vec{v}_{pec} = -\mathcal{H}(\tau) [f(\Omega_M, \Omega_\Lambda) D^+(a) A(\vec{r}) + g(\Omega_M, \Omega_\Lambda) D^-(a) B(\vec{r})] , \quad (2.52)$$

is derived where

$$f(a) \equiv \frac{d \ln(D^+(a))}{d \ln(a)} \quad (2.53)$$

is called the dimensionless growth rate and corresponds to the growing mode whereas  $g$  describes the decaying mode, which is defined similar to Equation (2.53) but instead of  $D^+(a)$   $D^-(a)$  is used.

### Dark matter bispectrum derivation

As mentioned in Chapter 1, the second part of this thesis is mainly concentrated on the bispectrum. The focus of this thesis lies on the three-point clustering statistics generated only by gravity. It is still under debate whether there are primordial non-Gaussianities, as claimed by Rossmanith et al. (2009) and Räth et al. (2011) or not. These primordial non-Gaussianities would be created during the phase of inflation and would also influence the growth of structure. In this case, primordial higher order statistics has to be taken into account for the modeling of the bispectrum at later epochs (Sefusatti, 2009; Sefusatti et al., 2011). However, primordial non-Gaussianities and their effects on the clustering are not any subject in this thesis.

Therefore, gravity is the only effect left which is able to introduce a non-Gaussian signature to the density field. In the last section, only the linear growth of structure was discussed but in reality structure growth becomes more non-linear the smaller the considered scales are. If the growth of structure would preserve the Gaussianity of the density field, there would be no need to use any clustering statistics beyond the two-point statistics due to Wick's theorem (Bernardeau et al., 2002). However, the non-linear behavior of the growth of structure leads to a more complicated theory, too. Several attempts were made to describe this situation at least for the two-point statistics (Jain & Bertschinger, 1994; Crocce & Scoccimarro, 2006; Matsubara, 2008) where only two  $k$ -modes have to be taken into account. For higher order clustering statistics more  $k$ -modes have to be considered for their coupling terms.

This thesis follows the description of the bispectrum in the weakly non-linear regime which is used in the literature (Bernardeau et al., 2002). It is assumed that in the weakly non-linear regime the deviations from the linear structure growth can be described by a perturbative series of the density field  $\delta(\vec{x}, \tau)$  which is defined by

$$\delta(\vec{x}, \tau) = \sum_{n=1}^{\infty} \delta_n(\vec{x}) \quad (2.54)$$

and the divergence of the velocity field

$$\nabla \cdot \vec{v}_{pec}(\vec{x}, \tau) = \sum_{n=1}^{\infty} (\nabla \cdot \vec{v}_{pec,n}(\vec{x})) \quad (2.55)$$

where  $\delta_n$  and  $\nabla \cdot \vec{v}_{pec,n}$  indicate the order in the initial density field where  $n = 1$  represents the linear case. The derivation of the behavior of the velocity field will also be examined in this section because the results are required for the bispectrum in redshift space (see Section 2.6.2).

This perturbative expansion was performed by Goroff et al. (1986) for the first time. Before any further derivation or calculation will be carried out, the continuity and the Euler equations will be Fourier transformed (Bernardeau et al., 2002). The subsequent mathematical steps will be simplified enormously and after some mathematical exercises the equations of motion in Fourier space are given by

$$\begin{aligned} \frac{\partial \delta(\vec{k}, \tau)}{\partial \tau} + \nabla \cdot \vec{v}_{pec}(\vec{k}, \tau) = \\ - \int d^3 \vec{k}_1 d^3 \vec{k}_2 \delta_D(\vec{k} - \vec{k}_{12}) \alpha(\vec{k}_1, \vec{k}_2) \nabla \cdot \vec{v}_{pec}(\vec{k}_1, \tau) \delta(\vec{k}_2, \tau) \end{aligned} \quad (2.56)$$

$$\begin{aligned} \frac{\partial (\nabla \cdot \vec{v}_{pec}(\vec{k}, \tau))}{\partial \tau} + \mathcal{H}(\tau) \nabla \cdot \vec{v}_{pec}(\vec{k}, \tau) + \frac{3}{2} \Omega_M \mathcal{H}^2(\tau) \delta(\vec{k}, \tau) = \\ - \int d^3 \vec{k}_1 d^3 \vec{k}_2 \delta_D(\vec{k} - \vec{k}_{12}) \beta(\vec{k}_1, \vec{k}_2) \nabla \cdot \vec{v}_{pec}(\vec{k}_1, \tau) \nabla \cdot \vec{v}_{pec}(\vec{k}_2, \tau) \end{aligned} \quad (2.57)$$

where

$$\alpha(\vec{k}_1, \vec{k}_2) \equiv \frac{\vec{k}_{12} \cdot \vec{k}_1}{k_1^2} \quad (2.58)$$

$$\beta(\vec{k}_1, \vec{k}_2) \equiv \frac{k_{12}^2 (\vec{k}_1 \cdot \vec{k}_2)}{2k_1^2 k_2^2} \quad (2.59)$$

and  $\vec{k}_{12} \equiv \vec{k}_2 + \vec{k}_1$ . These equations can be solved by inserting them in the perturbative expansions of Equation (2.54) and (2.55). It is assumed that the time evolution and the mode-coupling are linear functions. This would result in the following series expansion for the density field

$$\delta(\vec{k}, \tau) = \sum_{n=1}^{\infty} D^+(\tau) \delta_n(\vec{k}, \tau) \quad (2.60)$$

and for the divergence of the velocity field

$$\nabla \cdot \vec{v}_{pec}(\vec{k}, \tau) = -\mathcal{H}(\tau) \sum_{n=1}^{\infty} D^+(\tau) (\nabla \cdot \vec{v}_{pec,n}(\vec{k}, \tau)) \quad (2.61)$$

where only the fastest growing mode is considered. For simplicity it is common to perform the following steps of the derivation for an Einstein-de-Sitter (EdS) universe ( $\Omega = \Omega_M = 1$ ) (Bernardeau et al., 2002). The linear growth factor  $D^+(\tau)$  is then simply given by the scale factor  $a(\tau)$ . These series expansions (Equation (2.60) and Equation (2.61)) are inserted in the equations of motion (see Equation (2.57)) and give the following results

$$\delta_n(\vec{k}) = \int d^3 \vec{q}_1 \dots \int d^3 \vec{q}_n \delta_D(\vec{k} - \vec{q}_1 + \dots + \vec{q}_n) F_n(\vec{q}_1, \dots, \vec{q}_n) \delta_1(\vec{q}_1) \dots \delta_1(\vec{q}_n) \quad (2.62)$$

$$\nabla \cdot \vec{v}_{pec,n}(\vec{k}) = \int d^3 \vec{q}_1 \dots \int d^3 \vec{q}_n \delta_D(\vec{k} - \vec{q}_1 + \dots + \vec{q}_n) G_n(\vec{q}_1, \dots, \vec{q}_n) \delta_1(\vec{q}_1) \dots \delta_1(\vec{q}_n) \quad (2.63)$$

where  $F_n(\vec{q}_1, \dots, \vec{q}_n)$  and  $G_n(\vec{q}_1, \dots, \vec{q}_n)$  are homogeneous functions of the wave vectors and are built up from combinations of  $\alpha(\vec{k}_1, \vec{k}_2)$  and  $\beta(\vec{k}_1, \vec{k}_2)$ . The derivation of the mode-coupling kernels for  $n \geq 2$  was given the first time in Goroff et al. (1986):

$$F_n(\vec{q}_1, \dots, \vec{q}_n) = \sum_{m=1}^{n-1} \frac{G_m(\vec{q}_1, \dots, \vec{q}_m)}{(2n+3)(n-1)} \left( (2n+1) \alpha(\vec{k}_1, \vec{k}_2) F_{n-m}(\vec{q}_{m+1}, \dots, \vec{q}_n) + 2\beta(\vec{k}_1, \vec{k}_2) G_{n-m}(\vec{q}_{m+1}, \dots, \vec{q}_n) \right) \quad (2.64)$$

$$G_n(\vec{q}_1, \dots, \vec{q}_n) = \sum_{m=1}^{n-1} \frac{G_m(\vec{q}_1, \dots, \vec{q}_m)}{(2n+3)(n-1)} \left( 3\alpha(\vec{k}_1, \vec{k}_2) F_{n-m}(\vec{q}_{m+1}, \dots, \vec{q}_n) + 2n\beta(\vec{k}_1, \vec{k}_2) G_{n-m}(\vec{q}_{m+1}, \dots, \vec{q}_n) \right) \quad (2.65)$$

where  $\vec{k}_1 \equiv \vec{q}_1 + \dots + \vec{q}_m$ ,  $\vec{k}_2 \equiv \vec{q}_{m+1} + \dots + \vec{q}_n$  and  $F_1 = G_1 = 1$ . In this thesis, for both kernels  $n = 2$  is required for the theoretical predictions in real and redshift space (see Section 2.6.2) and are given by

$$F_2(q_1, q_2, \theta_{12}) = \frac{5}{7} + \frac{\cos(\theta_{12})}{2} \left( \frac{q_1}{q_2} + \frac{q_2}{q_1} \right) + \frac{2}{7} \cos^2(\theta_{12}) \quad (2.66)$$

$$G_2(q_1, q_2, \theta_{12}) = \frac{3}{7} + \frac{\cos(\theta_{12})}{2} \left( \frac{q_1}{q_2} + \frac{q_2}{q_1} \right) + \frac{4}{7} \cos^2(\theta_{12}) \quad (2.67)$$

where  $\theta_{12}$  is the angle between the modes  $\vec{q}_1$  and  $\vec{q}_2$ . The solution for the linear case was already given in the last section where the linearization of the basic equations (2.45), (2.46) and (2.47) lead to the linear growth factor  $D^+(a)$ . If the growth of structure was completely linear and there was not any mode-coupling generated by gravitational forces, the power spectrum would contain all necessary information about the distribution of matter in the Universe.

The  $2^{nd}$  order perturbation theory is the simplest and first non-vanishing order for the calculation of the bispectrum, it is also often called the tree-level approach. Then the real space dark matter bispectrum is given by

$$\begin{aligned} B_{dm}(k_1, k_2, \theta_{12}) &= 2F(k_1, k_2, \theta_{12})P_L(k_1)P_L(k_2) + 2F(k_2, k_3, \theta_{23})P_L(k_2)P_L(k_3) \\ &+ 2F(k_1, k_3, \theta_{13})P_L(k_1)P_L(k_3) \\ &= 2(F(k_1, k_2, \theta_{12})P_L(k_1)P_L(k_2) + \text{cyclic permutations}) . \end{aligned} \quad (2.68)$$

where  $k_1$  and  $k_2$  are some certain modes with their corresponding linear power spectra ( $P_L(k_1)$ ,  $P_L(k_2)$ ) and the angle  $\theta_{12}$  between the two modes. The dependence on cosmological parameters of the bispectrum is provided by the power spectra in Equation (2.69) and not by the mode-coupling process itself, as long as different cosmologies can be neglected in the calculation of the kernels. This will be shown in a later discussion but first a closer look at the non-linear structure growth is required.

### $3^{rd}$ order perturbation theory: power spectrum

The  $3^{rd}$  order perturbation theory is the next order in the series expansion of Equation (2.60) and Equation (2.61). The exact solution of the non-linear structure growth cannot be computed, because an infinite number of higher order terms would be required. The non-linear power spectrum is calculated as an ensemble average of the density and velocity field, up to third order here. A general definition for the ensemble average can be expressed as

$$(D^+(a(\tau)))^n \langle \Theta_i(\vec{k}) \Theta_j(\vec{k}') \rangle \equiv (2\pi)^3 \cdot \delta_D(\vec{k} + \vec{k}') P_{ij}(k, \tau) , \quad (2.69)$$

where the  $\Theta_i(\vec{k})$ s can be identified as the density field  $\delta_n(\vec{k})$  for  $i = \delta$  and as the divergence of the velocity field  $\nabla \cdot v_{pec}(\vec{k})$  for  $i = \theta$ . This kind of notation can be found in Nishimichi et al.

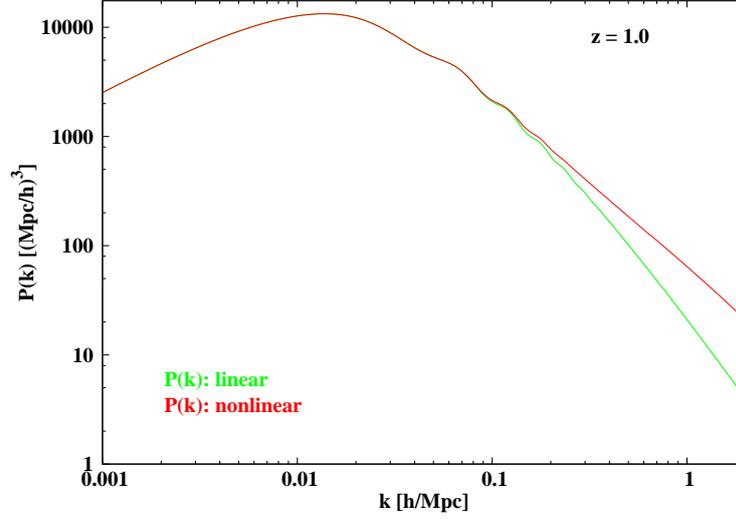


Figure 2.3:

The linear and non-linear power spectra are shown in green and red, respectively, at  $z = 1.0$ .

(2007). Up to third order terms the solution for Equation (2.69) can be written as

$$P_{ij}(k, \tau) = (D^+(a(\tau)))^2 P_L(k, \tau) + \underbrace{(D^+(a(\tau)))^4 \left( P_{ij}^{(22)}(k, \tau) + 2P_{ij}^{(13)}(k, \tau) \right)}_{\text{one loop corrections}} \quad (2.70)$$

where  $n$  and  $m$  ( $n, m = 1, 2$  and  $3$ ) specify higher order terms, which are used for the calculations of  $P_{ij}^{nm}(\vec{k}, \tau)$  and are given by

$$P_{\delta\delta}^{(22)}(k, \tau) = 2 (D^+(a(\tau)))^2 \int \frac{d^3q}{(2\pi)^3} \left( F_2(\vec{k} - \vec{q}, \vec{q}) \right)^2 P_L(|\vec{k} - \vec{q}|) P_L(q) \quad (2.71)$$

$$P_{\theta\theta}^{(22)}(k, \tau) = 2 (D^+(a(\tau)))^2 \int \frac{d^3q}{(2\pi)^3} \left( G_2(\vec{k} - \vec{q}, \vec{q}) \right)^2 P_L(|\vec{k} - \vec{q}|) P_L(q) \quad (2.72)$$

and

$$P_{\delta\theta}^{(13)}(k, \tau) = 3 \int \frac{d^3q}{(2\pi)^3} \left( F_3(\vec{k}, \vec{q}, -\vec{q}) + G_3(\vec{k}, \vec{q}, -\vec{q}) \right) P_L(k) P_L(q) \quad (2.73)$$

where  $P_{\delta\theta}^{(13)}(k, \tau) = P_{\theta\delta}^{(13)}(k, \tau)$ . A detailed discussion of the one loop corrections for  $P_{\delta\delta}(k)$  can be found in Bernardeau et al. (2002).

In Figure 2.3, the linear and non-linear power spectra are plotted as green and red solid line, respectively, at  $z = 1.0$ . The public available CAMB<sup>5</sup> routine (Lewis et al., 2000) is

<sup>5</sup><http://camb.info/>

used for the computation of the linear power spectrum. For the calculation of the  $3^{rd}$  order perturbation theory power spectrum the public available routines from Komatsu (2008)<sup>6</sup> is used. All following theoretical power spectra are calculated with these routines if not stated otherwise. At the smallest modes, the large scales in configuration space, the two power spectra cannot be distinguished from each other. At  $k \approx 0.1 h \text{ Mpc}^{-1}$  after the second wiggle the power spectra begin to differ. The non-linear power spectrum shows more signal at larger modes than the linear one. The boost of the amplitude in the non-linear power spectrum is originated from the  $3^{rd}$  order mode-coupling. The involved modes are now correlated with each other and do not evolve independently anymore, as it is the case in linear perturbation theory. The non-linear power spectrum is a good description at high redshifts (Jain & Bertschinger, 1994; Jeong & Komatsu, 2006). Therefore,  $3^{rd}$  order perturbation theory will also be used to model the non-linear growth of structure for the later investigations if not stated otherwise.

### Modifying the tree-level ansatz

In this section, it will be shown how the original tree-level ansatz, given by Equation (2.69), can be modified, in order to get a better description of the non-linear structure growth for the bispectrum. The result will be called modified tree-level ansatz.

The modification is related to the usage of the linear power spectra  $P_L(k)$  in the original tree-level ansatz given in Equation (2.69) and motivated by Guo & Jing (2009). The  $3^{rd}$  order perturbation theory power spectra (Jain & Bertschinger, 1994) will be inserted in Equation (2.69) for a better modeling of the bispectrum. It can surely be assumed that by this change in Equation (2.69) some of the non-linear features can be better dealt with as just by the F-kernel.

In Figure 2.4, the dark matter bispectrum in real space is given by the blue squares with error bars, which are the rms of the 50 realizations of the L-BASICC simulation at  $z = 1.0$ . The green solid line represents the original tree-level ansatz and in red the modified tree-level ansatz, in which the linear power spectra in Equation (2.4) are replaced by the  $3^{rd}$  order perturbation theory power spectra. The triangle configuration is set to  $k_1 = 0.094 h \text{ Mpc}^{-1}$ ,  $k_2 = 0.187 h \text{ Mpc}^{-1}$  and a varying  $k_3$ , which is determined by the angle  $\theta_{12}$  between  $k_1$  and  $k_2$  to ensure the closure of the triangle. The theoretical bispectrum, given by the original tree-level ansatz, constantly underestimates the amplitude of the L-BASICC bispectrum. The replacement of the linear power spectra in Equation (2.69) by the  $3^{rd}$  order perturbation theory power spectra is able to reproduce the amplitude more correctly. The statistical quantity  $\chi^2$  is able to discriminate between different models and combinations of their parameters (Press et al., 1996).  $\chi^2$  is defined as

$$\chi^2 = \sum_{i=1}^N \sum_{j=1}^N (y_{model}(x_i, a_1, \dots, a_M) - y_{data}(x_i)) (Covar(x_i, x_j))^{-1} (y_{model}(x_j, a_1, \dots, a_M) - y_{data}(x_j)) \quad (2.74)$$

<sup>6</sup><http://gyudon.as.utexas.edu/~komatsu/CRL/index.html>

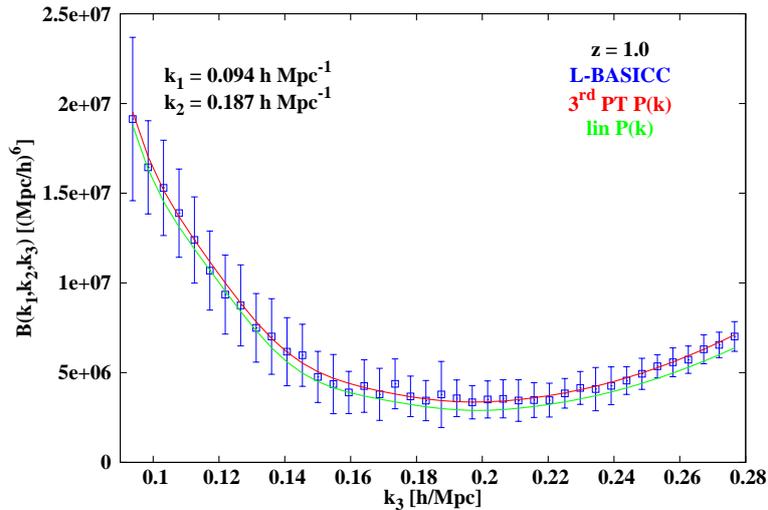


Figure 2.4: The dark matter bispectrum from the L-BASICC simulations is given by the blue squares with the error bars (the rms of the 50 realizations of the L-BASICC dark matter simulations at  $z = 1.0$ ). The original tree-level ansatz is indicated by the green solid line and modified tree-level ansatz by red. The configuration for the triangle was set to be  $k_1 = 0.094 h \text{ Mpc}^{-1}$ ,  $k_2 = 0.187 h \text{ Mpc}^{-1}$  and varying  $k_3$ .

where  $y_{\text{model}}(x_i, a_1, \dots, a_M)$  is the model which is described by  $M$  different parameters  $a_i$  with  $i \in [1, M]$  for the  $N$  measurement points  $x_i$  with  $i \in [1, N]$  and  $y_{\text{data}}(x_i)$  the corresponding measurement.

For the usage of the linear power spectra in Equation (2.69) a  $\chi^2$  of 9.05 can be found, whereas for the non-linear power spectra  $\chi^2 = 1.22$ . The model, which obtains the smaller  $\chi^2$ , describes the data in a better way than the model with the larger  $\chi^2$ . It can be concluded from this short discussion that some of the features, introduced by the non-linear structure growth, are better dealt with by the modified tree-level ansatz.

### Cosmology dependence of the F-kernel

As mentioned during the derivation of the real space dark matter bispectrum, the cosmology dependence of the kernels was not taken into account for simplicity during the derivation of the bispectrum. It was shown in Bouchet et al. (1992, 1995) that for Equation (2.66) the cosmology can be considered by replacing the  $5/7 \rightarrow (1 + \kappa)/2$  and  $2/7 \rightarrow (1 - \kappa)/2$  where  $\kappa = 3/7 \Omega_M^{-1/143}$ . A similar scheme for Equation (2.67) can be found in Bernardeau et al. (2002) where the  $3/7$  is replaced by  $\kappa$  and  $4/7$  by  $(1 - \kappa)$ . By looking at  $\kappa$  it can be noticed that the dependence on the cosmology is very weak and can almost be neglected.

In Figure 2.5, the  $F_2(k_1, k_2, \theta_{12})$  kernels for  $z = 1.0$  are shown on the left panel, calculated for an EdS universe (black line) and for the fiducial  $\Lambda$ CDM cosmology (red line). The corresponding bispectra for a configuration with  $k_1 = 0.048 h \text{ Mpc}^{-1}$  and  $k_2 = 2 \times k_1$ ,

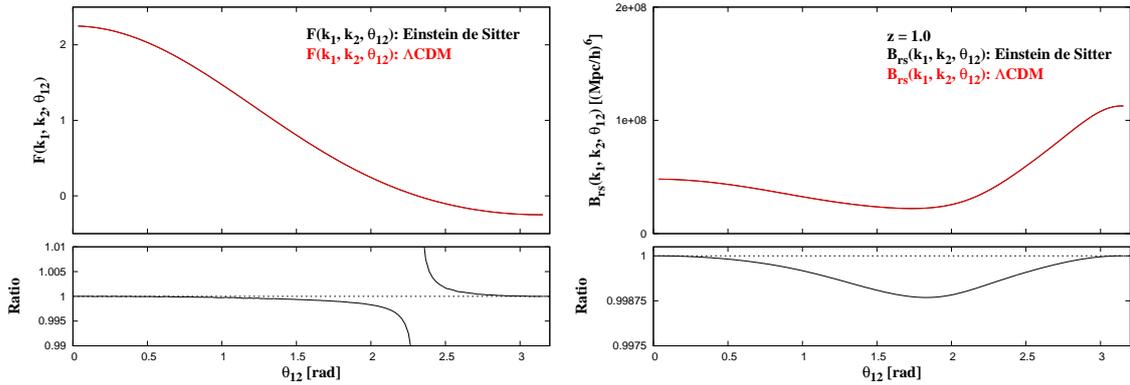


Figure 2.5: The  $F_2(k_1, k_2, \theta_{12})$  kernels calculated for an Einstein de Sitter (EdS) universe (black line) and for the fiducial  $\Lambda$ CDM cosmology (red line) at  $z = 1.0$  is given in the left panel. On the right panel the corresponding bispectra are shown with the same color coding. The configuration for the bispectra is given by  $k_1 = 0.048 h \text{ Mpc}^{-1}$  and  $k_2 = 2 \times k_1$ . Below each panel the ratio between the EdS and  $\Lambda$ CDM quantity can be seen.

with the same color coding as on the left panel are given on the right panel. Below each panel the ratio between the EdS and  $\Lambda$ CDM quantity is given. It can be noticed that for  $F_2(k_1, k_2, \theta_{12})$  the difference for small angles and for large angles is negligible. Between  $1.25 \lesssim \theta_{12} \lesssim 2.75$  the ratio between the quantities grows up to half of a percent without considering the zero crossing. Around the zero crossing, the largest deviation is visible. This behavior is usual around zero crossings where small values are divided through each other and small deviation results in a huge relative deviation. However, a behavior like that cannot be identified in the comparison plot of the bispectra because no zero crossing is present there. The difference between the bispectra is even smaller than for the kernels. It is less than 0.2 percent.

The discussion of the bispectra is required because the kernel must be calculated for every angle and then the involved  $k$ -modes are permuted as it can be seen in Equation (2.69). The different effects are added up and the net effect could lead to another dependence. From this discussion, it can be noticed that the dependence on the cosmological parameters of the  $F_2(k_1, k_2, \theta_{12})$  kernel is even less important.

However, the cosmology dependence should become more important at lower redshifts. In Figure 2.6, the same plots as in Figure 2.5 are shown but for  $z = 0.0$ . As it was expected, the lower the redshift the more important is the dependence of the kernel on the cosmological parameters. For this configuration, the kernel differs up to two percent and the bispectra differ up to 0.5 percent. Around the zero crossing the same situation is present as in Figure 2.5. This is still a very small effect and it is impossible to distinguish the two bispectra from each other. The deviation between the two bispectra can be up to 0.5 percent.

In this thesis, all kernels will be calculated for an EdS universe because of consistency reasons. The main reason is the calculation of the 3<sup>rd</sup> order perturbation theory power spectra,

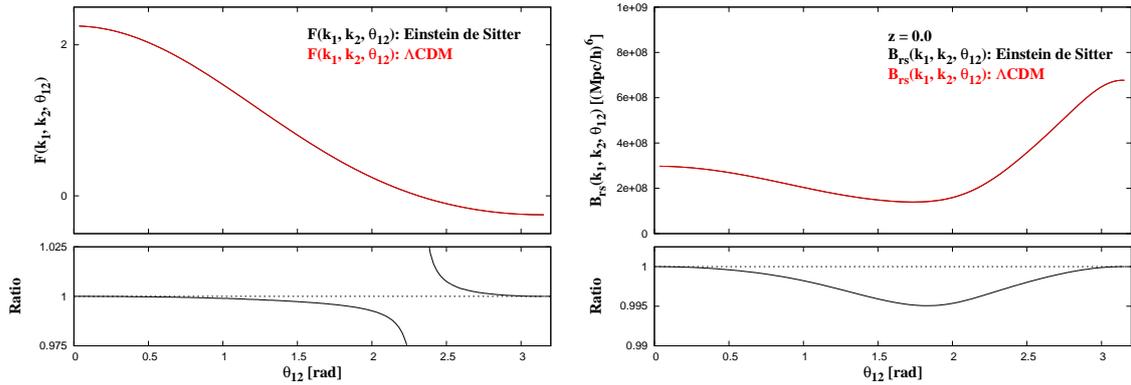


Figure 2.6: Description as in Figure 2.5, but for  $z = 0.0$ .

using the routines of Komatsu (2008), where the EdS kernels are implemented. Neglecting the cosmology dependence on the kernels is still common even for more sophisticated calculations of the non-linear clustering statistics (Nishimichi et al., 2007; Montesano et al., 2011; Sánchez et al., 2012).

So far, the derivations were carried out for dark matter particles only. In surveys, galaxies are observed and their clustering is not identical to the dark matter clustering. The effect of this, so-called bias, will be discussed in the following section.

### Biased Objects

Tracers like galaxies, and not dark matter particles, can be observed in galaxy surveys and are assumed to follow the underlying dark matter density field (Kaiser, 1984). In Chapter 1, the well known 2dFGRS (Cole et al., 2005) and the SDSS (Eisenstein et al., 2001) were introduced as examples for galaxy surveys. The linear bias, which is commonly assumed to be sufficient to describe the clustering for the two-point statistics on large scales, is defined as

$$b(r, z)^2 \equiv \frac{P_{galaxy}(k, z)}{P_M(k, z)} \equiv \frac{\langle \delta_{galaxy}^2 \rangle}{\langle \delta_M^2 \rangle} \quad (2.75)$$

where  $P_{galaxy}(k, z)$  is the power spectrum and  $\delta_{galaxy}$  the overdensity of galaxies at a certain redshift and  $k$ -mode and  $P_M(k, z)$  for the matter power spectrum and  $\delta_M$  the overdensity of matter. The clustering of the galaxies is influenced by gas physics, star formation and feedback processes, which are still not understood completely, whereas dark matter is only influenced by gravity. This will lead to a different clustering of galaxies compared to dark matter and cannot be fully described from first principles.

However, the differences can be accounted for in three increasingly more complicated different ways:

- linear bias,
- linear stochastic bias,

- and non-linear stochastic bias.

The linear bias only takes into account the linear fluctuations of the continuity and Euler equations (see Equation (2.45) and (2.46)). This means following Hui & Parfrey (2007) (originally done in Fry (1996)) that the comoving number density of dark matter and galaxies is conserved, which is expressed by the continuity equation. In other words, galaxies are not allowed to undergo mergers in the linear bias model.

Assuming that matter and galaxies are affected by gravity in the same way, then the Euler equation leads to the result that any initial velocity will decay with time and that the galaxies move with the underlying dark matter. This means no velocity bias will be present. If this statement was not true, it would not be a linear bias anymore.

In Wake et al. (2006) and Padmanabhan et al. (2008), it was shown that for **Luminous Red Galaxies** (LRGs) these statements are a good description for the bias between  $z = 0.55$  and  $z = 0.70$ , but Wake et al. (2008) shows that a merger rate of  $7.5 \pm 2.3\%$  is required to explain the clustering evolution of LRGs between  $z = 0.55$  and  $z = 0.19$ .

For the linear stochastic bias, also explored in Hui & Parfrey (2007), an additional parameter is introduced. By measuring the ensemble average of the product between the matter and galaxy density fluctuations in Fourier space or in configuration space this parameter is given by

$$b \cdot cc \equiv \frac{\langle \delta_{galaxy} \delta_M \rangle}{\langle \delta_M^2 \rangle} \quad (2.76)$$

where the cross-correlation between the mass and the galaxy distribution is given by a scale independent  $cc$ . With time, both  $cc$  and  $b$  tend towards unity but  $cc$  converges to unity much faster, see Hui & Parfrey (2007). In Swanson et al. (2008), it was shown that the cross-correlation on large scales is of order unity and thus in this thesis  $cc$  is set to unity for later considerations. In the non-linear case, a series expansion of  $\delta_{galaxy}$  in orders of  $\delta_M$  is carried out. Going to second order in the expansion it can be shown that a trend towards the linear case is present.

A similar analysis for  $3^{rd}$  order perturbation theory was carried out by McDonald (2006) and Jeong & Komatsu (2009) for the galaxy power spectrum. The assumption of a local function between the tracers and the underlying density field, which will also be used in this thesis for the three-point statistics, is the main ingredient of these works. The density fluctuations of the tracers  $\delta_{galaxy}$  can be expanded in terms of the dark matter density fluctuations  $\delta_M$  as shown in (Fry & Gaztañaga, 1993; McDonald, 2006) and is given by

$$\delta_{galaxy}(k) = \epsilon + \sum_{n=1}^{\infty} \frac{b_n \delta_M(k)^n}{n!} \quad (2.77)$$

where  $\epsilon$  is a stochastic parameter of the galaxy bias and the  $b_n$ s are called the galaxy bias parameters. One of the main tasks of this work is to determine the bias parameters  $b_1$  and  $b_2$ , also called linear and quadratic bias, respectively, by means of the bispectrum. If it was possible to observe the dark matter density field, the bias parameters would be  $b_1 = 1.0$  and  $b_2 = 0.0$  in this case.

Due to the effects introduced above the visible tracers have a different clustering. Therefore, Equation (2.69) must take this into account for the determination of the bias parameters. In Scoccimarro et al. (1998) and Bernardeau et al. (2002), the bispectrum for biased objects is given by

$$B_{halo/galaxy}(k_1, k_2, \theta_{12}) = b_1^3 B_{dm}(k_1, k_2, \theta_{12}) + b_2 b_1^2 (P_{dm}(k_1)P_{dm}(k_2) + 2 \text{ perm.}) \quad (2.78)$$

where  $P_{dm}(k_i)$  is the corresponding dark matter power spectrum for the mode  $k_i$ . In Fry & Gaztañaga (1993), McDonald (2006) and (Jeong & Komatsu, 2009) the power spectrum is utilized for the determination of the bias parameters, instead of the bispectrum. The extracted bias parameters from the two-point statistics are very degenerated with each other and reliable estimation of quadratic bias is not possible. In this thesis, it is still not clear so far, why the bispectrum is better suited for this task. In the following section, this will be discussed in detail.

### *Power spectrum vs. bispectrum*

In this section, it will be shown why the bispectrum is better suited to determine the bias parameters. It is already known that the cosmological parameters can be determined by the two-point statistics (Sánchez et al., 2006; Montesano et al., 2011) with a high precision. However, the two-point statistics suffers from the degeneracy between  $\sigma_8$  and the linear bias  $b_1$ . The three-point statistic is able to break this degeneracy. This is illustrated in Figure 2.7, where the two- and three-point statistics at  $z = 0.0$  are shown on the left and right panel, respectively.

The calculation was performed in real space, where for this discussion the growth of structure was assumed to follow linear perturbation theory for the power spectrum and  $2^{nd}$  order

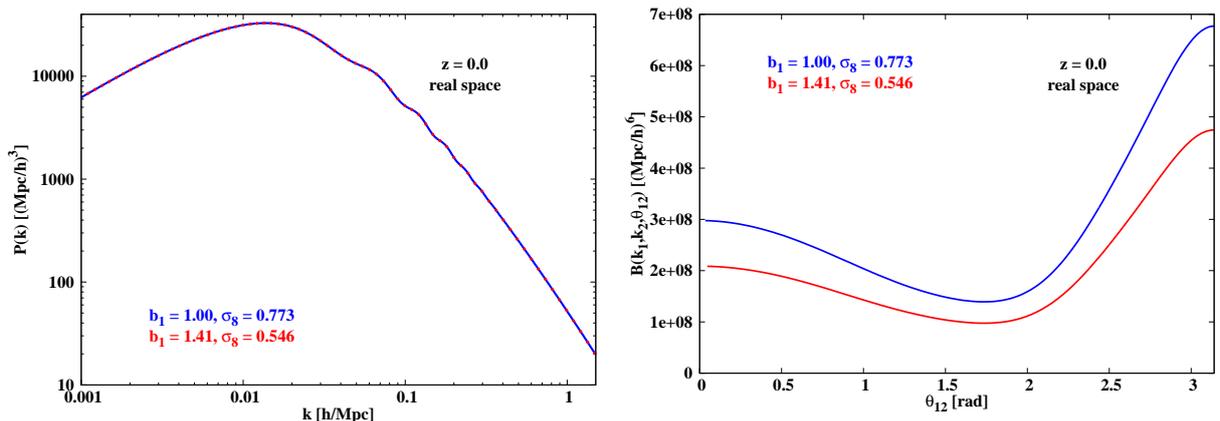


Figure 2.7: The power spectrum (left panel) given by different values for  $b_1$  and  $\sigma_8$  in such a way that the amplitude remains unchanged. The combinations of  $(b_1, \sigma_8)$  are set to (1.00, 0.773) and (1.41, 0.546) and are plotted in blue and red, respectively. The bispectrum (right panel) with  $k_1 = 0.048 h \text{ Mpc}^{-1}$  and  $k_2 = 2 \times k_1$  which is calculated for the same set of parameters is able to distinguish between the two the cases.

for the bispectrum with  $(b_1, \sigma_8) = (1.000, 0.773)$  (solid blue line) and  $(b_1, \sigma_8) = (1.41, 0.546)$  (red dots for the power spectrum and red solid line for the bispectrum). The parameters  $(b_1, \sigma_8)$  were chosen in such a way that the amplitude of their power spectra are the same. However, the amplitude of the bispectrum, right panel of Figure 2.7, is dependent on the given  $(b_1, \sigma_8)$  parameter combination. Therefore, the two cases can be distinguished. The triangle configuration for the computation of the bispectrum is chosen to be  $k_1 = 0.048 h \text{ Mpc}^{-1}$  and  $k_2 = 2 \times k_1$ .

However, one important point must be noted here, the quadratic bias  $b_2$ , see definition in Equation (2.77), was set to zero. If it is allowed to vary  $b_2$  in a way that the amplitude of the two bispectra almost coincide, then even higher order statistics have to be taken into account. However, an exact overlap will not be possible because  $b_2$  introduces a different shape dependence than the F-kernel term in Equation (2.78). In a later discussion in this section, this will be investigated in detail. But as seen in McDonald (2006) and Jeong & Komatsu (2009) the power spectrum is also influenced by the quadratic bias. This requires very detailed knowledge of the non-linear structure growth. More fine tuning would be required in order to match both amplitudes for the power spectra and the bispectra at the same time.

#### *Discussion of the bias parameters $b_1$ and $b_2$*

The effect of the linear bias  $b_1$  and the quadratic bias  $b_2$  will be discussed in this section. As described in Equation (2.77),  $b_1$  is defined as a constant multiplicative factor between the dark matter density field  $\delta_M(k)$  and the galaxy density field  $\delta_{galaxy}(k)$ . But that simple linear relation is not enough to describe the clustering of galaxies. The quadratic bias  $b_2$  is the next step in the series expansion of Equation (2.77) and relates  $\delta_M^2(k)$  with  $\delta_{galaxy}^2(k)$ . As mentioned in the previous section, the bispectrum is a suitable quantity to determine  $b_1$  and  $b_2$ .

In massive structures  $b_1$  is assumed to be higher than in less massive halos due to the high galaxy density. This is also true for  $b_2$  (Mo et al., 1997). The more time gravity is able to act on the baryonic as well as on the dark matter, the better they trace each other but never perfectly. Because on small scales, gas physics become the non-vanishing driving force, in addition to gravity, see discussion for biased objects. Therefore, it is very likely that bias parameters will always be needed to describe the difference between the galaxy clustering and the underlying matter density field.

In Figure 2.8, the effect of  $b_2$  is shown where the bias parameters were set to be  $(b_1, b_2) = (1.0, 1.0)$ . This set of parameters was chosen, because the dark matter bispectrum  $B_{rs, dm}(k_1, k_2, k_3)$  (red line in Figure 2.8) which is given by the first term of Equation (2.78) and the  $b_2 b_1^2$ -term (green line) are of comparable size. The only difference between the two terms is located in the kernel, see Equation (2.66), which is present in the first term but not in the second.

As shown in the Figures 2.5 and 2.6, where only the dark matter bispectrum was examined, the U-shape is produced by the F-kernel. In Figure 2.8, two different methods for presenting the bispectrum are given.

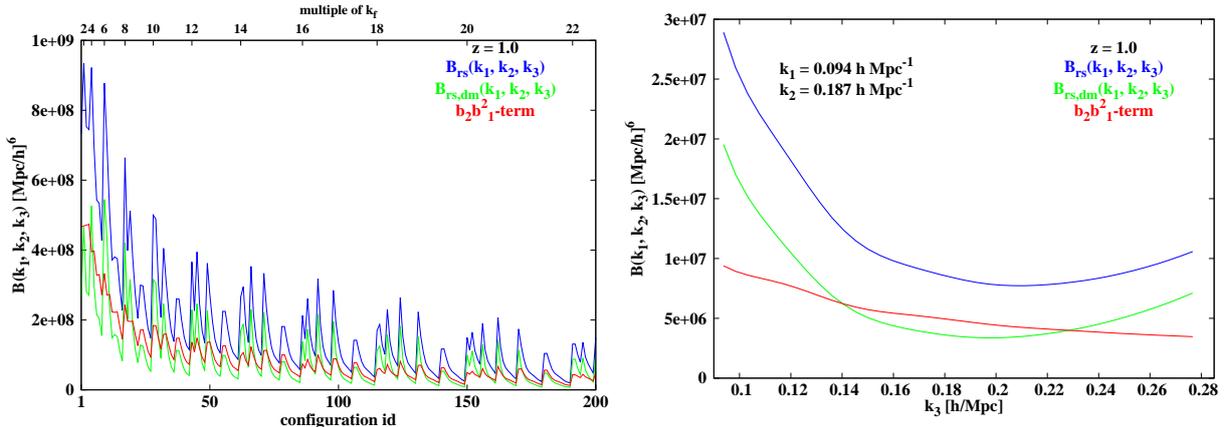


Figure 2.8: This figure shows the effect of the  $b_2 b_1^2$ -term in the bispectrum.  $B_{rs}(k_1, k_2, k_3)$  is given in blue for  $b_1 = 1.0$  and  $b_2 = 1.0$ , the dark matter  $B_{rs,dm}(k_1, k_2, k_3)$  in green ( $b_1 = 1.0$ ) and the  $b_2 b_1^2$ -term in red.

On the left panel the bispectrum is calculated for every configuration and sorted by the size of the side length of the triangle starting with largest  $k$ -mode,  $k_1 \leq k_2 \leq k_3$ . The lower x-axis is denoted *configuration id*, which defines a certain configuration, where the first configuration represents the largest scale. In Figure 2.8, the first configuration (*configuration id* one) is given by an equilateral triangle with a side length of  $k_{min} = 0.009 h \text{ Mpc}^{-1} = 2k_f$ , the side length of the following configurations are multiples of  $2k_f$ . The boxsize is set to be  $1340 h^{-1} \text{ Mpc}$ . The upper x-axis shows the multiples of  $k_f$  in relation to the *configuration id*, the tics mark the equilateral case. Below each tic, which also defines the maximum side length of the triangle up to that *configuration id*, all possible triangle configurations are considered. The larger the *configuration id*, the smaller are the considered scales (larger  $k$ -modes) of the triangle.

The right panel shows the commonly used U-shape representation. At the x-axis,  $k_3$  can be found. In the beginning of the calculation, two sides of the triangle are pre-defined (for Figure 2.8,  $k_1 = 0.094 h \text{ Mpc}^{-1}$ ,  $k_2 = 0.187 h \text{ Mpc}^{-1}$ ) and the third side length is varied for all possible closed triangle configurations. This can also be translated in an angle between the two pre-defined side lengths  $\theta_{12}$ , which then ranges from zero to  $\pi$ .

In both panels of Figure 2.8, the different shape dependencies of the two terms can be identified easily. The  $b_2 b_1^2$ -term is a declining function, starting from small angles ( $\theta \approx 0$ ), which corresponds to small  $k$ -modes and therefore large values for the power spectrum. The triangle in this situation is often called collapsed, where  $k_3 \approx k_1 - k_2$ . As  $\theta_{12}$  gets larger, the corresponding  $k_3$  gets larger and the values for the power spectrum become smaller. It can be stated, that the  $b_2 b_1^2$ -term is a declining function, when going from small to large angles. This can be noticed in the right panel of Figure 2.8. In the left panel it can also be noticed that for the two terms of Equations (2.78) two different shape dependencies are present. The estimation of  $b_1$  and  $b_2$  by means of the bispectrum is based on this difference, otherwise the estimation of  $b_1$  and  $b_2$  would be highly degenerated.

### 2.6.2 Redshift space

So far, the discussion was carried out under the assumption that the exact spatial positions of the particles or galaxies/halos are known. The only information an astronomer on earth can observe is the light from such an object from which the redshift of that object can be derived. The real redshift of the object  $z_{Hubble}$  is given by the Hubble flow. However, the gravitational potential around an object results in a net force, if matter is not distributed homogeneously. The object falls along a trajectory, which is given by the Einstein field equation (see Equation (2.1)), with the so-called peculiar velocity. Therefore, the redshift of an object is distorted by these peculiar velocities. This effect is also called redshift space distortions. The observed redshift is then given by

$$z_{obs} = z_{Hubble} + \frac{\vec{v}_{pec} \cdot \hat{x}_{\parallel}}{c} . \quad (2.79)$$

where only the line-of-sight component of these peculiar velocities ( $\vec{v}_{pec} \cdot \hat{x}_{\parallel}$ ) is required for measuring redshifts. From an observational point of view, it is impossible to distinguish between these two components. Via the redshift-distance relation (see Equation (2.4)) the observed redshift can be transferred into a distance and the two components can theoretically (but never in real life) be calculated as

$$\vec{s}(\vec{x}) = \vec{x} + \frac{\vec{v}_{pec} \cdot \hat{x}_{\parallel}}{H(z)} \quad (2.80)$$

where  $\vec{x}$  is the real spatial position of the object and the last term of Equation (2.80) is the distance, introduced by the line-of-sight component of the peculiar velocity. The exact positions of the objects are not known and cannot be recovered anymore. Therefore, a translation back to real space is not possible. For a correct data analysis, it is required to work with these redshift space distortions and derive their impact on the clustering statistics (e.g. power spectrum and bispectrum). There are two categories of redshift space distortions:

- coherent infall, motion of objects from underdense towards overdense regions  $\longrightarrow$  Kaiser effect (Kaiser, 1987)
- random motion within a cluster  $\longrightarrow$  Fingers of God (FoG) (Jackson, 1972)

The Kaiser effect (Kaiser, 1987) is a signature of the ongoing clustering within the Universe. Large structures are merging and have a motion towards each other. Depending on their position, before or behind the center of mass with respect to the observer's point of view, these objects appear red- or blue- shifted, respectively, if just compared to their Hubble flow. The observed structure looks more clustered than it really is. The signal of any clustering statistics is enhanced by this effect. In redshift space, the structure looks squashed compared to real space.

The second contribution is dominant on small scales, e.g. within a cluster. The FoGs

are originated by the random motions of objects within the cluster potential. They can have very high peculiar velocities which have a strong effect on small scales. The observed structure looks elongated or cigar-like (Jackson, 1972) and the clustering signal appears to be reduced by the FoG.

After the definition of the redshift space, it is important to evaluate its effects on the clustering statistics. For simplicity the power spectrum and the bispectrum in redshift space will be derived already for biased objects. The dark matter quantities can be recovered by simply setting  $b_1 = 1.0$  and  $b_2 = 0.0$ . First, the concept of the derivation will be given for the power spectrum and secondly, will be extended to obtain the redshift space bispectrum.

### Power spectrum in redshift space including bias

In this section, the redshift space power spectrum will be derived. A few main assumptions are required for this derivation as given in Kaiser (1987) and are listed in the following:

- the main part of the distortions is caused by large scale fluctuations,
- the large scale fluctuations are well described by linear perturbation theory (see Section 2.6.1),
- the large scale fluctuations are fully enclosed by the survey volume,
- the plane parallel approximation is valid (the displacements produced by the coherent infall are effectively parallel) and
- stationary reference frame to the CMB.

These assumptions are the main ingredients for the later derivation of the redshift space distortions. It can also be noticed from these assumptions, that the redshift space distortions mainly affect the line-of-sight direction. The more distant the object of interest is the less important is the effect transverse to the line-of-sight. This effect is shown in Hamilton (1993) where the structure looks like a banana if the distant observer approximation is not valid anymore.

In this thesis, this is not a problem because the examined simulations are boxes where the directions parallel and transverse to the line-of-sight are defined by the axes of the L-BASICC simulations. For example, the z-axis defines the direction parallel to the line-of-sight and the plane spanned by the x- and y-axis represents the direction transverse to the line-of-sight.

In this section, mainly the clustering on large scales is of interest and therefore the following derivation, which is fully given in Kaiser (1987), is focused on the so-called linear Kaiser effect. The extension to the non-linear Kaiser effect (Scoccimarro, 2004) is mathematically more complicated but the basic concept is the same. The result for the non-linear Kaiser effect will be given at the end of this section but not its derivation. The discussion here follows Dodelson (2003).

For the derivation, of the linear redshift space distortions it is assumed that the number

of galaxies within a given volume found in real space is the same as in redshift space. Of course, this assumption can be violated in reality for example magnitude limited data samples. If the change of the coordinate system does not alter the number of galaxies, then

$$n_s(\vec{s}) d^3 s = n(\vec{x}) d^3 x \quad (2.81)$$

where  $n(\vec{x}) = \bar{n}(1 + \delta)$  and  $n_s(\vec{s}) = \bar{n}(1 + \delta_s)$  are the galaxy number densities in real and redshift space, respectively, which can be calculated from the average number density  $\bar{n}$ . The change of the coordinate system is described by the Jacobian  $J$

$$J \equiv \left| \frac{d^3 x}{d^3 \vec{s}} \right| = \frac{dx}{ds} \frac{x^2}{s^2}. \quad (2.82)$$

The Jacobian can be evaluated by inserting Equation (2.80) in Equation (2.82) and after some mathematical exercises this results in

$$J = \left( 1 + \frac{\partial}{\partial x} \left( \frac{\vec{v}_{pec} \cdot \hat{x}_{\parallel}}{H(z)} \right) \right)^{-1} \left( 1 + \frac{\vec{v}_{pec} \cdot \hat{x}_{\parallel}}{H(z)x} \right)^{-2}. \quad (2.83)$$

For the computation of the galaxy clustering statistics only modes which fulfill the condition  $kx_{Survey} \gg 1$  are of interest. These modes are the best determined ones. Therefore, observers are mostly interested in these modes. Then, the first term of Equation (2.83) is the dominant part. A more detailed discussion on this topic can be found in Dodelson (2003). An expansion about  $v_{pec} = 0$  reduces Equation (2.83) to

$$J \simeq \left( 1 - \frac{\partial}{\partial x} \left( \frac{\vec{v}_{pec} \cdot \hat{x}_{\parallel}}{H(z)} \right) \right). \quad (2.84)$$

This result can be inserted in Equation (2.81) and can be rewritten as follows

$$1 + \delta_s = (1 + \delta) \left( 1 - \frac{\partial}{\partial x} \left( \frac{\vec{v}_{pec} \cdot \hat{x}_{\parallel}}{H(z)} \right) \right). \quad (2.85)$$

Afterwards, Equation (2.85) is expanded to first order and results in

$$\delta_s = \delta - \frac{\partial}{\partial x} \left( \frac{\vec{v}_{pec} \cdot \hat{x}_{\parallel}}{H(z)} \right). \quad (2.86)$$

Equation (2.86) shows, that the redshift space density field  $\delta_s$  is given by the real space density field  $\delta$  but corrected by a multiplicative factor, originated from the peculiar velocities of the galaxies.

In the remaining derivation, the distant observer approximation will be made in which it is assumed that the direction between two galaxies varies only negligibly. Therefore,  $\vec{v}_{pec} \cdot \hat{x}_{\parallel}$

in Equation (2.86) can be replaced by  $\vec{v}_{pec} \cdot \hat{z}$  and  $\hat{z}$  defines the radial direction to the center of the galaxy of interest.

The clustering quantities, in which this thesis is mainly interested in, are extracted in Fourier space. The result in Equation (2.86) needs to be Fourier transformed and after some mathematical exercises is given by

$$\delta_s(\vec{k}) = \delta(\vec{k}) + \int \frac{d^3k'}{(2\pi)^3} \delta(\vec{k}) \left( f(k' \cdot \hat{z})^2 \right) \int d^3x e^{i(\vec{k}' - \vec{k})\vec{x}} . \quad (2.87)$$

The last integral of Equation (2.87) can be identified as the Dirac delta function and simplifies the first integral to just the functional value of  $\vec{k}$ . Introducing  $\mu_{\vec{k}} \equiv k' \cdot \hat{z}$  and considering the above results one gets

$$\delta_s(\vec{k}) = \delta(\vec{k}) \left( 1 + f(z) \mu_{\vec{k}}^2 \right) \quad (2.88)$$

where the function  $f(z)$  (see also Equation 2.53) can be written as

$$f(z) = \left( \frac{\Omega_M \cdot (1+z)^3}{\Omega_M (1+z)^3 + \Omega_\Lambda (1+z)^{3 \cdot (1+w_{DE})}} \right)^\gamma . \quad (2.89)$$

To allow for the dark energy to be different from a cosmological constant (in which case  $\gamma = 0.55$ ),  $\gamma = 0.55 + 0.05 \cdot (1 + w_{DE}(z))$  for  $w_{DE}(z) > -1$  and  $\gamma = 0.55 + 0.02 \cdot (1 + w_{DE}(z))$  for  $w_{DE}(z) < -1$  (Linder, 2007) was used in all calculations in this thesis. The derivation of the redshift space power spectrum is the aim of this section. From the discussion in Section 2.4 it is known, that the power spectrum is a spherically averaged quantity, in redshift space it is simply given by

$$P_{zs}(k) = P(k) \left( 1 + \frac{2}{3} \beta(z) + \frac{1}{5} \beta^2(z) \right) \quad (2.90)$$

where  $\beta(z)$  is

$$\beta(z) = \frac{f(z)}{b} . \quad (2.91)$$

The dark matter case can be recovered by setting  $b$  to unity. In Scoccimarro (2004) also the non-linear Kaiser effect is derived in detail. This will not be fully discussed here, the concept is similar to the derivation above. The main difference is an expansion of the velocity dispersion  $\sigma_{12}^2$  of real space about redshift space, as it is called in Scoccimarro (2004). On large scales, the linear limit was used for the pairwise velocity, and the non-Gaussian terms, which are a result of the exact derivation, can be neglected. The spherically averaged power spectrum in redshift space can then be written as

$$P_{zs}(k) = b^2 \left( P_{\delta\delta}(k) + \frac{2}{3} \beta(z) P_{\delta\theta}(k) + \frac{1}{5} \beta^2(z) P_{\theta\theta}(k) \right) . \quad (2.92)$$

So far, the small scale redshift space distortions, the FoG, were still not considered in this discussion. They can easily be included in the derived framework above. In this thesis, the description of Smith et al. (2008) will be used for modeling the FoG. Then, the redshift space power spectrum is given by

$$P_s(k, \mu) = \underbrace{e^{-(f(z)\sigma_v k \mu)^2}}_{\text{FoG}} \underbrace{(b^2 P_{\delta\delta}(k) + 2b_1 f(z) \mu^2 P_{\delta\theta}(k) + f^2(z) \mu^4(z) P_{\theta\theta}(k))}_{\text{non-linear Kaiser effect}} \quad (2.93)$$

where  $\sigma_v$  is the velocity dispersion of an object. The spherically averaged redshift space power spectrum  $P_s(k)$  is given by

$$P_{zs}(k) = \frac{1}{4\pi} \int_{-1}^1 \mu \, d\mu \int_0^{2\pi} d\phi P_{zs}(k, \mu) . \quad (2.94)$$

In the presence of the FoGs, this equation cannot be solved analytically anymore as in the cases above. The calculation has to be performed numerically.

### Bispectrum in redshift space

In this section, the redshift space bispectrum will be derived and follows the work of Heavens et al. (1998). The linear Kaiser effect was expanded to describe the three-point statistic in Fourier space. The redshift space bispectrum is then given by

$$B_{zs}(k_1, k_2, \theta_{12}, \mu_1, \mu_2, \mu_3) = F_{zs}(k_1, k_2, \theta_{12}, \mu_1, \mu_2, \mu_3) P(k_1) P(k_2) + 2 \text{ perm.} \quad (2.95)$$

where  $F_{zs}(k_1, k_2, \theta_{12}, \mu_1, \mu_2, \mu_3)$  is the mode-coupling term in redshift space and has the following form

$$\begin{aligned} F_{zs}(k_1, k_2, \theta_{12}, \mu_1, \mu_2, \mu_3) = & b_1^3 (1 + \beta(z) \mu_1^2) (1 + \beta(z) \mu_2^2) (F_2(k_1, k_2, \theta_{12}) + \beta(z) \mu_3^2 G_2(k_1, k_2, \theta_{12})) \\ & + b_1 \beta(z)^2 (\mu_1^2 \mu_2^2 + \frac{\mu_1 \mu_2}{2} (\mu_1^2 \frac{k_1}{k_2} + \mu_2^2 \frac{k_2}{k_1})) \\ & + \frac{b_1 \beta(z)}{2} (\mu_1^2 + \mu_2^2 + \mu_1 \mu_2 (\frac{k_1}{k_2} + \frac{k_2}{k_1})) + \frac{b_2}{2b_1^2} . \end{aligned} \quad (2.96)$$

The  $\mu_i$ -variables are defined as the cosine between  $k_i$  and the line-of-sight direction which is assumed to be parallel to the z-axis. The first and second expressions are the already derived mode-coupling terms of the density (see Equation (2.66)) and the divergence of the velocity fields (see Equation (2.67)).

So far, only the large scale redshift space distortions for the bispectrum were considered. The second contribution to the redshift space distortions originates from the small scale random motions of objects within a gravitationally bound system, the FoG effect (Jackson, 1972). Then, the bispectrum is given by

$$B_{zs}(k_1, k_2, \theta_{12}, \mu_1, \mu_2, \mu_3) = \frac{F_{zs}(k_1, k_2, \theta_{12}, \mu_1, \mu_2, \mu_3) P(k_1) P(k_2) + 2 \text{ perm.}}{\exp\left(\frac{1}{2}(f(z)\sigma_v)^2 [(\mu_1 k_1)^2 + (\mu_2 k_2)^2 + (\mu_3 k_3)^2]\right)} . \quad (2.97)$$

In Equation (2.97), the orientation of the triangles has to be taken into account. In Equation (2.78), the integration over all possible orientations of a given triangle configuration was performed analytically. The same integration will be carried out in redshift space as well. Under the assumptions of homogeneity and isotropy (which are not strictly given in redshift space) an integration over all possible orientations result in the spherical averaged bispectrum in redshift space.

The redshift space bispectrum is dependent on six variables as it can be seen in  $B_{zs}(k_1, k_2, \theta_{12}, \mu_1, \mu_2, \mu_3)$ . But it has to be pointed out that the  $\mu_i$ -variables are not independent of each other, because the triangle must be closed for the calculation of the bispectrum. Therefore,  $B_{zs}(k_1, k_2, \theta_{12}, \mu_1, \mu_2, \mu_3)$  is only dependent on five variables, where the orientation of  $B_{zs}(k_1, k_2, \theta_{12}, \mu_1, \mu_2, \mu_3)$  is described by the  $\mu_i$ -variables which will be expressed in spherical coordinates as suggested in Smith et al. (2008). The spherical averaged bispectrum will then be given by

$$B_{zs}(k_1, k_2, \theta_{12}) = \frac{1}{4\pi} \int_0^\pi \sin(\gamma_2) d\gamma_2 \int_0^{2\pi} d\gamma_1 B_{zs}(k_1, k_2, \theta_{12}, \mu_1, \mu_2, \mu_3) \quad (2.98)$$

where the  $\mu_i$ -expressions can be calculated in following way

$$\mu_1 = \cos(\gamma_2) \quad (2.99)$$

$$\mu_2 = \sqrt{1 - \cos^2(\gamma_2)} \sin(\gamma_1) \sin(\theta_{12}) + \cos(\gamma_2) \cos(\theta_{12}) \quad (2.100)$$

$$\mu_3 = -\frac{\mu_1 k_1}{k_3} - \frac{\mu_2 k_2}{k_3}. \quad (2.101)$$

Again Equation (2.98) is the result for biased objects as it was assumed throughout the whole section. If the bispectrum should be evaluated for dark matter, the bias parameters should be set to  $b_1 = 1.0$ , like for the two-point statistics, and  $b_2 = 0.0$ .

Due to the inclusion of the FoG an analytical expression for Equation (2.98) cannot be found, the integration has to be performed numerically. This will be very time consuming. The whole procedure can be sped up if the FoG are neglected and only the the coherent infall is taken into account. Then, the angle averaging integrals can be carried out analytically. Due to the complexity of this solution it was moved to the Appendix A.1.

In the following section, a similar discussion as for real space will be given on the cosmology dependence of the F-kernels, this time in redshift space.

### Cosmology dependence of the F-kernels in redshift space

The cosmology dependence of the F-kernel will be examined again but this time in redshift space. On the left panel in Figure 2.9 the angle averaged redshift space kernels  $F_{zs}(k_1, k_2, \theta_{12})$  for dark matter at  $z = 1.0$  are plotted and on the right panel the corresponding bispectra. The EdS case is indicated by the black solid line and the  $\Lambda$ CDM case by the red solid line. Below each panel the ratios between the quantities were evaluated like in real space.

It can be noticed that the real and redshift space ratios for both, the kernels and the

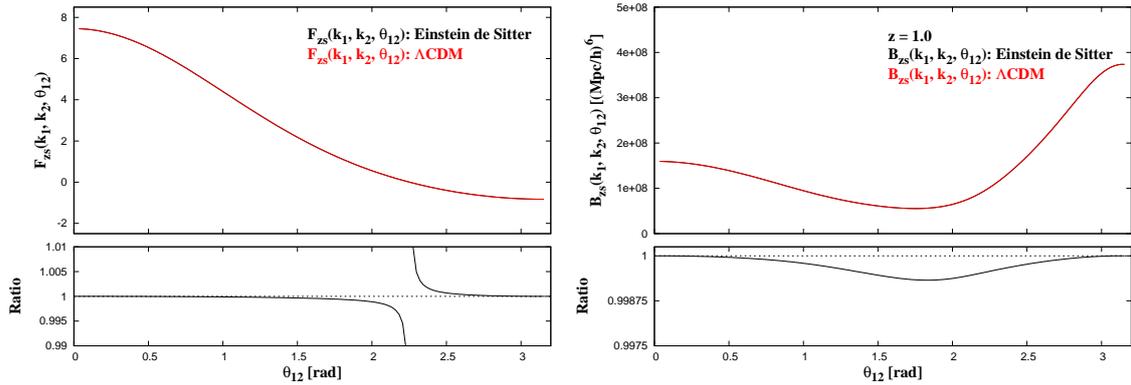


Figure 2.9: The angle averaged  $F_{zs}(k_1, k_2, \theta_{12})$  kernels calculated for an EdS universe (black line) and for the fiducial  $\Lambda$ CDM cosmology (red line) at  $z = 1.0$  are plotted on the left panel. On the right panel the corresponding bispectra are shown with the same color coding. The configuration for the bispectra is chosen to be  $k_1 = 0.048 h \text{ Mpc}^{-1}$  and  $k_2 = 2 \times k_1$ . Below each panel the ratio between the EdS and  $\Lambda$ CDM quantity is shown.

bispectra, have a very similar behavior. The redshift space bispectrum is less affected by the cosmological dependence of the kernels as in real space. The difference between the bispectra is around 0.1 percent in the worst case and the kernels differ up to 0.5 percent without considering the zero-crossing. Due to the similarity of the ratios in real and redshift space, the  $z = 0.0$  case was not examined. But it can be concluded that at  $z = 0.0$  the dependence on cosmological parameters is also negligible in redshift space.

### 2.6.3 Reduced bispectrum

In the literature also another quantity, the so-called reduced  $B(k_1, k_2, \theta)$  or  $Q(k_1, k_2, \theta)$ , is used to determine the bias parameters  $b_1$  and  $b_2$ .  $Q(k_1, k_2, \theta)$  is defined as

$$Q(k_1, k_2, \theta) \equiv \frac{B(k_1, k_2, \theta)}{P(k_1)P(k_2) + P(k_1)P(k_3) + P(k_2)P(k_3)} \quad (2.102)$$

and cannot be measured only for biased objects but also for dark matter. If the dark matter reduced bispectrum,  $Q_{dm}(k_1, k_2, \theta)$ , is known, the reduced bispectrum for biased objects,  $Q_{halo/galaxy}(k_1, k_2, \theta)$ , can be written as

$$Q_{halo/galaxy}(k_1, k_2, \theta) \equiv \frac{1}{b_1} \left( Q_{dm}(k_1, k_2, \theta) + \frac{b_2}{b_1} \right). \quad (2.103)$$

Due to its definition, the reduced bispectrum is almost independent on cosmological parameters, time evolution and redshift space distortions at lowest order, the tree-level ansatz. Because of the division by the corresponding power spectra only the shape dependence of the considered triangle configuration remains. This is the reason why this representation of the three-point statistic in Fourier space is preferred in the literature.

However, the predictive power of  $Q_{halo/galaxy}(k_1, k_2, \theta)$  is reduced by the division with a large the  $b_1$  parameter. Usually objects with a large  $b_1$ -value are also rare and as a consequence the signal-to-noise ratio is smaller compared to objects with a smaller  $b_1$ -value. This can be explained by the first term which reduces the effect of a possible shape dependence of a certain triangle configuration. The last term of Equation (2.103) is not able to change the shape of the reduced bispectrum. It just adds a constant signal at all modes to the reduced bispectrum. It is expected that  $Q_{halo/galaxy}(k_1, k_2, \theta)$  is more comparable to a flat straight line the larger  $b_1$  is because possible features are flattened out. In Section 3.2.3, an analysis to estimate  $b_1$  and  $b_2$  will be performed on  $Q_{halo/galaxy}(k_1, k_2, \theta)$  and the results will be compared to the analysis carried out on  $B(k_1, k_2, k_3)$ .

In Section 2.5, also the power spectrum was discussed. One reason was the shotnoise correction of the bispectrum. The evaluation of the reduced bispectrum delivers another reason for measuring the power spectrum during the extraction of the bispectrum.

### 2.6.4 Extracting the bias parameters

In this section, it was shown that the bispectrum is a powerful tool to extract the linear bias  $b_1$  and the quadratic bias  $b_2$  due to their different effects on the shape of the bispectrum. This fact will be used to extract these bias parameters in real and redshift space. Therefore, the range of validity of the model must be found out. By extracting  $b_1$  and  $b_2$  from a data set where they are already exactly known this can be done. The dark matter L-BASICC simulations are such a well-known "data set" where  $b_1 = 1.0$  and  $b_2 = 0.0$ . The extraction of the bias can be performed by varying  $b_1$  and  $b_2$  until the best match with the measurement is achieved. The Monte Carlo Markov Chain (MCMC) Metropolis Hastings ansatz (Metropolis et al., 1953; Spergel et al., 2003) will be used for the extraction of  $b_1$  and  $b_2$ . After the results of the dark matter bias parameters estimation are understood and the range of validity of the model is known and biased objects such as halos or galaxies can be examined. These proposed investigations will be carried out in Section 3.2.

This section can be concluded by now because the theoretical framework for the two- and three-point statistics in Fourier space were derived. In the following discussion, the two- and three-point correlation functions which are configuration space counterparts of the power and bi-spectra will be examined in detail.

## 2.7 Correlation functions

The two- and three-point statistics in configuration space, the two-point correlation function  $\xi(r)$  and the three-point correlation function  $\zeta(r_{12}, r_{23}, \theta)$ , respectively, will be discussed in this section with a focus on the two-point statistic. First,  $\xi(r)$  and  $\zeta(r_{12}, r_{23}, \theta)$  will be defined (Section 2.7.1). Secondly, a theoretical description for two-point statistic will be given in Section 2.7.2.  $\zeta(r_{12}, r_{23}, \theta)$  will not be discussed in detail in this work because the later analysis for extracting the bias parameters  $b_1$  and  $b_2$  will be performed in Fourier space. If  $\zeta(r_{12}, r_{23}, \theta)$  needs to be modeled, the bispectrum can be Fourier trans-

formed back to configuration space. However, for this Fourier transform a wide range of bispectra have to be computed in order to receive a smooth and aliasing free bispectrum.

### 2.7.1 Definition of the correlation functions

Before the two- and three-point correlation function can be used, the definition of these two quantities will be given, starting with the lower order. It is worth noting that also in configuration space the two-point statistic is the main ingredient for the computation of the higher point statistics.

The two-point correlation function  $\xi(r)$  is defined as the excess probability compared to a Poisson process of finding two particles at a certain spatial separation  $r$  (Bernardeau et al., 2002)

$$dP_{12}(r) = \bar{n}^2 (1 + \xi(r)) dV_1 \cdot dV_2 \quad (2.104)$$

where  $\bar{n}$  is the mean density. These particles can be galaxies or really just particles, which represent dark matter or dark matter halos. Clustered objects are indicated by  $\xi(r) > 0$  whereas objects are anticorrelated when  $\xi(r) < 0$ . Equation (2.104) can also be expressed as the conditional probability of having a particle at  $dV_1$  with a probability of  $\bar{n}_1 \cdot dV_1$  and of finding another particle at distance  $r$  in  $dV_2$ . Similarly, the definition of the three-point correlation function  $\zeta(r_{12}, r_{23}, \theta)$  (Bernardeau et al., 2002) is given by

$$dP_{123}(r_{12}, r_{23}, \theta) = \bar{n}^3 (1 + \xi(r_{12}) + \xi(r_{23}) + \xi(r_{13}(\theta)) + \zeta(r_{12}, r_{23}, \theta)) dV_1 \cdot dV_2 \cdot dV_3 \quad (2.105)$$

and describes the excess probability compared to a Poisson process of finding three particles separated by  $r_{12}$ ,  $r_{23}$  and  $r_{13}(\theta)$ , where  $\theta$  is defined to be the angle between  $r_{12}$  and  $r_{23}$ . By this choice, the corners of the triangles are given by the spatial positions of the particles. It is worth noting, that the triangle must be closed for the extraction of  $\zeta(r_{12}, r_{23}, \theta)$ , which is sometimes called connected  $\zeta(r_{12}, r_{23}, \theta)$  (Bernardeau et al., 2002). The definition of  $dP_{123}$  clearly visualizes the statement from the beginning of this section, that the two-point correlation  $\xi(r)$  is required for the determination of the three-point correlation function  $\zeta(r_{12}, r_{23}, \theta)$ .

A Gaussian density field would have the result  $\zeta(r_{12}, r_{23}, \theta) = 0$  which means that the whole statistical information of the probability distribution, or for this thesis of the density field, is contained in  $\xi(r)$ . This was shown in Fourier space (see Section 2.5).

The above definitions can be rewritten as follows

$$\xi(r) = \langle \delta(\vec{x}) \delta(\vec{x} + \vec{r}) \rangle \quad (2.106)$$

where the two-point correlation function is only dependent on  $|\vec{r}|$  or just  $r$  because of spherical symmetry. This is the manifestation of homogeneity and isotropy of the density field on large scales. Equation (2.106) can also be understood as a convolution of the density field with itself. The three-point statistics is defined very similarly by

$$\zeta(r_{12}, r_{13}, \theta) = \langle \delta(\vec{x}) \delta(\vec{x} + \vec{r}_{12}) \delta(\vec{x} + \vec{r}_{23}) \rangle \quad (2.107)$$

where again homogeneity and isotropy reduce the originally nine required variables in order to define the triangle uniquely to three, namely  $r_{12}$ ,  $r_{23}$  and  $\theta$ .

So far, the definitions of the clustering statistics for configurations space, which will be used in this thesis, are defined and were discussed. In the following paragraph, it will be discussed how to extract the two-point correlation function from a data set ( $\zeta(r_{12}, r_{23}, \theta)$  will not be considered in later investigation in this thesis anymore). In Peebles (1980), an estimator was introduced to measure  $\xi(r)$  like it is the case in Fourier space, see Section 2.5.

The two-point correlation functions needed in this work are estimated from  $N$ -body simulations by the today commonly used Landy&Szalay estimator (Landy & Szalay, 1993)

$$\xi_{est}(r) = \frac{\langle DD(r) \rangle - 2\langle DR(r) \rangle + \langle RR(r) \rangle}{\langle RR(r) \rangle} \quad (2.108)$$

where  $\langle DD(r) \rangle$ ,  $\langle DR(r) \rangle$  and  $\langle RR(r) \rangle$  are the normalized number of all possible pairs in the data, data-artificial random and within the artificial random set, respectively. The random set is required to include the effects of the survey geometry and the selection function of the survey e.g. selection criteria. The normalization factors are given by

$$N_{DD} = \frac{N_D(N_D - 1)}{2} \quad (2.109)$$

$$N_{DR} = N_D N_R \quad (2.110)$$

$$N_{RR} = \frac{N_R(N_R - 1)}{2} \quad (2.111)$$

where  $N_D$  and  $N_R$  are the number of particles in the data and random set, respectively. There are also other estimators in the literature (Peebles, 1980; Hamilton, 1993) but as it was shown in Kayo et al. (2004) and Kerscher et al. (2000) that the Landy&Szalay estimator converges to the direct estimate the fastest. This means that the Landy&Szalay estimator predicts the two-point correlation function correctly on smaller scales compared to the other proposed estimators (for the same the data set, this scale is strongly dependent on the number of random points contained in the artificial catalog, see Kayo et al. (2004)).

### 2.7.2 Theory of the correlation functions

The estimation of clustering quantities is a very important step for extracting the information they contain about the Universe. However, it must be compared to some theoretical model to obtain e.g. cosmological or bias parameters from the data set. Here, the theoretical discussion will result in a model for extracting the equation of state parameter of dark energy  $w_{DE}$  and the linear bias  $b$  (see Equation (2.75)), even in the presence of redshift errors. Therefore, the anisotropic two-point correlation function  $\xi(r_p, \pi)$  will be introduced, where the distance between the particles is separately binned parallel ( $\pi$ ) and perpendicular to the line-of-sight ( $r_p$ ), where  $r = \sqrt{\pi^2 + r_p^2}$ . After the detailed discussion

of  $\xi(r_p, \pi)$  it will also be explained how to extract  $w_{DE}$  and  $b$  from a data set or in this thesis from numerical  $N$ -body simulations. The following discussion can also be found in Schlagenhauser et al. (2012).

### Two-point correlation function

In this section, the two-point correlation function will not only be derived in redshift space but also the measurement uncertainties of redshifts in a photometric redshift survey, such as Pan-STARRS (Chambers & Pan-STARRS Team, 2004), the upcoming DES (Tucker et al., 2010), or PAU (Benítez et al., 2009) will be incorporated. In the later discussion to this situation will be referred to as redshift error space. A detailed investigation of the correlation function and how it is affected by redshift space distortions and redshift errors was already performed. Therefore, such an investigation will not be performed again in this thesis.

#### $\xi(r_p, \pi)$ in real space

The theoretical means for the two-point correlation function which are required for the analysis in Section 3.1 will be given here. The discussion will be started in real space. The peak of  $\xi(r)$  at around  $110 h^{-1}$  Mpc, which is originated from the BAOs, is an important feature for the later extraction of  $w_{DE}$  and  $b$ .

By measuring the apparent extent of the BAO peak parallel and perpendicular to the line-of-sight, the redshift evolution of  $H(z)$ , see Equation (2.5), and  $D_A(z)$ , see Equation (2.8), can be measured via a simple geometrical relation (Blake & Glazebrook, 2003). The distances between the galaxies are split in a component parallel and perpendicular to the line-of-sight  $\pi$  and  $r_p$ , respectively. This is the so-called anisotropic two-point correlation function  $\xi(r_p, \pi)$ . In this thesis, small angles or high redshifts are assumed so that  $r^2 = r_p^2 + \pi^2$ . The single broad peak of  $\xi(r)$  translates into a ring with a radius of around  $110 h^{-1}$  Mpc for  $\xi_{rs}(r_p, \pi)$ , where the subscripted  $rs$  should indicate real space. In Figure 2.10, the BAO ring can easily be identified but for a full understanding some comments are needed.

The derivation of the three-point statistics was mainly performed by means of  $2^{nd}$  order perturbation theory, whereas for the two-point statistics the  $3^{rd}$  order perturbation theory is used as shown in Section 2.6.1. These results in Fourier space can be inserted in

$$\xi_{ij}(r) = \int \frac{d^3k}{(2\pi)^3} P_{ij}(k) \exp(-i\vec{k}\vec{r}) = \frac{1}{2\pi^2} \int P_{ij}(k) \frac{\sin kr}{kr} k^2 dk \quad (2.112)$$

where  $i$  and  $j$  can be  $\delta$  and  $\theta$  in order to specify which quantity is meant<sup>7</sup>.

For Figure 2.10 only the  $(i, j) = (\delta, \delta)$  is required and also the biased case is shown. A

<sup>7</sup>As a side note the same would be true for the three-point correlation function  $\zeta(r_{12}, r_{23}, \theta)$  and the bispectrum  $B(k_1, k_2, \triangleleft(\vec{k}_1, \vec{k}_2))$ . They are related with each other by

$$\zeta(r_{12}, r_{23}, \theta) = \int \frac{d^3k_1}{(2\pi)^3} \int \frac{d^3k_2}{(2\pi)^3} B(k_1, k_2, \triangleleft(\vec{k}_1, \vec{k}_2)) \exp(-i\vec{k}_1\vec{r}_{12} - i\vec{k}_2\vec{r}_{23}) . \quad (2.113)$$

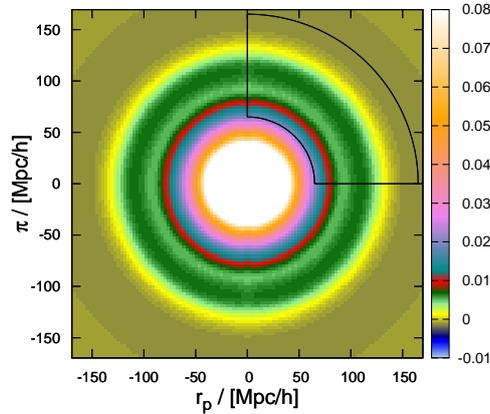


Figure 2.10: The model of this thesis of the two-point correlation function of dark matter halos in real space,  $\xi_{rs}(r_p, \pi)$ , evaluated at  $z = 0.5$  for  $b = 2.65$ . The BAO peak is now a ring at about  $r = \sqrt{r_p^2 + \pi^2} \sim 110 h^{-1} \text{Mpc}$ , here corresponding to the dark green feature. The black lines indicate the range in which the fit to the model is performed: Distances smaller than  $|r| \leq 64 h^{-1} \text{Mpc}$  and larger than  $|r| \geq 165 h^{-1} \text{Mpc}$  will not be taken into account in the analysis. Only one quadrant is used in the fit, since the information in the other three quadrants is redundant. The color bar on the right shows the values of  $\xi_{rs}(r_p, \pi)$ .

detailed discussion of the bias can be found in Section 2.6.1. This thesis is only interested in large scales, on which the linear bias is just defined as a constant multiplicative factor, as in Equation (2.75). A constant multiplicative factor of the power spectrum is also just a constant multiplicative factor for the correlation function. By inserting Equation (2.75) in Equation (2.112) the biased correlation function is given by

$$\xi_{halo/galaxy}(r_p, \pi) = b^2 \xi_{dm}(r_p, \pi) , \quad (2.114)$$

where  $\xi_{dm}(r_p, \pi)$  and  $\xi_{halo/galaxy}(r_p, \pi)$  are the dark matter and halo/galaxy  $\xi_{rs}(r_p, \pi)$ , respectively. The discussion for  $\xi(r_p, \pi)$  is focused on the L-BASICC halo catalog at  $z = 0.5$ . Therefore, for simplicity  $\xi_{rs}(r_p, \pi) = \xi_{halo/galaxy}(r_p, \pi)$ . As a next step the redshift space for  $\xi(r_p, \pi)$  will be derived.

### $\xi(r_p, \pi)$ in redshift space

The derivation of the redshift space distortions was explained in Section 2.6.2 and will not be repeated here. Only the application to  $\xi(r_p, \pi)$  will be given in this section. In the later discussion, the redshift space anisotropic two-point correlation function is called

$\xi_{zs}(r_p, \pi)$ .

On large scales coherent infall of objects onto large structures dominates the redshift-space distortions, the Kaiser effect (Kaiser, 1987). The formulation is based on the distant observer approximation, meaning that the structure is distorted in a plane-parallel fashion, and therefore only the line-of-sight component is affected. The Kaiser effect in its original linear form includes the linear power spectrum; for more details it is also referred to Hamilton (1992). In order to obtain unbiased and more accurate constraints on the cosmological parameters, a more sophisticated description is required for the redshift space distortions (Cole, Fisher & Weinberg, 1994) e.g. the non-linear Kaiser effect (Scoccimarro, 2004). The detailed derivation for  $\xi_{zs}(r_p, \pi)$  is given in Appendix A.2. Following the theoretical derivation of the linear Kaiser effect for  $\xi_{zs}(r_p, \pi)$  in Hamilton (1992) and Hamilton (1993) the anisotropic redshift space two-point correlation function can be expressed as

$$\xi_{zs}(r_p, \pi) = \xi_0(r)P_0(\mu) + \xi_2(r)P_2(\mu) + \xi_4(r)P_4(\mu), \quad (2.115)$$

where  $P_i(\mu)$  are the Legendre polynomials (see Appendix A.3) and the multipoles  $\xi_i(r)$  are given by

$$\xi_0(r) = b^2\xi_{\delta\delta}(r) + \frac{2}{3}fb\xi_{\delta\theta}(r) + \frac{1}{5}f^2\xi_{\theta\theta}(r), \quad (2.116)$$

$$\xi_2(r) = \frac{4}{3}fb[\xi_{\delta\theta}(r) - \bar{\xi}_{\delta\theta}(r)] + \frac{4}{7}f^2[\xi_{\theta\theta}(r) - \bar{\xi}_{\theta\theta}(r)], \quad (2.117)$$

$$\xi_4(r) = \frac{8}{35}f^2 \cdot [\xi_{\theta\theta}(r) + \frac{5}{2}\bar{\xi}_{\theta\theta}(r) - \frac{7}{2}\bar{\bar{\xi}}_{\theta\theta}(r)], \quad (2.118)$$

where  $\xi_{\delta\delta}(r)$ ,  $\xi_{\delta\theta}(r)$  and  $\xi_{\theta\theta}(r)$  are the dark matter density-density, density-velocity and velocity-velocity two-point correlation functions, respectively, Fourier transformed from the corresponding power spectra, and  $f$  is given by Equation (2.89). The barred two-point correlation functions in Equations (2.116)–(2.118) are given by

$$\bar{\xi}_{\delta\theta}(r) \equiv 3r^{-3} \int_0^r \xi_{\delta\theta}(r')r'^2 dr' \quad (2.119)$$

$$\bar{\xi}_{\theta\theta}(r) \equiv 3r^{-3} \int_0^r \xi_{\theta\theta}(r')r'^2 dr' \quad (2.120)$$

$$\bar{\bar{\xi}}_{\theta\theta}(r) \equiv 5r^{-5} \int_0^r \xi_{\theta\theta}(r')r'^4 dr' . \quad (2.121)$$

If the linear power spectrum was used instead of the  $3^{rd}$  order perturbation theory power spectra, the derived formulas above would result in the linear Kaiser effect.

In Figure 2.11, the contours of  $\xi_{zs}(r_p, \pi)$  modeled including the non-linear Kaiser effect are shown. Redshift space distortions destroy spherical symmetry and change the shape and location of the BAO ring. The structure, affected by the coherent infall of objects, looks squashed along the line-of-sight, which also increases the clustering signal compared to  $\xi_{rs}(r_p, \pi)$ . In addition to the change in the amplitude, redshift space distortions also contain information about the cosmological parameters and the bias, as both growth rate

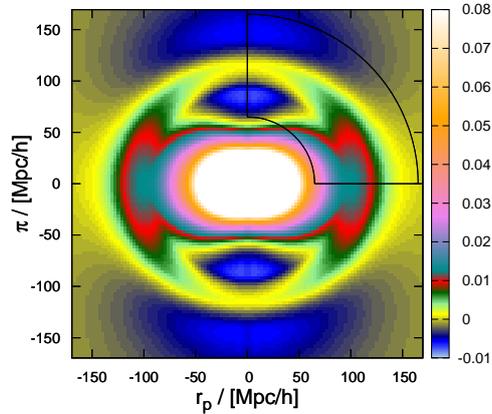


Figure 2.11:  $\xi_{zs}(r_p, \pi)$  of dark matter halos with non-linear Kaiser effect at redshift  $z = 0.5$  for  $b = 2.65$ . As in Figure 2.10 the black lines indicate the range in which the model is fitted.

and linear bias enter into the multipoles which describe them (Equations 2.116–2.118). So far, the model is able to describe redshift space clustering as long as the redshifts (which translate into distances between the observer and the objects) are determined without large redshift errors, as in spectroscopic surveys, where they are negligible. However, collecting a large number of spectroscopic redshifts is time consuming, especially if the aim is to observe a large volume at high redshift. A faster alternative is to estimate redshifts from deep photometric data, which makes it necessary to include the effect of large redshift errors on  $\xi_{zs}(r_p, \pi)$  in the model. How this can be done will be explained in following section.

#### *$\xi(r_p, \pi)$ in redshift error space*

Traditionally photometric redshifts are derived from observed fluxes in five or more broad to medium band filters (Baum, 1962). The probability distribution of the fitted redshifts depends on the spectral type, magnitude and redshift of the observed objects, the filter set and the fitting scheme (e.g. a neural network or a comparison with a library of template spectra, which are used to perform synthetic photometry). However, several authors have found the redshift error distribution in realistic surveys to be very close to Gaussian. For example, Cunha et al. (2009) calculated the full probability distribution function for  $\sim 78$  million SDSS DR7 galaxies using photometric observables and weighted sampling from a spectroscopic subsample of the data, and Saglia et al. (2012) estimated photometric redshifts for objects found in the Medium-Deep Fields of Pan-STARRS1 using available spectroscopic surveys (including SDSS spectra) as training and/or verification sets. From a direct comparison between spectroscopic and photometric redshifts obtained for the same objects, they find the width of the distribution of all galaxies in the sample

to be  $\sigma_z \lesssim 5\%$ , with  $\sim 1\%$  extreme outliers (defined as the fraction of objects for which  $|z_{phot} - z_{spec}| > 0.15 \times (1 + z_{spec})$ ), and  $\sigma_z/(1+z) = 0.024$  and an insignificant fraction of only 0.4% outliers for the object types which will eventually be used to measure BAOs in Pan-STARRS, namely LRGs at  $z \lesssim 0.5$ .

Hence, as a first order approximation a single Gaussian peak is adopted here to simulate the influence of photometric redshift errors on the measurement of  $\xi(r_p, \pi)$ . Since the uncertainty of the redshift estimation mainly results in a distortion of the distance along the line-of-sight, redshift errors have a similar effect as peculiar velocities dominant on small scales, and to first order they do not depend on the local density field.

The main effect of photometric redshift errors is that the line-of-sight component of the distance to a galaxy is smeared out by an amount  $\delta x_{||}$ . This description in Blake & Bridle (2005) and Phleps et al. (2007) will be followed here for modeling the redshift errors where the redshift errors are assumed to be Gaussian distributed such that the spatial displacement  $\delta x$  is given by

$$f(\delta x_{||}) \propto \exp \left[ -0.5 \left( \frac{\delta x_{||}}{\sigma_x} \right)^2 \right], \quad (2.122)$$

where  $\sigma_x$  is the comoving distance corresponding to the rms  $\sigma_z$  of the redshift error probability distribution function at the considered redshift. Equation (2.122) describes the probability distribution function for one single galaxy, but the two-point correlation function is a pairwise statistic, hence in order to simulate the impact on the model  $\xi_{zerr}(r_p, \pi)$  the *pairwise* error distribution function has to be computed. For two galaxies with errors  $\delta x_{||,i}$  and  $\delta x_{||,j}$  the rms of their pairwise error would be

$$\delta x_{||,ij} = \sqrt{\delta x_{||,i}^2 + \delta x_{||,j}^2}. \quad (2.123)$$

If  $N_g$  galaxies are contained in the survey then there are  $N_p = \frac{N_g(N_g-1)}{2}$  pairs. Every galaxy is contained in  $N_g - 1$  pairs, and the overall convolving function  $f_p(\pi)$  is a sum of the Gaussians corresponding to all pairs  $ij$ :

$$f_p(\delta\pi) = \frac{1}{N_p} \sum_{n=1}^{N_p} \exp \left[ -0.5 \left( \frac{\pi}{\delta x_{||,ij,n}} \right)^2 \right]. \quad (2.124)$$

This smearing function is estimated by randomly drawing  $N_g$  values  $\delta x_{||,n}$  from Equation (2.122), from which all possible pairwise errors  $\delta x_{||,ij,n}$  are calculated. Equation (2.124) is the probability distribution of the smearing of the correlation signal along the line-of-sight due to the redshift errors. By convolving  $\xi_{zs}(r_p, \pi)$  with this function, the effect of photometric redshifts on the anisotropic two-point correlation function is modeled. The resulting  $\xi_{zerr}(r_p, \pi)$  for redshift errors with an rms of  $\sigma_z = 0.03$  is shown in Figure 2.12. The redshift errors have the same effect on large scales as the FoG have on small scales, only the order of magnitude is much larger. The clustering signal is smeared out and the structure looks extremely elongated, the BAO ring is almost completely washed out and can barely be identified; with increasing size of the redshift errors the signal-to-noise of

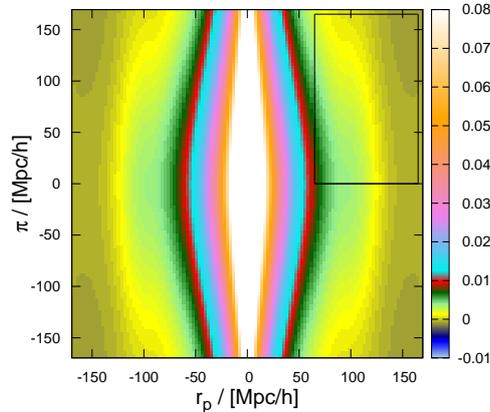


Figure 2.12: The model of  $\xi_{zerr}(r_p, \pi)$  at  $z = 0.5$  for  $b = 2.65$  convolved with a pairwise redshift error distribution, which assumes Gaussian redshift errors with an rms of 0.03. Because the clustering signal is now smeared out along the line-of-sight, in order to capture (to zeroth order) the same information as in the fit of the two-point correlation function in real and redshift space, the spherical shell in which the model is fitted (see Figures (2.10) and (2.11)) is replaced by a cylinder (as indicated by the black rectangular box).

the BAO ring decreases rapidly, until the signal completely vanishes and cannot be used anymore to constrain any cosmological parameters.

### Cosmological test

The peak of  $\xi(r)$  or the ring of  $\xi(r_p, \pi)$  is a standard ruler as explained in Section 2.2. In this section, all distortions of the BAOs in configuration space were discussed. By their inclusion the BAOs can still be used as a standard ruler and will help to distinguish between different cosmological models.

If a wrong cosmology was assumed, incorrect cosmological distances would be calculated, see Section 2.1. This also means that the standard ruler will be wrongly evaluated. In order to find the true cosmology the cosmological parameters have to be varied until the true value of the standard ruler is found. This will be performed in the following way: from the angular extent of the standard ruler on the sky its physical or comoving size is computed for the fiducial cosmological parameters. If the true cosmology was assumed than the size of standard ruler is calculated correctly. If this is not the case the true cosmology has to be found with the help of scaling relations where the cosmological parameters are changed until the true size or position of the standard ruler is found. The scaling relations are given by

$$r_p^t = \frac{D_A^t(z)}{D_A^a(z)} \cdot r_p^a \quad (2.125)$$

and

$$\pi^t = \frac{H^a(z)}{H^t(z)} \cdot \pi^a, \quad (2.126)$$

where the quantities for the assumed cosmology are superscripted with  $a$  and the true cosmology with  $t$ . For  $\xi_{rs}(r_p, \pi)$  this means that by assuming a wrong cosmology the size of the BAO ring along and perpendicular to the line-of-sight is not the same. The correct cosmology is found when these two sizes coincide (Alcock & Paczynski, 1979). In redshift or redshift error space, isotropy is destroyed and this statement is not fully correct anymore. However, with the theoretical concepts derived in this chapter it is possible to account for their effects on the two-point clustering and the scaling relations can be utilized to extract the correct cosmological parameters.

If only one cosmological parameter shall be varied it is obvious that the remaining parameter should be known very well. In Section 3.1,  $w_{DE}$  will be extracted from the L-BASICC simulations by means of this cosmological test.

### 2.7.3 Redshift errors for the three-point statistic?

Before this chapter will be concluded, a qualitative discussion will be given, why the examination of redshift errors on the three-point statistics is excluded from this thesis. In Figure 2.13, a cuboid with a side length of  $200 h^{-1}$  Mpc for the x- and z- direction and  $500 h^{-1}$  Mpc for the y-direction at  $z = 0.5$  of the L-BASICC simulation number 48, in which  $\xi_{48}(r)$  behaves mostly like the mean  $\xi(r)$ , is projected along the z-axis. On the top panel, the blue dots represent the position of the halos in real space and the red ones the halos in redshift space. Their positions are distorted along the y-axis of the simulation box, which was defined to be the line-of-sight direction. The shift can easily be identified.

In redshift space, the structures look squashed, as it is expected. These structures are still visible and it is possible to extract a clear clustering signal from this simulation. On the bottom panel, Gaussian redshift errors with a rms of  $\sigma_z = 0.03$  were applied to the same L-BASICC simulation in redshift space with periodic boundary conditions (of course along the y-axis). Almost no structures can be detected anymore. As shown for the BAO peak in  $\xi(r_p, \pi)$ , prominent features are washed out.

The signal of the three-point clustering statistics will also be reduced by redshift errors. The signal-to-noise is anyway lower the higher the clustering statistics is and with redshift errors the situation gets worse. The triangles which are counted by the three-point correlation function will look elongated in redshift error space. The spherical average for the calculation of the three-point statistics would not make any sense because homogeneity and isotropy are destroyed by the redshift errors anyway. Therefore, such an analysis is excluded from this thesis.

As mentioned earlier, the three-point statistics is not as thoroughly investigated as the two-point statistics. The basic effects of non-linear structure growth, biasing and peculiar velocities must be examined first, before further complication are incorporated in the model.

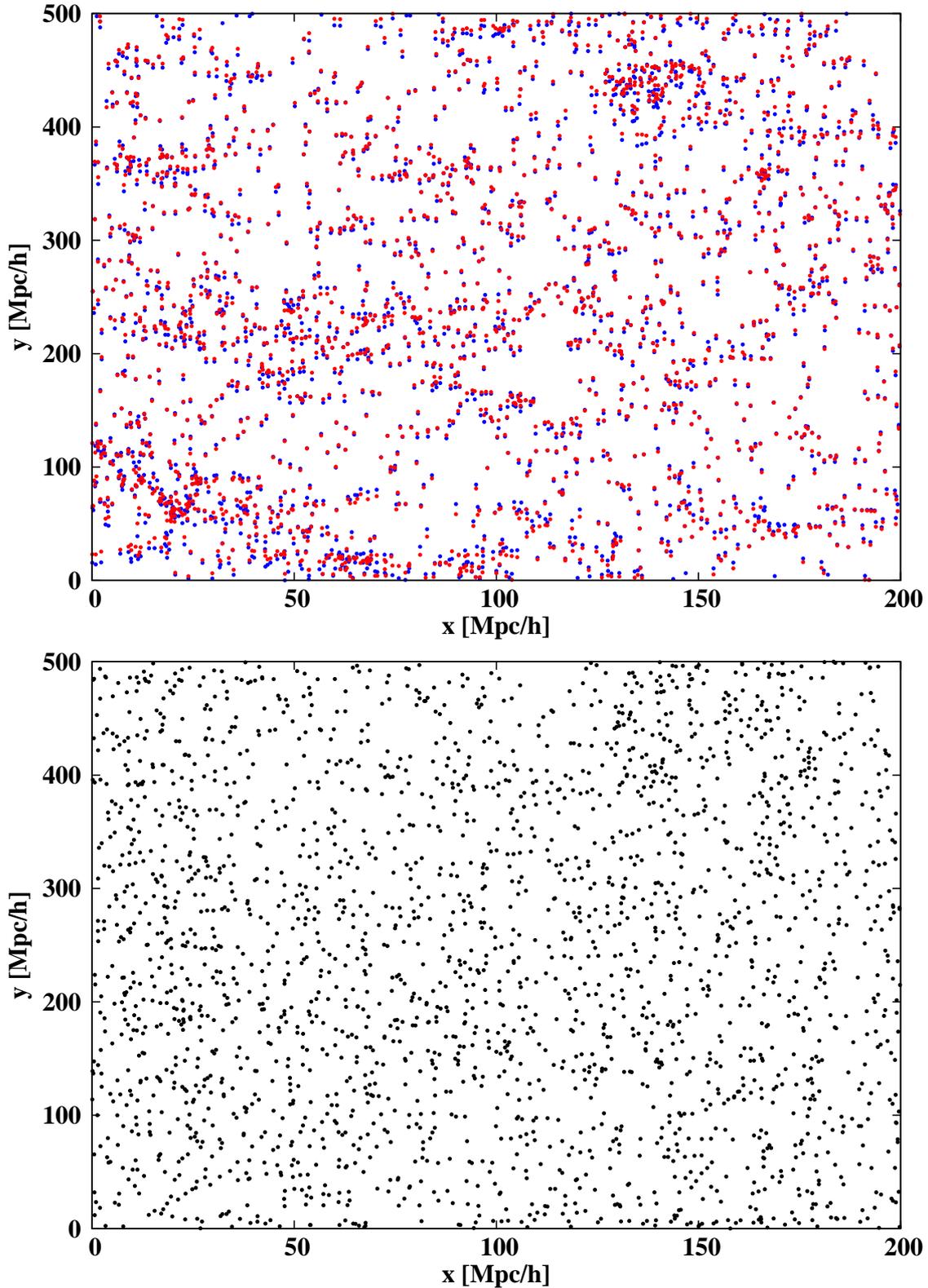


Figure 2.13: A cuboid projected along the  $z$ -axis with a side lengths  $(l_x, l_y, l_z) = (200 h^{-1} \text{ Mpc}, 500 h^{-1} \text{ Mpc}, 200 h^{-1} \text{ Mpc})$  was taken from the dark matter halo L-BASICC simulation box number 48 at  $z = 0.5$ . Real and redshift space are combined in the upper panel and are represented by the blue and the red dots, respectively. The lower panel shows the effect of Gaussian distributed redshift errors with a rms of  $\sigma_z = 0.03$ .

# Chapter 3

## Investigation of $\xi(r_p, \pi)$ and $B(k_1, k_2, k_3)$ : Extracting the parameter of the equation of state of dark energy and the bias parameters

This chapter is divided into two smaller main sections. The first section will be focused on the two-point statistics whereas the second is mainly interested in the bispectrum. Section 3.1 will describe how to extract the equation of state parameter of dark energy and the linear bias from  $\xi(r_p, \pi)$  as well as from  $w(r_p)$ . A huge part of this section is from Schlagenhauser et al. (2012). In Section 3.2, the extraction of the linear and the quadratic bias from the bispectrum will be investigated in detail.

### 3.1 Two-point statistics: Determination of $w_{DE}$ and $b$

This first big section will present the results of the estimation of the equation of state parameter of dark energy  $w_{DE}$  and the linear bias  $b$  by means of  $\xi(r_p, \pi)$  and  $w(r_p)$ . In the discussion of the two-point statistics, the linear bias is called  $b$  and is defined by Equation (2.75), while for the three-point statistics the linear bias  $b_1$  is associated with the first term of the series expansion of Equation (2.77).

It is clear from this definitions that the  $b$ -variable is an effective quantity which contains contributions from all  $b_n$ -parameters of the series expansion. The anisotropic two-point correlation function  $\xi(r_p, \pi)$ , see Section 3.1.2, and the projected correlation function  $w(r_p)$ , see Section 3.1.3, will be utilized for the determination of  $w_{DE}$  and  $b$ . These two quantities were selected for this task because they have special properties.

$\xi(r_p, \pi)$  is a two-dimensional function and its information content is not combined into a few bins like it is the case for  $\xi(r)$ . More information can be used to constrain cosmological parameters, in this thesis  $w_{DE}$  and  $b$ . This underlines the importance of a careful modeling of this quantity, see Section 2.7.2.

$w(r_p)$  is the projected correlation function and the result when  $\xi(r_p, \pi)$  is integrated along the line-of-sight component (the  $\pi$ -direction). Ideally, this integration should be performed to infinity in order to recover the three-dimensional real space correlation function  $\xi(r)$ . Then,  $w(r_p)$  would be better suited for extracting  $w_{DE}$  and  $b$ . Therefore,  $w(r_p)$  will also be examined.

The exact procedure for performing the analysis will be discussed in Section 3.1.1. The investigation is carried out at  $z = 0.5$  only. This is the expected mean redshift of Pan-STARRS (Cai et al., 2009), on which the model will be applied to in a future analysis. In the Sections 3.1.2 and 3.1.3,  $w_{DE}$  and  $b$  are extracted from the L-BASICC dark matter halo catalogs by means of  $\xi(r_p, \pi)$  and  $w(r_p)$ , respectively. This whole section will be concluded with Section 3.1.4 in which the introduced model of  $\xi(r_p, \pi)$  will be qualitatively compared to similar approaches.

### 3.1.1 Constraining $w_{DE}$ and $b$

The capability of the model of  $\xi(r_p, \pi)$  will be demonstrated, which was described in the previous section, to constrain  $w_{DE}$  and  $b$ , by fitting the correlation function computed from the L-BASICC simulations. The anisotropic two-point correlation function is calculated using the estimator of Landy & Szalay (1993) in each single box up to a distance of  $\pm 300.0 h^{-1}$  Mpc, and in bins of  $3.0 h^{-1}$  Mpc in both  $r_p$  and  $\pi$ , for real, redshift and redshift error space, respectively. In the latter case “redshift errors”, that is, small offsets, are added to the coordinate which has been designated the line-of-sight. These offsets have been randomly drawn from a Gaussian error distribution function with a given rms. From the 50 estimates of  $\xi(r_p, \pi)$  the mean is calculated.

The model  $\xi(r_p, \pi)$  is also evaluated from  $(-300 h^{-1} \text{ Mpc}, -300 h^{-1} \text{ Mpc})$  to  $(300 h^{-1} \text{ Mpc}, 300 h^{-1} \text{ Mpc})$ , in steps of  $(0.5 h^{-1} \text{ Mpc}, 0.5 h^{-1} \text{ Mpc})$ . In order to facilitate an accurate comparison with the simulation, which has been calculated with a six times larger bin size, one averages over the model bins when fitting  $w_{DE}$  and  $b$ .

While keeping  $H_0$ ,  $\Omega_m$  and  $\Omega_\Lambda$  fixed at the values determined by the L-BASICC simulation,  $w_{DE}$  and  $b$  are varied, which both change the amplitude of the two-point correlation function and the shape of the redshift space distortions (see Section 2.6.2 and 2.7.2). When the correlation function of the simulation boxes is calculated, a fiducial cosmology has to be assumed – for simplicity the input parameters of the L-BASICC were chosen, i.e. the correct cosmology (and thus the correct correlation function with the acoustic peak at the position predicted by theory). For any other choice of cosmological parameters the measured redshifts and positions on the sky translate differently into distances perpendicular and parallel to the line-of-sight, which leads to a distortion (Alcock & Paczynski, 1979). Therefore when the model is fitted to the data, the distances have to be rescaled according to the Equations (2.126) and (2.125) for the parallel and perpendicular to the line-of-sight component, respectively.

The model and the halo catalog correlation functions are compared within a spherical shell of  $64 h^{-1} \text{ Mpc} \leq \sqrt{r_p^2 + \pi^2} \leq 165 h^{-1} \text{ Mpc}$  in both real and redshift space. In redshift error space the shape of that shell is distorted due to the smearing of the redshift

errors, so in order to sample similar scales the spherical shell was replaced by a cylinder, as indicated by a rectangle in Figure 2.12. The corners of the rectangle are defined by  $[64 h^{-1} \text{ Mpc}, 0 h^{-1} \text{ Mpc}]$  (left lower corner),  $[64 h^{-1} \text{ Mpc}, 165 h^{-1} \text{ Mpc}]$  (left upper corner),  $[165 h^{-1} \text{ Mpc}, 0 h^{-1} \text{ Mpc}]$  (right lower corner) and  $[165 h^{-1} \text{ Mpc}, 165 h^{-1} \text{ Mpc}]$  (right upper corner).

The model was tested for two different cases, in one case the information contained in the amplitude was taken into account, and in the other case it was analytically marginalized over the amplitude using the scheme described in Lewis & Bridle (2002). In the latter case only the shape of the model is examined.

The fit is performed by means of a Monte Carlo Markov Chain (MCMC) after averaging and rescaling in order to find the best-fitting values for  $w_{DE}$  and  $b$ , and to estimate their errors. Due to the limited number of realizations it is unfortunately not possible to calculate correct (invertible) covariance matrices (the full covariance matrix would take  $100 \times 100$  pixels into account, hence at least  $100 \times 100 + 1$  independent simulation boxes are required). If only the variance of the correlation function is used when fitting the *mean*  $\xi(r_p, \pi)$ , the resulting errors of  $w_{DE}$  and  $b$  (in real and redshift space of the order of  $\sim 2\%$ ) are clearly underestimated.

Instead it is assumed that the scatter of 50 fits to the *single* realizations can at least partly account for cosmic variance and the otherwise ignored contribution of the off-diagonal elements in the covariance matrix. Therefore, the variance of the best-fitting values inferred in this way can be thought to represent a more realistic estimate of the errors of the fit parameters, while leaving the actual best-fitting values unchanged.

Indeed, although the resulting values of  $w_{DE}$  and  $b$  reassuringly do not depend on the way of fitting, the size of the errors does. Throughout this thesis with errors is meant that they have been calculated from the variance of the 50 values of  $w_{DE}$  and  $b$ , which are about an order of magnitude larger than those inferred from the fit to the mean.

The range in which the parameters are fitted are restricted to  $-1.6 \leq w_{DE} \leq -0.4$  and  $1 \leq b \leq 20$ . 40 000 steps are sufficient for the chain to converge towards the best-fitting values and to explore the likelihood. Since the calculation of the  $3^{rd}$  order perturbation theory power spectrum is time consuming, to speed up the analysis the MCMC is ran on a grid, for which a library of correlation functions  $\xi(r)$  is constructed for  $-1.6 \leq w_{DE} \leq -0.4$  in steps of  $-0.001$ . Once the random sampling process of the MCMC has chosen a new value for  $w_{DE}$ , the appropriate  $\xi(r)$  for the nearest value of  $w_{DE}^{lib}$  is selected from the library and used as starting point for the calculation of the model  $\xi(r_p, \pi)$  (and  $w_{DE}$  set to  $w_{DE}^{lib}$ ).

### 3.1.2 Anisotropic two-point correlation function

In this section, the results of the fits of  $\xi_{rs}(r_p, \pi)$  (real space),  $\xi_{zs}(r_p, \pi)$  (redshift space) and  $\xi_{zerr}(r_p, \pi)$  (redshift error space) will be discussed in detail. The analytic model will be tested by fitting  $w_{DE}$  and  $b$  against  $\xi(r_p, \pi)$  of the halo catalog from the L-BASICC simulations at  $z = 0.5$  by following the description of Section 3.1.1.

## Real Space

In order to investigate the validity of the  $3^{rd}$  order perturbation theory, first the model is tested against the dark matter anisotropic two-point correlation function of the L-BASICC simulations, where  $b = 1.0$ . The best-fitting parameters are found to be  $w_{DE} = -0.992 \pm 0.091$ ,  $b = 0.998 \pm 0.073$ , from which it can be concluded that the non-linear structure growth is modeled accurately enough to obtain unbiased estimates of these parameters. If the description of the clustering of collapsed objects like galaxies – or dark matter halos – which are biased tracers of the dark matter density field, is wanted, the bias has to be included (see Section 2.6.1) in this calculation.

Since in real space the linear bias is only a multiplicative factor (for the two-point statistics) which boosts the amplitude of the dark matter correlation function, but does not alter its shape, any information about the bias is contained in the amplitude at a given radius  $r = \sqrt{r_p^2 + \pi^2}$ . Including the amplitude in the fit of  $\xi_{rs}(r_p, \pi)$  of the dark matter halos in the 50 L-BASICC boxes one obtains  $w_{DE} = -1.010 \pm 0.117$  and  $b = 2.641 \pm 0.183$ , if the amplitude is not taken into account, the bias cannot be fitted, but the same value for  $w_{DE}$  is obtained when the amplitude information is disregarded. The measured value of  $w_{DE}$  is in good agreement with the fiducial value of the simulation. Since the simulation has only medium-resolution, it does not contain very small halos (the lowest mass halo has ten particles, which corresponds to a minimum mass of  $M_{\min} = 1.76 \times 10^{13} h^{-1} M_{\odot}$ ), hence a relatively large mean bias measured for all dark matter halos is expected.

In Figure 3.2, the contours of the best-fitting model (black solid line) are plotted over the  $\xi_{rs}(r_p, \pi)$  of the L-BASICC dark matter halos the values of  $\xi_{rs}(r_p, \pi)$  are color coded in a logarithmic fashion; only the scales taken into account in the fit are shown. Figure 3.1 shows cuts through the  $\xi_{rs}(r_p, \pi)$  plane along constant  $r_p$  for both data and best-fitting model (including the  $1\sigma$  uncertainty limits calculated from the variance of the correlation functions in the single L-BASICC boxes). The deviations between the model and the L-BASICC simulations are small compared to the errors. The obtained value of  $w_{DE}$  is in good agreement with the fiducial value of the simulation.

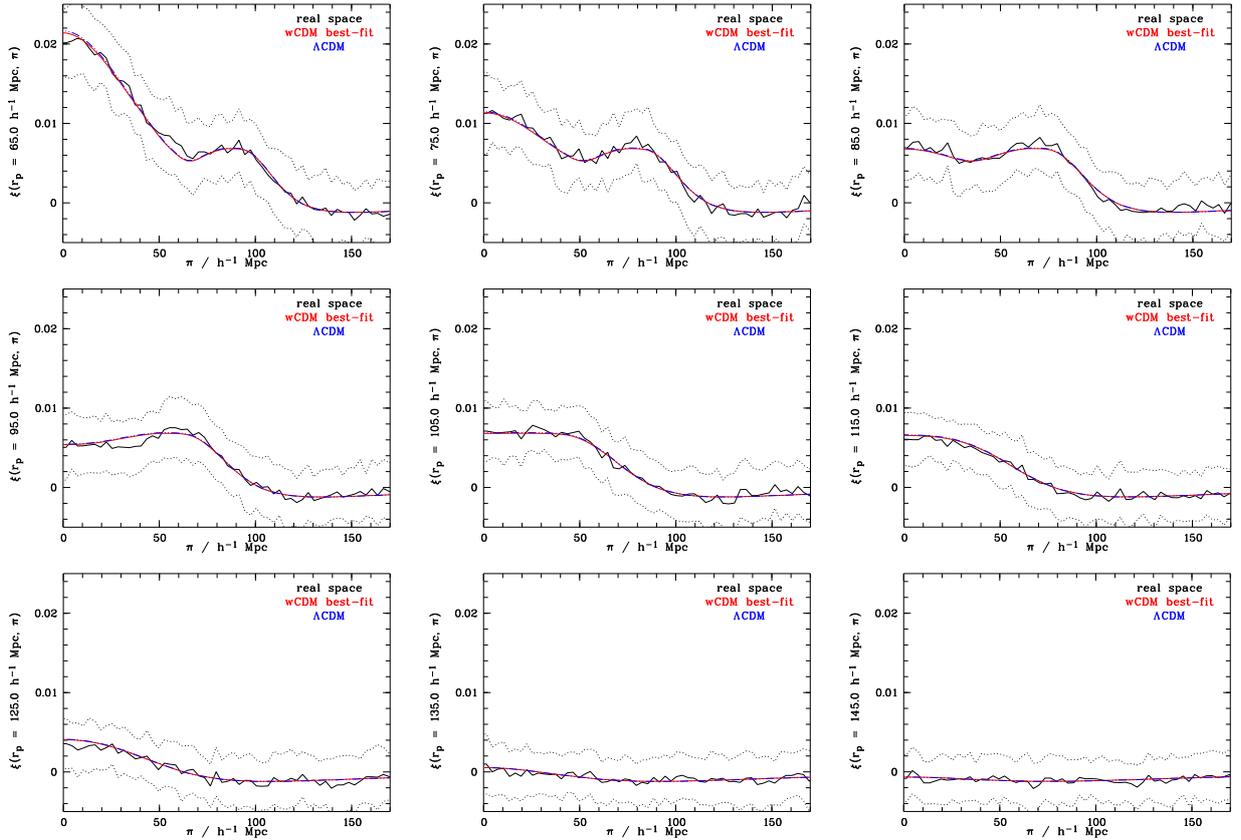


Figure 3.1: Cuts through the real space anisotropic two-point correlation function  $\xi_{rs}(r_p, \pi)$  of the L-BASICC dark matter halos along fixed  $r_p$ , black solid lines: mean, dotted lines:  $1\sigma$ -deviation calculated from the variance of the 50 boxes, red solid line: best-fitting wCDM model, blue dot-dot-dashed line:  $\Lambda$ CDM case.

### Redshift Space

In redshift space, the exact positions of the galaxies (and therefore the correlation function) are distorted due to the additional Doppler shift induced by their peculiar velocities, and thus mainly affects the line-of-sight components of  $\xi(r_p, \pi)$ . On large (BAO) scales coherent infall dominates, which in previous models of the anisotropic two-point correlation function was assumed to be linear and modeled following the description in Kaiser (1987). In this thesis, the non-linear Kaiser effect (Scoccimarro, 2004) is applied to the model of  $\xi_{zs}(r_p, \pi)$ . The validity of the approach of Scoccimarro (2004) has been tested by Jennings et al. (2011), who found a good match to simulations. Since the size and angular dependence of the effect depends on the bias of the objects (the bias is also contained in the quadrupole and hexadecapole needed to evaluate the model  $\xi_{zs}(r_p, \pi)$ , see Section 2.6), in redshift space, it is possible to infer the value of  $b$  from the shape of  $\xi_{zs}(r_p, \pi)$  alone, in contrast to real space.

The results of the MCMC analysis (again fitted in the range  $64.0 \leq \sqrt{r_p^2 + \pi^2} \leq 165.0 h^{-1}$

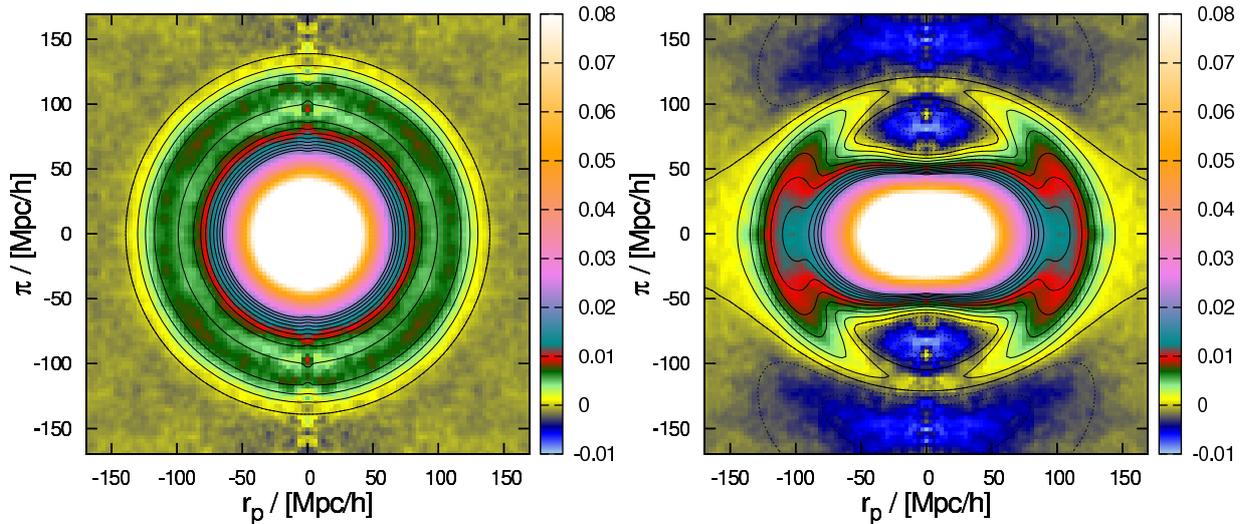


Figure 3.2: Contours from the best-fit model plotted over the anisotropic two-point correlation function calculated from the L-BASICC dark matter halo catalogs in real (left) and redshift space (right), respectively.

Mpc) are summarized in Table 3.1. The comparison of the best-fitting model (black solid line) with the logarithmically color coded L-BASICC  $\xi_{zs}(r_p, \pi)$  is shown in Figure 3.2.

	$w_{DE}$	$b$
shape only	$-1.012 \pm 0.139$	$2.518 \pm 0.646$
shape and amplitude	$-1.020 \pm 0.147$	$2.633 \pm 0.222$

Table 3.1:  $w_{DE}$  and  $b$  from the model  $\xi_{zs}(r_p, \pi)$ : Mean and variance of the 50 L-BASICC boxes, fit in the range  $64.0 \leq \sqrt{r_p^2 + \pi^2} \leq 165.0 h^{-1}$  Mpc. The fit has been carried out for two cases, one where only the shape was input to the fit and the amplitude was marginalized over, and one where both shape and amplitude have been taken into account.

The two-point correlation function of the L-BASICC halo catalogs can be well described by the model. The model contours match the L-BASICC  $\xi_{zs}(r_p, \pi)$  almost perfectly (see Figure 3.3), and the resulting values of  $w_{DE}$  and  $b$  are in good agreement with the real space estimates, too. As expected, the errors in redshift space are larger than in real space. Also the error on  $b$  is larger if the information contained in the amplitude is ignored, the determined values of  $w_{DE}$  and  $b$  are however consistent and do not depend on whether the amplitude is taken into account or not.

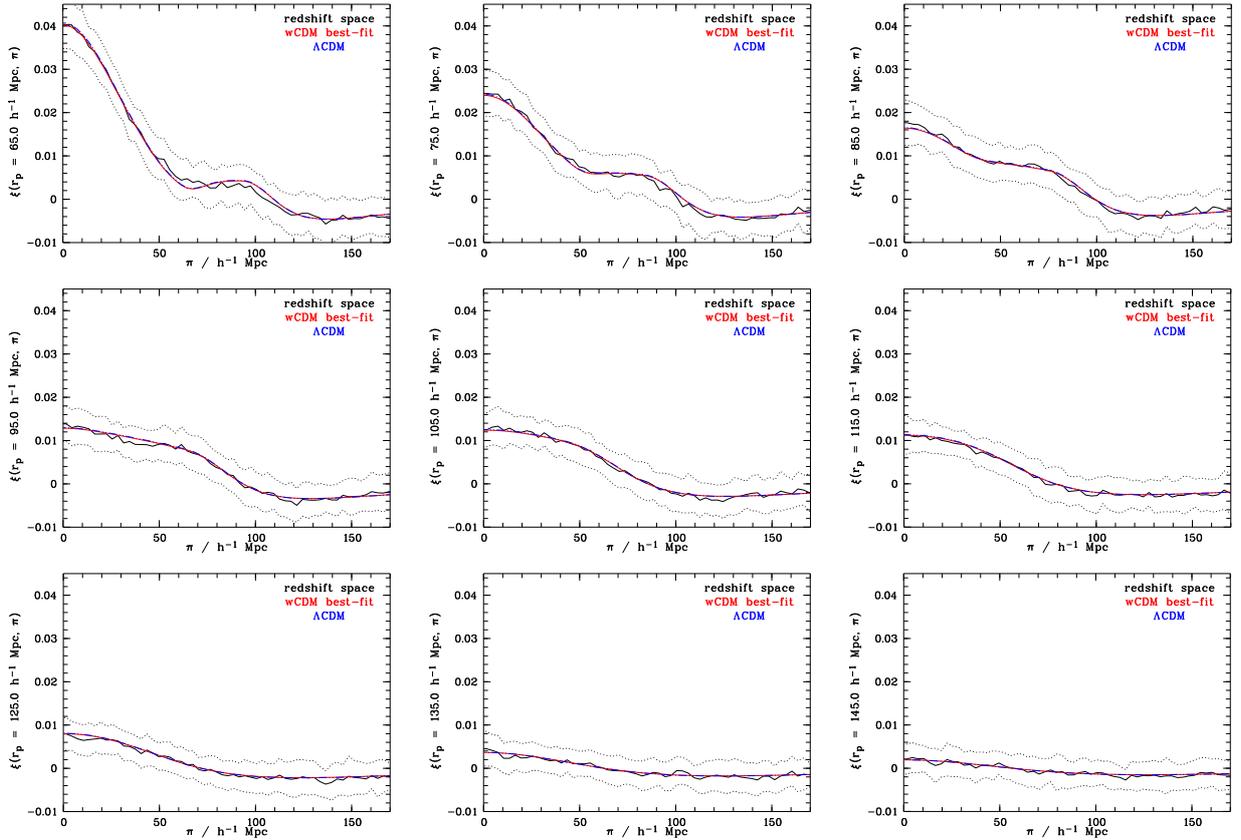


Figure 3.3: Cuts through the redshift space anisotropic two-point correlation function  $\xi_{zs}(r_p, \pi)$  of the L-BASICC dark matter halos along fixed  $r_p$ , black solid lines: mean, dotted lines:  $1\sigma$ -deviation calculated from the variance of the 50 boxes, red solid line: best-fitting wCDM model, blue dot-dot-dashed line:  $\Lambda$ CDM case.

### Redshift Error Space

One of the advantages of using the anisotropic two-point correlation function  $\xi(r_p, \pi)$  to infer cosmological parameters is that in the presence of redshift space distortions the clustering measurement perpendicular to the line-of-sight remains almost unaffected, while distortions along the line-of-sight can be modeled and thus properly taken into account. Due to this fact, it is a perfect tool to use in the case of photometric redshifts, the large errors of which lead to a rather dramatic distortion, as explained in Section 2.6.

In order to investigate the effect of photometric redshift errors on the estimate of  $w_{DE}$  and  $b$ , the influence of a Gaussian redshift error distribution with a rms of  $\sigma_z = 0.015, 0.03, 0.06,$  and  $0.12$ , respectively, was simulated on the measurement. As described in Section 3.1.1 the spherical shell in which the fit was carried out has now been replaced by a cylinder (indicated by the rectangular box in the  $r_p, \pi$ -plane in Figure (2.12)) in order to compare (to zeroth order) the same information. It was found out that in the last case a large part of the clustering signal is smeared out to distances much larger than the  $300 h^{-1}$  Mpc

where the model is calculated for, the BAO ring disappears, and the noise increases such that an accurate estimate of  $w_{DE}$  and  $b$  becomes impossible.

While still fitting  $\xi(r_p, \pi)$  only up to  $165 h^{-1}$  Mpc, extending the model to distances  $\pi = 2000 h^{-1}$  Mpc before convolving it with the pairwise redshift error distribution allows one to recover some of the clustering signal for redshift errors (at least for  $\sigma_z \leq 0.06$ ), but its information content is limited due to the low signal-to-noise of the data on these scales. The values of  $w_{DE}$  and  $b$  which are found for  $\sigma_z = 0.015, 0.03, 0.06$ , and  $0.12$  are summarized in Table 3.2; contours of the corresponding models of  $\xi_{zs}(r_p, \pi)$  are shown in comparison to the logarithmically color coded measurement from the data in Figure 3.4.

$\sigma_z$		$w_{DE}$	$b$
0.015	only shape	$-0.965 \pm 0.298$	$3.660 \pm 2.694$
	shape and amplitude	$-0.980 \pm 0.296$	$2.704 \pm 0.333$
0.030	only shape	$-0.883 \pm 0.313$	$6.309 \pm 3.603$
	shape and amplitude	$-0.966 \pm 0.363$	$2.622 \pm 0.415$
0.060	only shape	$-1.081 \pm 0.344$	$5.336 \pm 3.931$
	shape and amplitude	$-1.036 \pm 0.402$	$2.609 \pm 0.512$
0.120	only shape	$-1.316 \pm 0.348$	$4.904 \pm 4.537$
	shape and amplitude	$-1.199 \pm 0.424$	$2.295 \pm 0.554$

Table 3.2:  $w_{DE}$  and  $b$  from the fit of the model  $\xi_{zerr}(r_p, \pi)$  to the 50 L-BASICC boxes for  $\sigma_z = 0.015, 0.03, 0.06$ , and  $0.12$  (from top to bottom). Again the fit has been carried out for two cases, one where only the shape was input to the fit and the amplitude was marginalized over, and one where both shape and amplitude have been taken into account.

In Figure 3.5, Figure 3.6, Figure 3.7, and Figure 3.8 again cuts along constant values of  $r_p$  through  $\xi_{zerr}(r_p, \pi)$  are shown for  $\sigma_z = 0.015, 0.03, 0.06$ , and  $0.12$ , respectively. The model fits the data well, although slightly worse as expected than in real and redshift space (see Figures 3.1 and 3.3).

Figure 3.9 shows the values of  $w_{DE}$  and  $b$  and their corresponding errors for the fit of the model to the 50 L-BASICC boxes including increasing widths of the redshift errors. As long as the errors are smaller than  $\sigma_z \approx 0.06$ , the measurement is unbiased. The increase of the errors in redshift error space can be expected: Due to the convolution with the pairwise redshift error distribution not only the clustering signal is smeared out along the line-of-sight, but so is the noise and thus the errors on the cosmological parameters inferred from the measurement increase. Therefore, within a given bin in the  $\xi_{zerr}(r_p, \pi)$  plane the variance is increased compared to redshift space, and the larger the redshift errors, the larger the increase. The loss of information contained in the multipoles due to the distortion by the convolution with the pairwise redshift error distribution function also means that the linear bias  $b$  cannot be determined accurately by using the shape of  $\xi_{zerr}(r_p, \pi)$  only.

Large redshift errors also increase the probability for the MCMC not to converge within

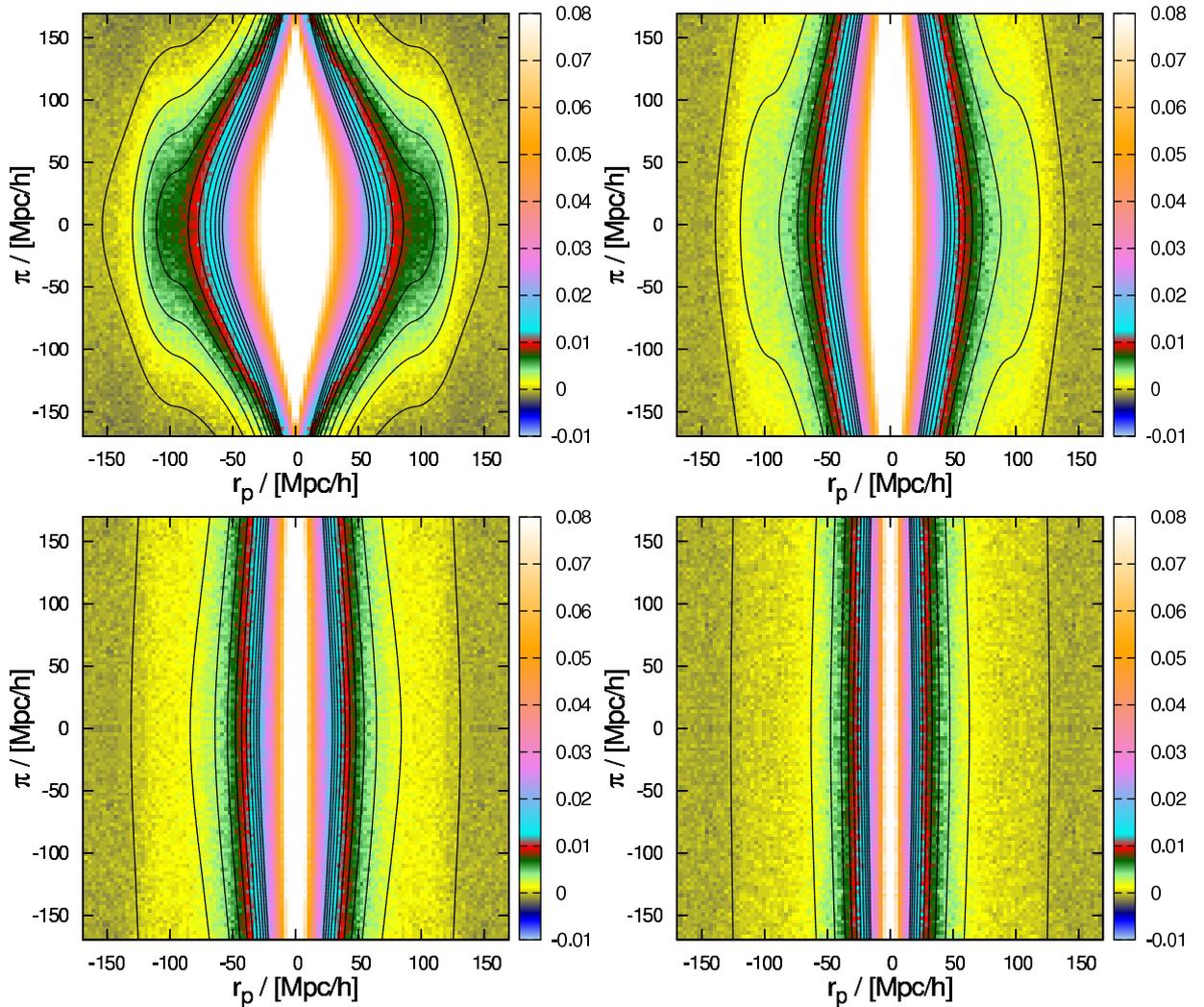


Figure 3.4: L-BASICC  $\xi_{zerr}(r_p, \pi)$  and contours from the model  $\xi_{zerr}(r_p, \pi)$  for  $\sigma_z = 0.015$  (top left),  $\sigma_z = 0.03$  (top right)  $\sigma_z = 0.06$  (bottom left) and  $\sigma_z = 0.12$  (bottom right).

the allowed parameter space – cosmic variance is still large even in boxes of the size of the L-BASICC simulations, and in some of them the BAO ring is almost invisible even in real space. In such cases, redshift errors finally destroy all of the information that might have been there before, and the fit fails. The larger the redshift errors the more catastrophic failures are produced. Even if the information contained in the amplitude is used, the fraction of boxes where the correlation function can not be fit can be as high as  $\sim 40\%$  for  $\sigma_z = 0.06$ . For  $\sigma_z = 0.12$  the model tends to yield biased results, as can be seen from the “best-fitting” values.

The exact size of the redshift error at which the values of  $w_{DE}$  and  $b$  can not be measured accurately anymore and their errors become unacceptably large certainly depends on the exact shape of the redshift error distribution function (a more Lorentz-like distribution

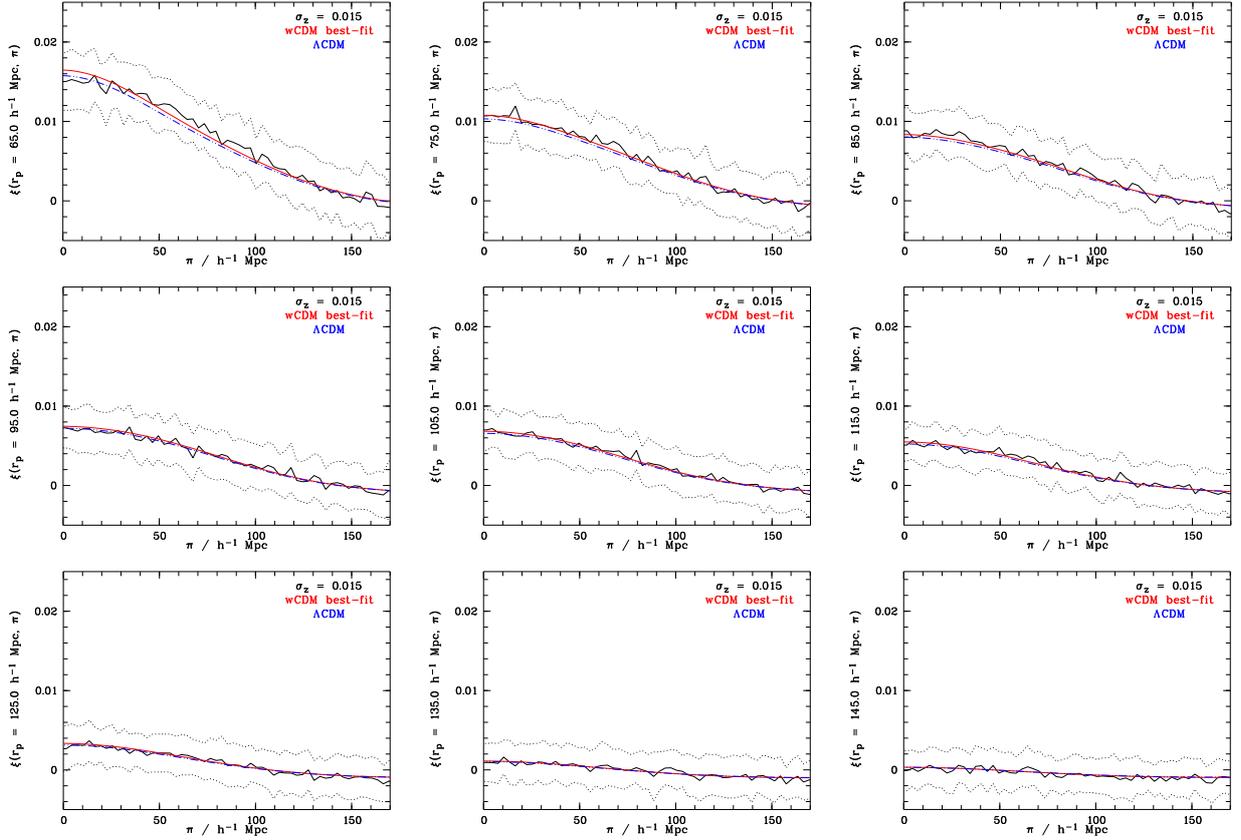
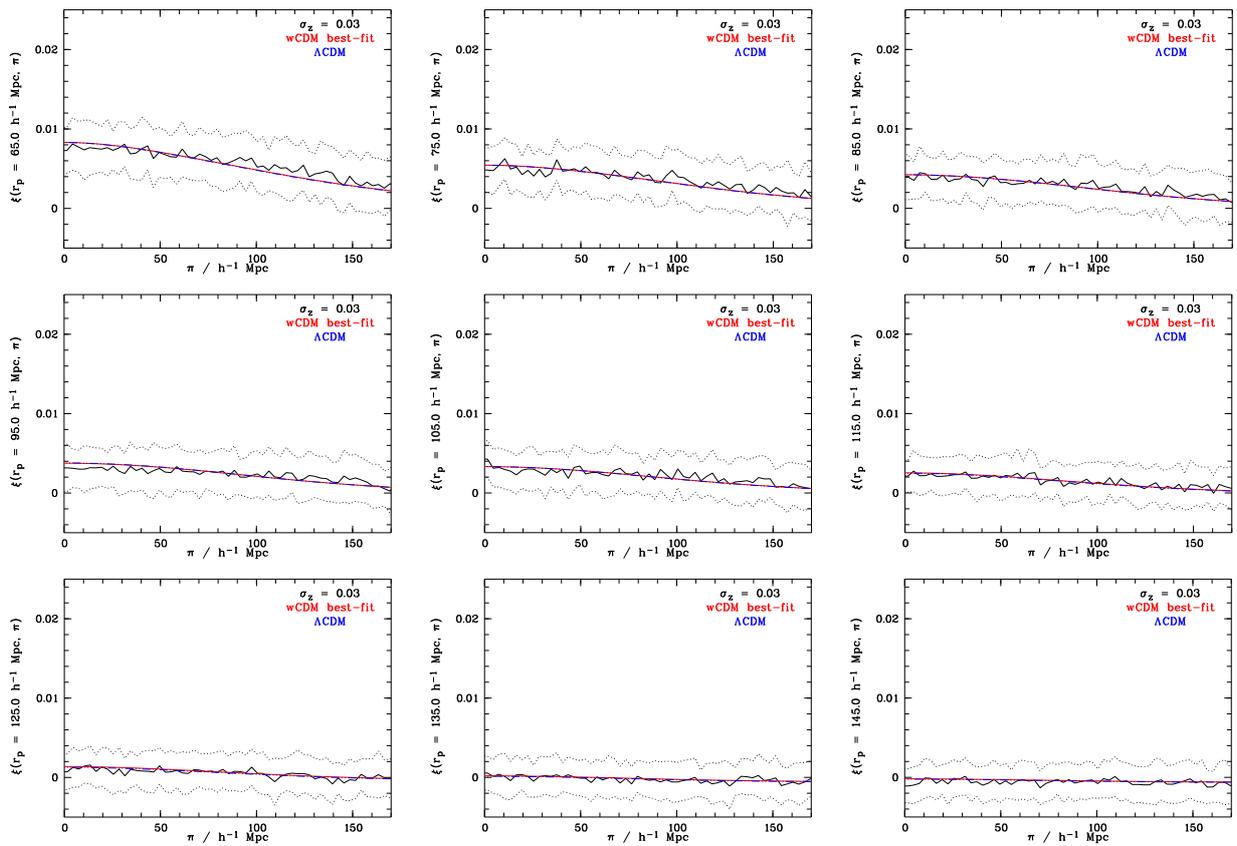


Figure 3.5: Cuts through the redshift error space anisotropic two-point correlation function  $\xi_{zerr}(r_p, \pi)$  of the L-BASICC dark matter halos along fixed  $r_p$ , for redshift errors of  $\sigma_z = 0.015$ , black solid lines: mean, dotted lines:  $1\sigma$ -deviation calculated from the variance of the 50 boxes, red solid line: best-fitting wCDM model, blue dot-dot-dashed line:  $\Lambda$ CDM case.

with broad wings will have a larger impact on  $\xi(r_p, \pi)$  than a Gaussian with a comparable width of the core), and it most certainly also depends on the volume and/or number density of the survey: since the reason the fit fails is mainly that the BAO feature vanishes in the increasing noise, the larger the signal-to-noise on large scales, the larger the redshift errors may be at which the disappearance of the BAO ring occurs. There may also be the possibility to improve the signal-to-noise ratio for photometric data by using the full probability distribution function of the redshifts in combination with a set of spectroscopic redshifts in the same area and redshift range, a method that has been shown to be able to improve the clustering signal strength in a manner equivalent to increasing the survey size by a factor 4-5 by Myers et al. (2009).

Figure 3.6: As in Figure 3.5, but for  $\sigma_z = 0.03$ .

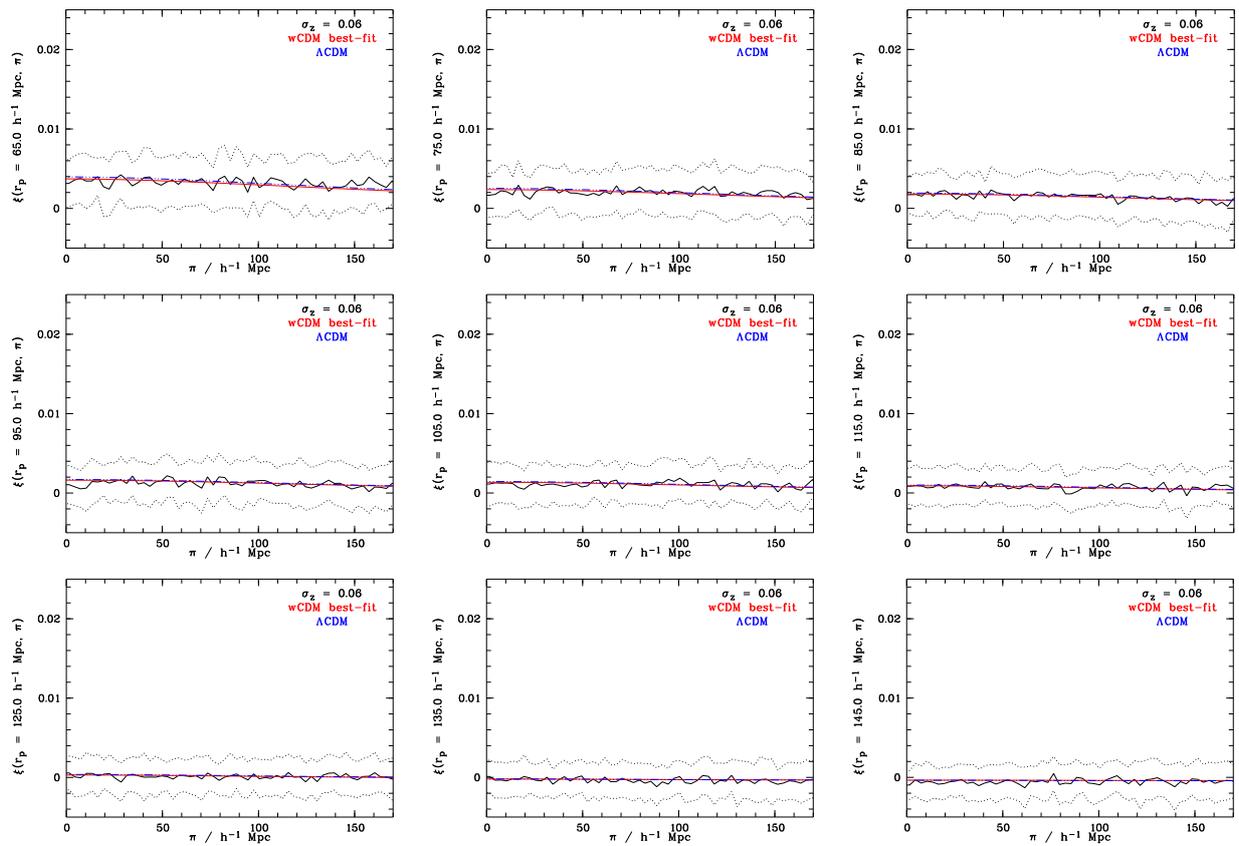
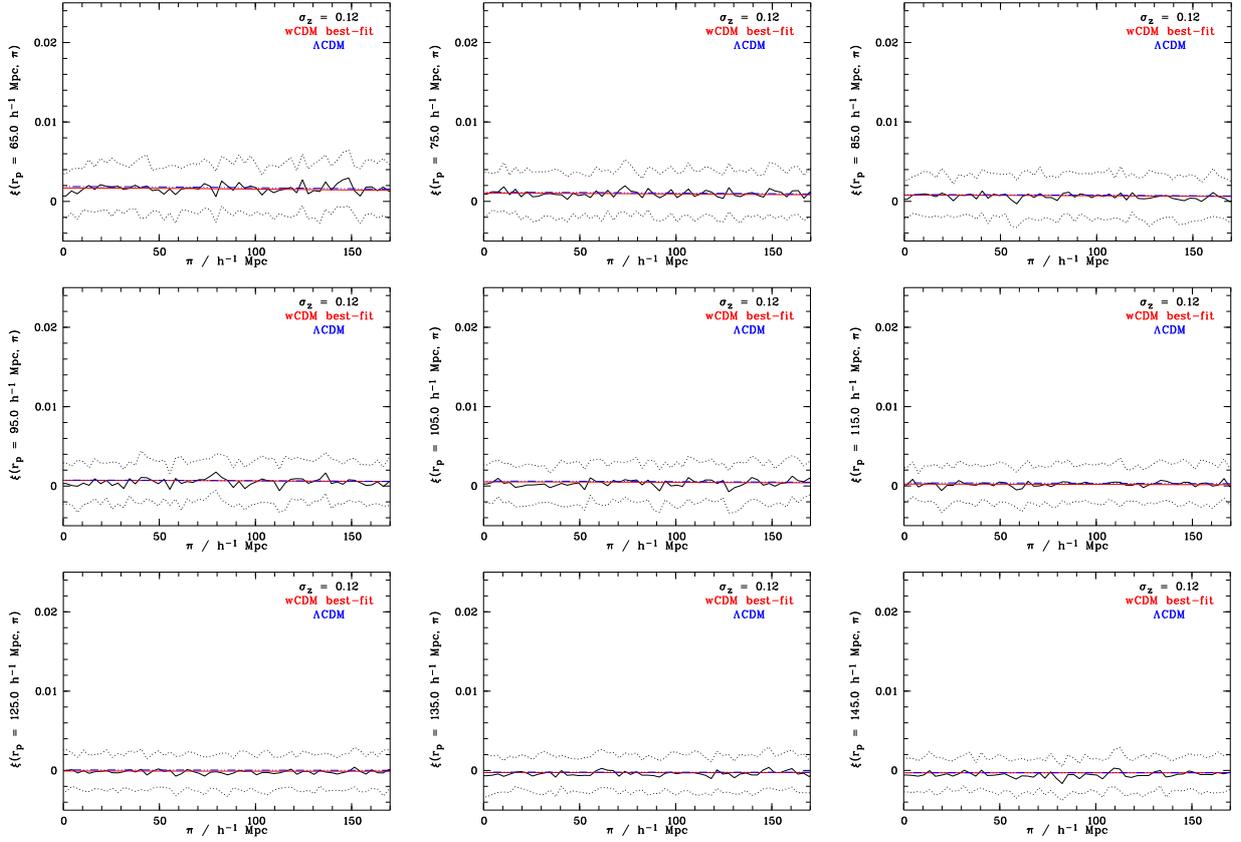
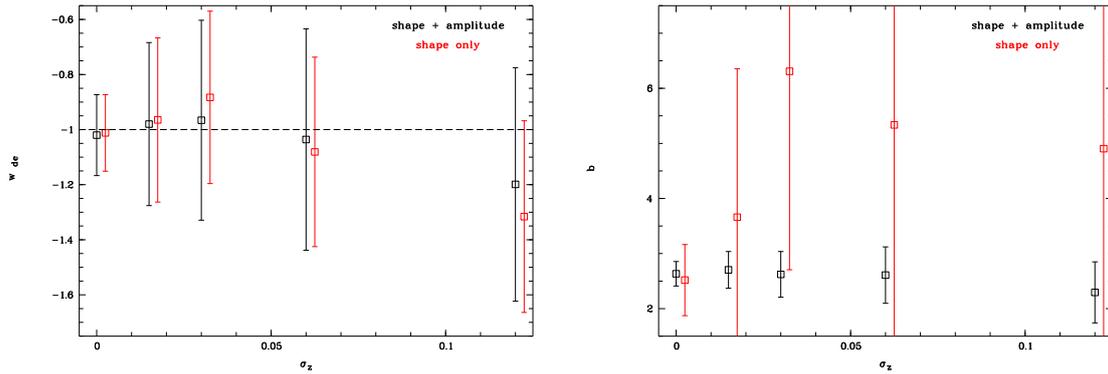


Figure 3.7: As in Figure 3.5, but for  $\sigma = 0.06$ .

Figure 3.8: As in Figure 3.5, but for  $\sigma_z = 0.12$ .Figure 3.9: Fitted values of the dark energy equation of state parameter  $w_{DE}$  (left) and the bias  $b$  (right) against the width of the redshift errors applied to the L-BASICC halos and the model. In black the amplitude and the shape of  $\xi(r_p, \pi)$  were taken into account for the fit, whereas in red only the shape was considered.

### 3.1.3 Projected two-point correlation function

For the analysis of the two-point correlation function on small scales often the projected correlation function has been calculated, which is in theory independent of any radial distortions (Peebles, 1980; Davis & Peebles, 1983). For small angles  $r^2 = r_p^2 + \pi^2$ . Thus the projected correlation function is defined as

$$w(r_p) = \int_{-\infty}^{\infty} \xi(r_p, \pi) d\pi. \quad (3.1)$$

Note that  $w(r_p)$  has dimensions of length. If it was possible in practice to integrate to infinity, it would in principle be possible to recover the three-dimensional real space correlation function  $\xi(r)$ , and  $w(r_p)$  would be far better suited to infer cosmological parameters from the spatial distribution of galaxies than  $\xi(r_p, \pi)$ , which suffers from redshift space distortions. However, since integrating out even to very large distances without significantly increasing the noise is not feasible (in particular if the signal is smeared out and the amplitude diminished by large redshift errors), the integration limits have to be finite and even rather small, see also Norberg et al. (2009) for an illustration of this. This means that a part of the clustering signal, which depends on the pairwise redshift probability distribution function as well as on the real and the assumed cosmology, can not be recovered. This is illustrated in Figure 3.10, where  $w(r_p)$  is shown for different widths of the assumed redshift errors ( $\sigma_z = 0.015, 0.03, 0.06, \text{ and } 0.12$ , respectively) and two different choices of the integration limits,  $\Delta\pi = 163.5 h^{-1} \text{ Mpc}$  and  $\Delta\pi = 298.5 h^{-1} \text{ Mpc}$ .

The resulting shape of  $w(r_p)$  depends strongly on both the width of redshift errors and the size of the integration limits: if the integration limits are very large, most of the signal can be recovered and the difference between real and redshift space and the correlation function affected by errors is, although small, still visible. It is not advisable to choose  $\Delta\pi = 298.5 h^{-1} \text{ Mpc}$ , as the measurement will be dominated by noise. On the other hand,

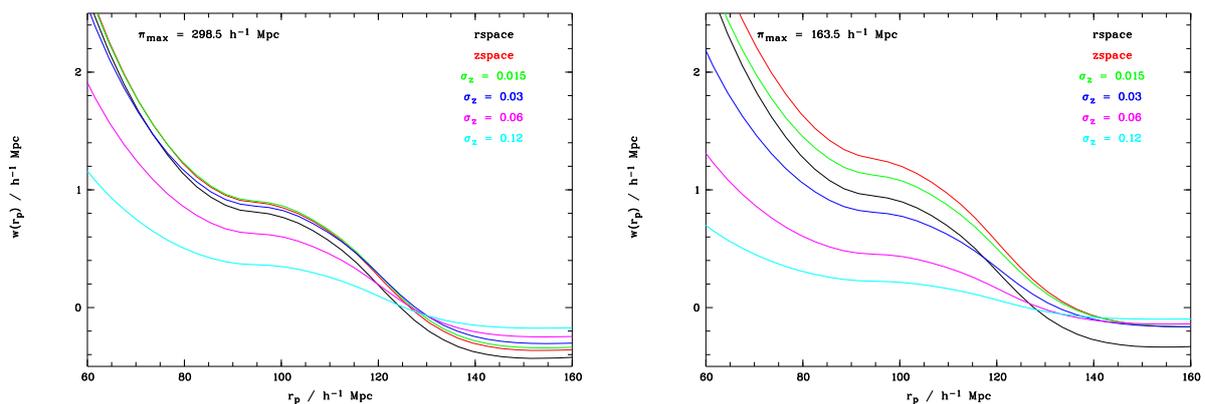


Figure 3.10: The projected correlation function  $w(r_p)$  for real and redshift space (black and red lines, respectively), and for different widths of the assumed redshift errors, left: for integration limits of  $\Delta\pi = 298.5 h^{-1} \text{ Mpc}$ ; right:  $\Delta\pi = 163.5 h^{-1} \text{ Mpc}$ .

if the BAO ring is supposed to be fully included in the integration, the limits cannot be much smaller than  $\Delta\pi \approx 150 h^{-1} \text{ Mpc}$  – in which case the resulting  $w(r_p)$  is extremely dependent on the size of the redshift errors (i.e. the fraction of the signal which can be recovered).

Figures 3.11 and 3.12 show the projected correlation function  $w(r_p)$  of the L-BASICC dark matter halos integrated up to  $\pi_{max} = 298.5 h^{-1} \text{ Mpc}$  and  $\pi_{max} = 163.5 h^{-1} \text{ Mpc}$ , respectively, for redshift errors of  $\sigma_z = 0.015$ ,  $\sigma_z = 0.03$ ,  $\sigma_z = 0.06$ , and  $\sigma_z = 0.12$ , as well as the best-fitting model for each case. The amplitude has to be taken into account, otherwise the fit fails: there is not enough information in the shape alone. The corresponding fitted values of the dark energy equation of state parameter  $w_{DE}$  and the bias  $b$  and their errors are listed in Table 3.3, and shown in Figure 3.13 as a function of  $\sigma_z$ .

Although the fit is not biased for  $\sigma_z \lesssim 0.06$ , the errors are, as expected, much larger and more quickly increasing with increasing redshift errors than for the corresponding fits of  $\xi(r_p, \pi)$ . Hence, it is concluded that for the analysis of the large scale two-point correlation function as a means to constrain cosmological parameters from photometric data,  $\xi(r_p, \pi)$  is better suited than the projected correlation function  $w(r_p)$ .

$\sigma_z$	$w_{DE}(\pi_{max} = 298.5 h^{-1} \text{ Mpc})$	$b(\pi_{max} = 298.5 h^{-1} \text{ Mpc})$
0.0	$-1.018 \pm 0.326$	$2.477 \pm 0.717$
0.015	$-1.018 \pm 0.419$	$2.683 \pm 0.551$
0.03	$-0.980 \pm 0.433$	$2.741 \pm 0.551$
0.06	$-1.017 \pm 0.437$	$2.725 \pm 0.608$
0.12	$-1.197 \pm 0.463$	$2.835 \pm 0.862$
$\sigma_z$	$w_{DE}(\pi_{max} = 163.5 h^{-1} \text{ Mpc})$	$b(\pi_{max} = 163.5 h^{-1} \text{ Mpc})$
0.0	$-1.078 \pm 0.386$	$2.683 \pm 0.470$
0.015	$-0.962 \pm 0.363$	$2.702 \pm 0.446$
0.03	$-0.979 \pm 0.372$	$2.706 \pm 0.456$
0.06	$-1.034 \pm 0.420$	$2.626 \pm 0.550$
0.12	$-1.143 \pm 0.576$	$2.764 \pm 0.804$

Table 3.3: The best-fitting values of  $w_{DE}$  and  $b$ , as deduced from the projected correlation function  $w(r_p)$ , with integration limits of  $\pi_{max} = 298.5 h^{-1} \text{ Mpc}$  (top rows) and  $\pi_{max} = 163.5 h^{-1} \text{ Mpc}$  (bottom rows), respectively, for Gaussian redshift errors with  $\sigma_z = 0.015$ ,  $\sigma_z = 0.03$ ,  $\sigma_z = 0.06$ , and  $\sigma_z = 0.12$ .

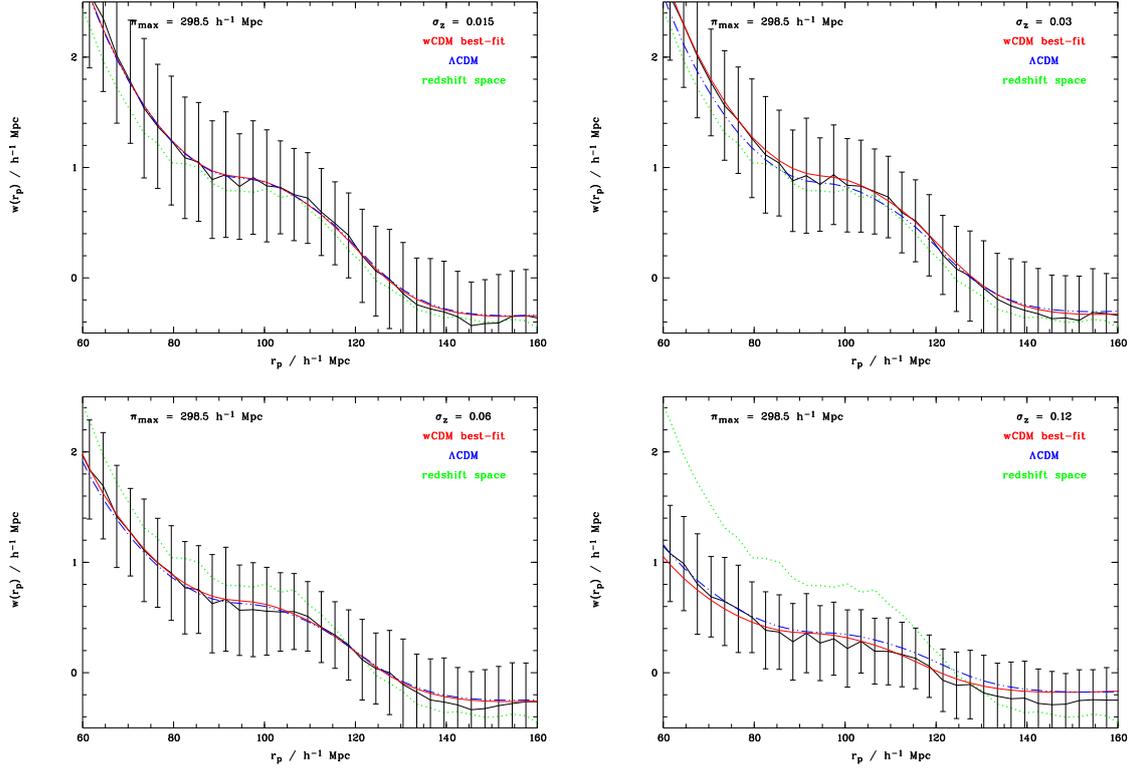


Figure 3.11: The projected correlation function  $w(r_p)$  of the L-BASICC dark matter halos integrated up to  $\pi_{max} = 298.5 h^{-1}$  Mpc for redshift errors of  $\sigma_z = 0.015$  (top right),  $\sigma_z = 0.03$  (top left),  $\sigma_z = 0.06$  (bottom right) and  $\sigma_z = 0.12$  (bottom left), black solid lines: mean, error bars:  $1\sigma$ -deviation calculated from the variance of the 50 boxes, red solid line: best-fitting wCDM model, blue dot-dot-dashed line:  $\Lambda$ CDM case, green dotted line: redshift space ( $\sigma_z = 0.00$ ).

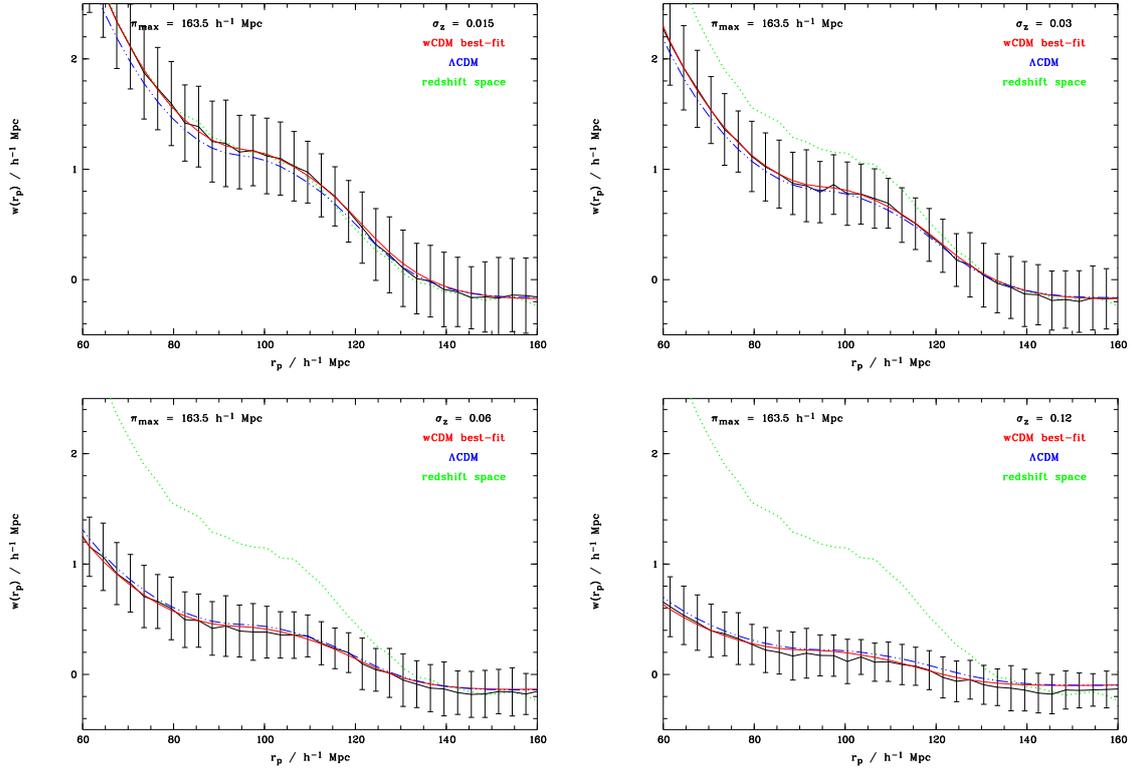


Figure 3.12: The projected correlation function  $w(r_p)$  of the L-BASICC dark matter halos integrated up to  $\pi_{max} = 163.5 h^{-1} \text{ Mpc}$  for redshift errors of  $\sigma_z = 0.015$  (top right),  $\sigma_z = 0.03$  (top left),  $\sigma_z = 0.06$  (bottom right) and  $\sigma_z = 0.12$  (bottom left), black solid lines: mean, error bars:  $1\sigma$ -deviation calculated from the variance of the 50 boxes, red solid line: best-fitting  $w\text{CDM}$  model, blue dot-dot-dashed line:  $\Lambda\text{CDM}$  case, green dotted line: redshift space ( $\sigma_z = 0.00$ ).

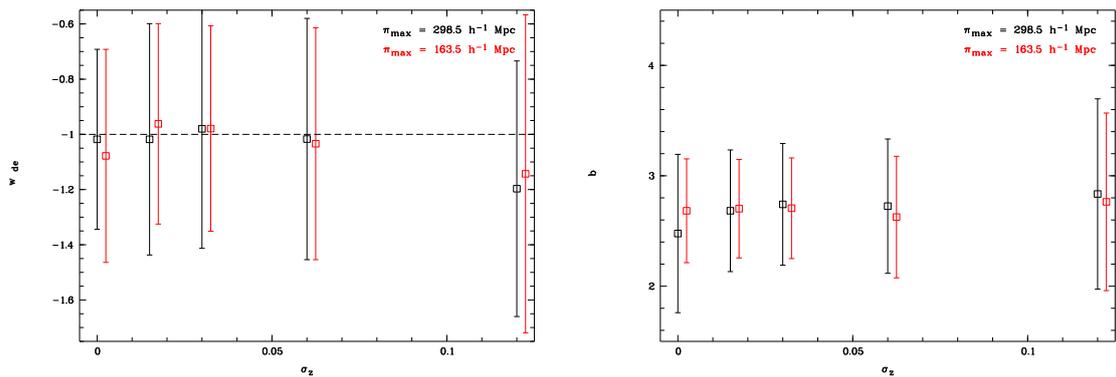


Figure 3.13: Fitted values of the dark energy equation of state parameter  $w_{DE}$  (left) and the bias  $b$  (right) against the width of the redshift errors applied to the L-BASICC halos and the model. In black  $w(r_p)$  was integrated up to  $\pi_{max} = 298.5 h^{-1} \text{ Mpc}$  for the fit, whereas in red  $\pi_{max} = 163.5 h^{-1} \text{ Mpc}$ .

### 3.1.4 Other work in the literature

In the last decade, substantial effort has been invested into the accurate modeling of the power spectrum or two-point correlation function of galaxies, in order to derive tight constraints on cosmological parameters. While the available and anticipated data sets have become larger, it has become clear that non-linear structure growth, peculiar velocities and galaxy biasing have to be described as precisely as possible. A variety of different methods has been developed and employed to tackle these problems, and the properties of different statistics have been investigated. Since the examination of this thesis (which is extracted from (Schlagenhauser et al., 2012)) is the first systematic investigation of the influence of redshift errors on the measurement of the dark energy equation of state parameter  $w_{DE}$  using  $\xi(r_p, \pi)$  (in Cai et al. (2009) the impact of photometric redshift errors on the *power spectrum* was investigated), a direct comparison with other work in the literature is not possible. Instead, the differences of the introduced  $\xi(r_p, \pi)$  here to existing models in which redshift errors can potentially be included will be highlighted.

In Gaztañaga et al. (2008),  $\xi(r_p, \pi)$  was modeled in order to estimate the position of the radial acoustic peak ( $\xi(r_p = 0 \text{ } h^{-1} \text{ Mpc}, \pi)$ ) and infer the bias  $b$  and  $\Omega_M$  from the SDSS DR6 LRG sample Adelman-McCarthy et al. (2008). For the modeling linear perturbation theory and linear redshift space distortions on large scales was used in Gaztañaga et al. (2008), whereas the model in this work is non-linear in both the description of structure growth and the Kaiser effect. While the FoGs are not included in the model (since they do not occur in a dark matter halo catalog without substructure, such as used to compare the model with), the one dimensional velocity dispersion  $\sigma_v$  is considered by them, in order to measure  $\beta = \Omega_M^\gamma/b$  from it, as discussed in detail in Cabré & Gaztañaga (2009).

A second study of  $\xi(r_p, \pi)$  was carried out by Okumura et al. (2008). Also in their approach both structure growth and Kaiser effect were treated linearly, but they take the wide angle effect (Szalay et al., 1998) and the high- $z$  distortion effect (Matsubara & Suto, 1996) into account, which are combined in Matsubara (2004). The scales examined were set to  $60.0 \text{ } h^{-1} \text{ Mpc} < s < 160.0 \text{ } h^{-1} \text{ Mpc}$ , in spherical shells like done in this thesis for real and redshift space, similar to the choice in this thesis. Fitting their model to the anisotropic two-point correlation function of the SDSS DR5 LRG sample (Eisenstein et al., 2005) not only  $w_{DE}$  and  $b$  were constrained but also several other cosmological parameters were determined. It remains to be tested if taking the wide angle effect into account improves the accuracy of the model in this work such that when the redshift errors become larger than the  $\sigma_z \gtrsim 0.06$  at which point the fit becomes extremely inaccurate, it is still possible to retrieve reliable constraints, but this analysis is beyond the scope of this thesis).

In Chuang et al. (2010),  $\xi(r_p, \pi)$  was modeled using halo fit (Smith et al., 2003) to take into account the non-linear structure growth. The redshift space distortions are included by the linear Kaiser effect on large scales and by the FoG on small scales. In Chuang et al. (2010),  $H(z)$  and  $D_A(z)$  was constrained instead of  $w_{DE}$ .

Instead of fitting an analytic model to observed data it is also possible to fit numerical  $N$ -body simulations (i.e. mocks, where the dark matter halos are populated with galaxies either using halo occupation modeling or a semi-analytic treatment of galaxy evolution,

to which the same selection function, mask and survey geometry has been applied). Then all kinds of clustering statistics can be calculated from the mocks and compared to the observed data. One example for such an attempt is the work of Kazin et al. (2010a) and Kazin et al. (2010b), although they did not measure  $\xi(r_p, \pi)$ , but studied the detectability of the BAO peak in  $\xi(r)$  as well as of the radial peak in SDSS LRG DR7 sample (Zehavi et al., 2005). The advantage of using a mock to compare the clustering statistics with is that non-linear clustering growth and redshift space distortions do not have to be modeled analytically, but occur naturally in the simulation. Also if many mocks are generated, the covariance matrix can be calculated. However, the big disadvantage is of course that a proper fit which takes differences in the non-linear clustering growth due to different values of  $w_{DE}$  into account requires to cover the full parameter space with a large number of mocks, which is extremely time consuming.

In Padmanabhan & White (2008), it was suggested to utilize the multipole moments for estimating cosmological parameters. Although this requires a smaller number of mock catalogs to calculate the covariance matrix than would be needed for  $\xi(r_p, \pi)$ , it can not be used in the presence of photometric redshift errors.

In Kazin et al. (2011), the analysis of the multipole expansion by Padmanabhan & White (2008) was transformed from Fourier space to configuration space. They compared the multipole expansion including the hexadecapole with so-called "clustering wedges" ( $\xi(\Delta_\mu, s)$ , where  $\mu = s_{||}/s$  and  $s_{||}$  is the radial component of separation  $s$ ) to constrain  $H$  and  $D_A$ , in order to break the degeneracy between these two parameters usually found when only using the monopole. The "clustering wedges" are able to provide constraints at least at the same level of accuracy as the multipole expansion. Their fit is based on  $N$ -body simulations, but they also argue that for practical use an analytic modeling based on physical principles is needed.

Although the model introduced in this thesis includes neither the large angle effect, nor Fingers of God (which are not present in the data on which it was tested against), since non-linear clustering growth and non-linear velocities are taken into account, and it possible to fit the linear bias by making use of the information contained in the redshift space distortions, it seems to be competitive. Also redshift errors can easily be included, which makes it a valuable tool to apply to photometric data, if the redshift errors are small enough ( $\sigma_z \lesssim 0.06$ ).

## 3.2 Bispectrum

After the theoretical derivation of the three-point statistics in real and redshift space, the model has to be tested against e.g. the L-BASICC simulations. This part of the thesis will be focused on the Fourier space quantity, the bispectrum. Such an investigation is required to find out the range of validity of the model and its limitations. Throughout this section, such an examination will be performed in detail, beginning with a special triangle configuration, the equilateral one. In Section 3.2.1, the problems the model has will be described in order to understand the results in the following sections. In Section 3.2.3, this investigation will be extended to all possible triangle configurations up to a predefined  $k$ -mode. Finally, Section 3.2 will be concluded with a discussion of other publications (see Section 3.2.5). The whole analysis of the bispectrum will be carried out with the rms of the 50 realizations of the simulation and not the variation of the mean. This procedure was chosen to get a better understanding of the errors which can be assumed in a real survey. As a reminder, the volume of one box is  $V_{\text{box}} = 2.41 h^{-3} \text{Gpc}^3$ , whereas the whole available volume is fifty times larger, namely  $120.31 h^{-3} \text{Gpc}^3$ , which is completely unrealistic for the size of a galaxy survey. For example the effective volume of the SDSS III DR 9 BOSS CMASS sample is given by  $2.2 \text{Gpc}^3$  (Anderson et al., 2012) which is almost the size of one L-BASICC box.

### 3.2.1 Equilateral configurations: dark matter

The equilateral triangle configuration can be described by only one  $k$ -mode, because all sides have the same length. This is a big advantage of equilateral configurations. If an effect is scale dependent or the theory breaks down at a certain  $k$ -mode, there will not be any signal from other  $k$ -modes. Therefore, a comparison of the dark matter simulation with the theoretical prediction is much easier and will clearly show at which  $k$ -mode the non-linear structure growth cannot be modeled correctly by the tree-level ansatz anymore. In Figure 3.14, the dark matter bispectra from the L-BASICC simulations for equilateral triangles are shown at redshifts of  $z = 0.5$  (left panel) and  $z = 1.0$  (right panel). The binsize for the extracted bispectra from the L-BASICC simulations is chosen to be  $2 \times k_f$  which will also be used for all other bispectra measurements. The results for the  $z = 0.0$  case will be given in the following section where all configurations are taken into account. For each redshift the real (blue color) and redshift (red color) space are plotted together in one panel where the solid lines represent the theory and the squares with the error bars the estimation of the bispectra from the dark matter L-BASICC simulations.

It can easily be seen, that for both cases in real space the dark matter estimations and the theoretical modeling start to deviate at around a  $k \approx 0.15 h \text{Mpc}^{-1}$ . If the ratio between the L-BASICC dark matter estimation of the bispectrum for equilateral triangles and the model (see Figure 3.15) is computed, this discrepancy can be identified more clearly. The model underestimates the simulation bispectrum for  $z = 0.5$  at  $k \approx 0.10 h \text{Mpc}^{-1}$ , whereas for  $z = 1.0$  the situation remains unchanged and a consistent and unbiased description of the model can be achieved up to  $k \approx 0.15 h \text{Mpc}^{-1}$ .

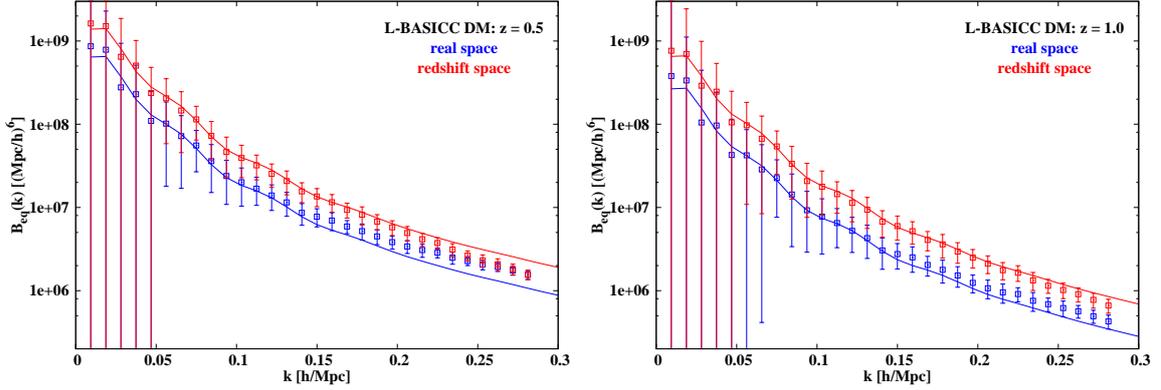


Figure 3.14: The equilateral dark matter bispectrum  $B_{eq}(k)$  at  $z = 0.5$  and  $z = 1.0$  are given on the left and right panel, respectively. The real space bispectrum is shown in blue and redshift space in red, where the theory is given by the solid lines and the L-BASICC estimation by the squares with the error bars.

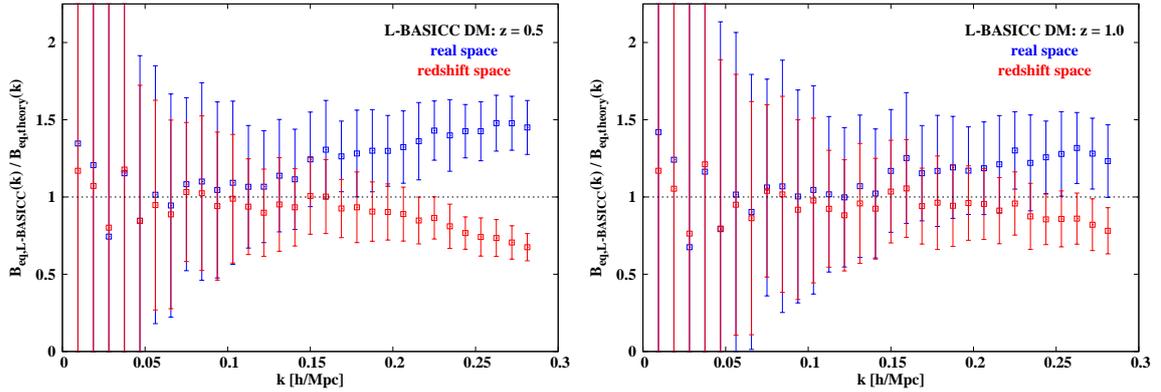


Figure 3.15: Ratio of  $B_{eq}(k)$  between the theory and the extracted quantity from the dark matter L-BASICC simulations, where  $z = 0.5$  can be found on the left panel and  $z = 1.0$  on the right panel. The real space ratio is indicated in blue and redshift space in red.

In redshift space, the theory overpredicts the signal and deviations occur on larger  $k$ -modes. This can be noticed by looking at Figure 3.14. This statement is true for  $z = 0.5$  where the ratio is around unity up to  $0.15 h \text{ Mpc}^{-1}$ . After that  $k$ -mode the ratio is biased towards values smaller than unity. For  $z = 1.0$  the situation remains unchanged up to  $0.15 h \text{ Mpc}^{-1}$  compared to real space. Afterwards, the ratio is lower than unity. In the following paragraph, these findings will be discussed in detail.

### Explanations of the results of the dark matter $B_{eq}(k)$

The deviations identified in the previous discussion are caused by the limitations of the theoretical model. Two possible explanations can be given:

- non-linear structure growth and
- random motions within a halo (also called FoG).

So far, the modeling of the structure growth is valid in the weakly non-linear regime. By extending the analysis to too high  $k$ -modes the non-linearities induced by gravity become more and more important but are not described sufficiently by the theory anymore.

In real space, it can be assumed that this is the only effect which causes deviations. This statement can be made clearer by Figure 3.15. At a certain  $k$ -mode, the real space ratios are above the unity line which means the model predicts a too low amplitude. There are no other effects present in real space (e.g. biasing or redshift space distortions) for the direct comparison performed in this section.

However, it could also be argued that the dark matter particles in the simulations are tracers of the underlying matter density field. This could cause deviations from the dark matter expectation values ( $b_1 = 1.0$  and  $b_2 = 0.0$ ). But this cannot be investigated in more detail because higher resolution  $N$ -body simulations would be required which were not available for this thesis.

Of course, in redshift space also redshift space distortions have to be considered. The signal of the redshift space bispectrum from the L-BASICC simulations is reduced. From the above statements one would expect the opposite trend because the non-linearities of the structure growth are also present in redshift space. However, the random motions of dark matter particles within a gravitationally bound system (a halo) are not taken into account. These random motions would act as a smoothing at a certain scale which depends on the strength of the random motions. This leads to a damping of the signal as expected from the theory section. The same effect can be seen in the power spectrum (Matsubara, 2008) where only two  $k$ -modes are involved. For the bispectrum three  $k$ -modes are required. Therefore, the damping effect of the FoG should be noticed at comparably smaller  $k$ -modes. This effect can be noticed more directly by comparing the redshift space boost-factor obtained from the theory with the one of dark matter L-BASICC estimation. In Figure 3.16, where the ratio of the redshift space and real space equilateral bispectra is shown, a scale dependence of the redshift space distortions can be identified clearly at  $k \gtrsim 0.10 h \text{ Mpc}^{-1}$  for both cases. This trend is not related to non-linear structure growth. Due to the calculation of the ratios no modeling is involved, non-linearities are present in real and redshift space. The vertical line in this figure shows the theoretical Kaiser boost-factor whereas the L-BASICC Kaiser boost-factor seems to decrease at larger  $k$ -modes or, in other words, the signal of the redshift space bispectra are damped. This damping is a clear hint towards the FoG.

In the absence of random motions the independence of the theoretical boost-factor on the  $k$ -mode is exactly what one would expect in the case of equilateral triangles. The angle

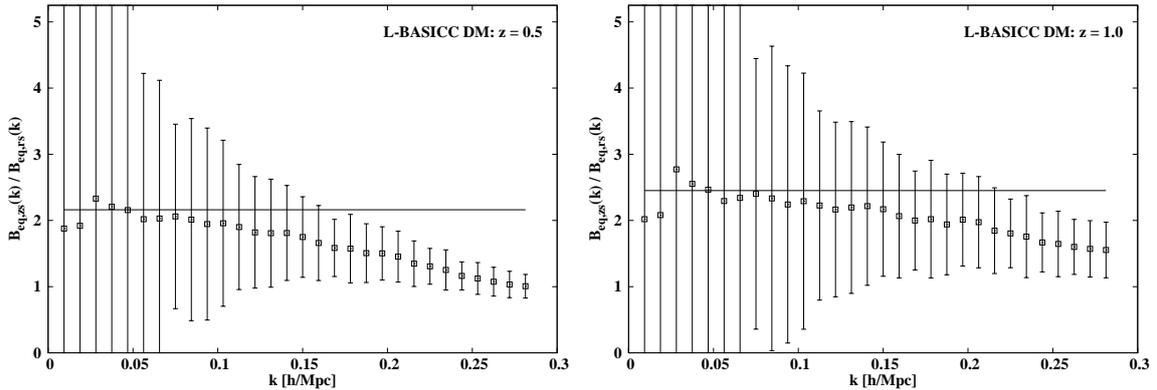


Figure 3.16: Ratio between the redshift ( $B_{eq,zs}(k)$ ) and real space dark matter bispectra ( $B_{eq,rs}(k)$ ) for  $z = 0.5$  and  $z = 1.0$  are plotted on the left and right panel, respectively. The vertical line indicates the theoretical expectation value, the Kaiser boost-factor whereas the squares with the error bars are the L-BASICC ratios.

averaging of the redshift space bispectrum in Equation (2.98) affects all sides of the triangle in the same way. If more general configurations are taken into account, the boost-factor will be dependent on the considered configurations. The sides of the triangle have different lengths. Therefore, they are affected differently by redshift space distortions.

It can be concluded that the non-linearities can be correctly modeled up to  $k \approx 0.10 h \text{ Mpc}^{-1}$  for all considered redshifts. The only effect not taken into account so far are the above mentioned random motions. They would introduce a  $k$ -dependence due to their increasing importance at larger  $k$ -modes and results in an additional damping which is able to explain the difference in the ratios. This means that without a proper modeling of this effect, modes larger than  $0.10 h \text{ Mpc}^{-1}$  cannot be taken into account for the bias estimation at low redshifts which is one of the goals of this thesis. Modeling the FoG will be examined in Section 3.2.3.

### Bias parameter extraction: dark matter

The validity of the model cannot only be tested by a direct comparison but it can also be checked by extracting the bias parameters and compare the results with the expectation values. By performing a MCMC (Metropolis et al., 1953; Spergel et al., 2003) the linear bias  $b_1$  and the quadratic bias  $b_2$  can be extracted from the L-BASICC simulations. In this case, the expectation values are already known,  $b_1 = 1.0$  and  $b_2 = 0.0$ .

The MCMC was carried out with  $0.0 \leq b_1 \leq 5.0$  and  $-20.0 \leq b_2 \leq 20.0$  as flat priors. The exploration of the parameter space is performed with a so-called jumping function, which is assumed to be a Gaussian (Spergel et al., 2003). In this analysis, the rms of this jumping function was set to  $\sigma_{b_1} = 0.3$  and  $\sigma_{b_2} = 5.0$ . In Table 3.4, the results of the fitting procedure, where the maximal  $k$ -mode for the fit is set to be  $0.10 h \text{ Mpc}^{-1}$ , are given.

Due to the large errors for the extracted bias parameters further conclusive statements on these results are not possible. In Section 3.2.3, the uncertainties of  $b_1$  and  $b_2$  become

smaller because more configurations will be used throughout the extraction process.

$z_{obs}$		$b_1$	$b_2$
0.5	rs	$1.106^{+0.643}_{-0.736}$	$1.017^{+2.259}_{-1.887}$
	zs	$0.907^{+0.386}_{-0.435}$	$1.429^{+2.048}_{-1.868}$
1.0	rs	$0.886^{+0.530}_{-0.528}$	$1.221^{+2.085}_{-1.809}$
	zs	$1.069^{+1.023}_{-1.011}$	$4.824^{+7.787}_{-6.220}$

Table 3.4: The estimated bias parameters  $b_1$  (column 3) and  $b_2$  (column 4) from the dark matter  $B_{eq}(k)$  with their corresponding one-sigma confidence levels for real (rs, in column 2) and redshift space (zs, in column 2) and the investigated redshift in column 1.

Under the assumption that the difference in Figure 3.16 for  $k$ -modes smaller than  $0.10 h \text{ Mpc}^{-1}$  is generated by random motions it would be possible to account for this effect. As mentioned above, an additional damping is needed. However, such an additional damping at such large scales would also result in a much stronger damping on smaller  $k$ -modes. Therefore, a smaller amplitude for the L-BASICC bispectrum should be measured. The deviation between the model and the L-BASICC boost-factor should be present at smaller  $k$ -modes compared to the situation now.

Then, the question arises, why the redshift space equilateral bispectrum is in better agreement with the model than in real space. There must be present an additional effect in redshift space, which is not properly taken into account in the model. Any additional boost would be damped by the FoG. The non-linear Kaiser effect (Scoccimarro, 2004) could provide such a boost in the amplitude. However, this effect would be more pronounced if random motions would not be present. For the dark matter halos, this would be the case, which will be examined in the next paragraph.

The velocity dispersion  $\sigma_v$  of the particles, which define the FoG, could be set as an extra fitting parameter. By this damping, the overshooting of the non-linear Kaiser effect compared to the modeling with the linear Kaiser effect could be corrected for to some extent. This kind of analysis will not be performed in this section due to the few data points (limited number of different triangles) which can be used for the MCMC. Adding more fitting parameters do not help to extract well determined bias parameters. Therefore, the effect of random motions on the bias parameters estimation will be examined again in Section 3.2.3 where all triangle configurations are taken into account.

### 3.2.2 Equilateral configurations: dark matter halos

The effect of random motions can also be examined by looking at the equilateral bispectrum for the L-BASICC halo catalogs. In this case, the small scale random motions are not present anymore because the FoF-halos are just points in the L-BASICC halo catalogs. From the discussion in Section 2.6.2 it is clear, that substructure is required for the FoG effect. By the lack of substructure, the FoGs do not affect the halo bispectrum in redshift

space anymore. If this assumption is correct, the same examination as for the dark matter particles will give more consistent results with the theoretical prediction.

The redshift space boost-factor comparison for the halos requires a fitting procedure first, where the fitting range was set to be  $k = 0.10 h \text{ Mpc}^{-1}$  for both redshifts. For the analysis of biased objects the flat priors of the MCMC are set to be  $0.0 \leq b_1 \leq 8.0$  and  $-20.0 \leq b_2 \leq 30.0$  and also the searching area in the parameter space was extended to  $\sigma_{b_1} = 0.5$  and  $\sigma_{b_2} = 3.0$ . These parameter ranges are used for both redshifts.

The results of this analysis are given in Table 3.5. The  $b_1$  parameter is weakly determined but it can be noticed that larger values for  $b_1$  can be extracted at higher redshifts with a one-sigma significance. The quadratic bias is still undetermined. A detailed discussion of estimating  $b_1$  and  $b_2$  will be performed in Section 3.2.3 where all triangle configurations are considered. This will help to obtain tighter constraints for  $b_1$  and  $b_2$ .

$z_{obs}$		$b_1$	$b_2$
0.5	rs	$1.892^{+0.712}_{-0.686}$	$5.598^{+5.538}_{-4.994}$
	zs	$1.943^{+0.736}_{-0.706}$	$5.425^{+5.596}_{-4.878}$
1.0	rs	$3.056^{+1.011}_{-0.871}$	$6.288^{+5.308}_{-4.951}$
	zs	$3.151^{+1.069}_{-0.913}$	$6.292^{+5.421}_{-5.074}$

Table 3.5: The estimated bias parameters  $b_1$  (column 3) and  $b_2$  (column 4) from the dark matter halo  $B_{eq}(k)$  with their corresponding one-sigma confidence levels for real (rs, in column 2) and redshift space (zs, in column 2) and the investigated redshift in column 1.

The comparison between the estimation of the equilateral bispectra in real (blue squares) and redshift (red squares) space as well as the corresponding theoretical counterpart is shown in Figure 3.17. The divergence between the estimation and the theory starts at around  $k \approx 0.10 h \text{ Mpc}^{-1}$ , which is almost the same value as for the dark matter case and hence has been chosen to be the end of the fitting range. For  $z = 0.5$  the situation seems to be better than for  $z = 1.0$ , even if the real space bispectrum is overestimated by the theory.

Assuming that the bias parameters are scale independent, the physical effects are the same on all scales. No additional scale dependence is introduced as long as the considered  $k$ -modes represent large scales (at least  $\gtrsim 60.0 h^{-1} \text{ Mpc}$ , see Section 3.1). This can be noticed for real space where the agreement between theory and measurement is extended up to  $k \approx 0.15 h \text{ Mpc}^{-1}$ . At smaller scales, differences between the L-BASICC and the theoretical bispectrum can be identified. This deviation is present in real and redshift space. The theoretical redshift space bispectrum for the halos underestimates the L-BASICC estimation. This is a contrary trend compared to the dark matter case. All of these statements can also be derived from Figure 3.18 where the ratios of the L-BASICC and the theoretical bispectrum are shown for the two considered redshifts.

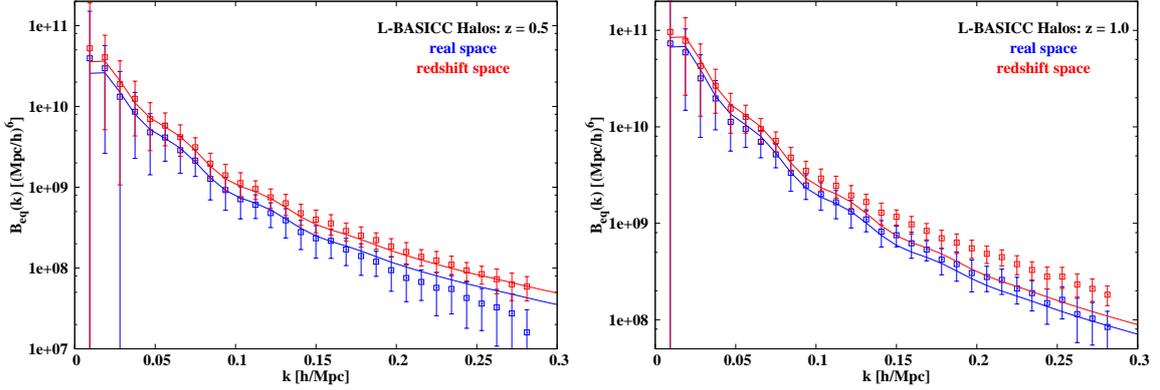


Figure 3.17:  $B_{eq}^h(k)$  at  $z = 0.5$  and  $z = 1.0$  are plotted on the left and the right panel, respectively. The solid lines show the theoretical model and the squares with the error bars the L-BASICC bispectra. Real space is illustrated in blue and redshift space in red.

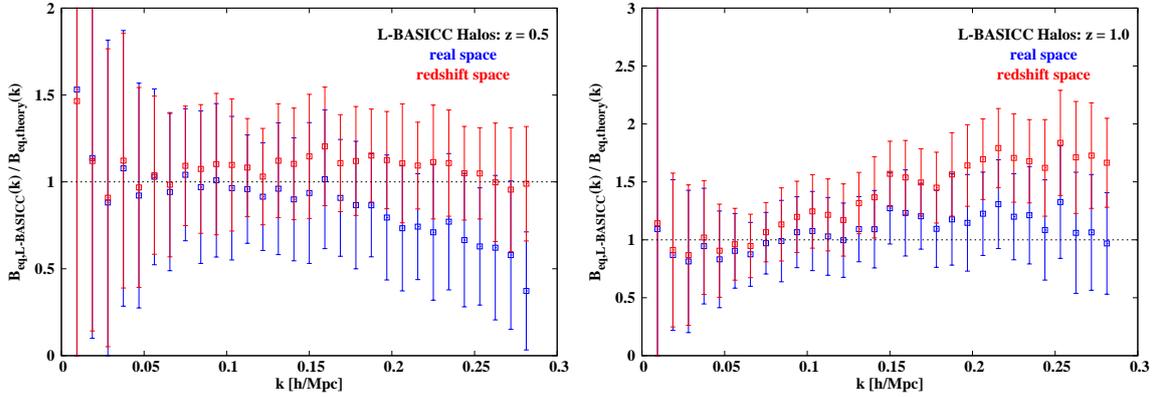


Figure 3.18: Ratio between the theory and the estimation from the L-BASICC halo catalogs ( $B_{eq,L-BASICC}(k)/B_{eq,theory}(k)$ ):  $z = 0.5$  (left panel) and  $z = 1.0$  (right panel). Real space is given by the blue color and redshift space by red.

### Explanations of the results of the dark matter halo $B_{eq}(k)$

First of all, one can have the impression that the low redshift estimation is better described by the theory than the high redshift case. In order to understand this, it is necessary to introduce the mass function.

The mass function is a statistics which describes the comoving number density of bound objects, such as halos, of mass  $M$  at redshift  $z$  (Sheth & Tormen, 1999). It is a rapidly declining function when going from low to high mass halos. The probability of finding a high mass halo is higher at lower redshifts than at higher redshifts because particles had more time to organize themselves in bound objects by following the local gravitational potential.

The mass of a dark matter particle for the L-BASICC simulations is  $M_{dm} = 1.75 \times 10^{12} h^{-1} M_{\odot}$  and the smallest halos consist of 10 dark matter particles, see Section 2.3. There-

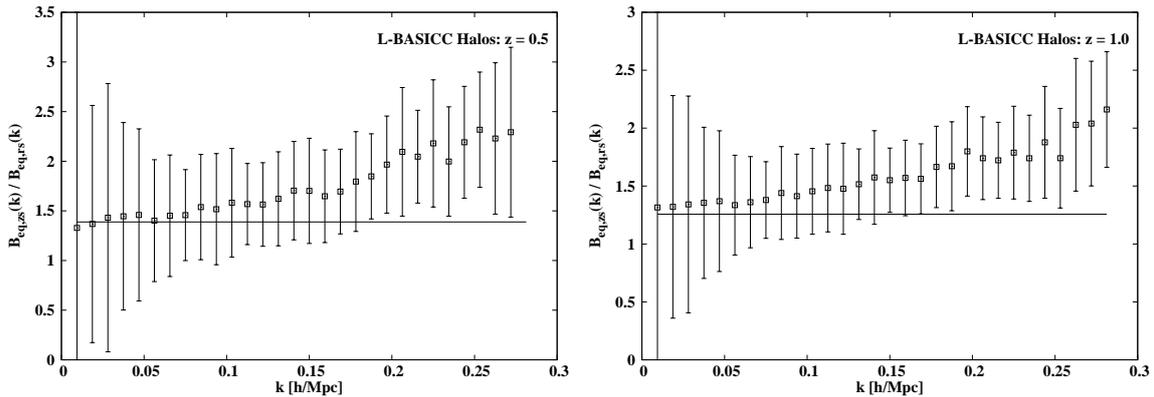


Figure 3.19: Ratio between the redshift and real space bispectra  $B_{eq,zs}(k)/B_{eq,rs}(k)$  (the Kaiser boost-factor) for  $z = 0.5$  and  $z = 1.0$  are plotted on the left and right panel, respectively. The vertical line indicates the theoretical expectation value whereas the squares with the error bars are the L-BASICC ratios.

fore, the mass of the smallest halos is given by  $M_{halo}^{min} = 1.75 \times 10^{13} h^{-1} M_{\odot}$ . This also sets the mass scale from which the mass function of the L-BASICC simulations can be examined. At a redshift of  $z = 1.0$ , halos with a mass of  $M_{halo}^{min}$  are already very massive with an already very non-linear structure growth history compared to lower redshifts. If this point is considered in the discussion, then the measured bispectrum will look much more non-linearly evolved. The underestimation of the theory is a good indication that the non-linearities are not described sufficiently for  $k$ -modes larger than  $0.15 h \text{ Mpc}^{-1}$  in real space and  $0.10 h \text{ Mpc}^{-1}$  in redshift space. The same discussion and statements can be made by looking at Figure 3.18. It can be noticed that the ratio for  $z = 0.5$  is closer to unity than for the  $z = 1.0$  case. This is not surprising from the discussion above.

A closer look at the redshift space bispectrum is required to explain the deviation between the L-BASICC and the theoretical bispectrum. The signal of the redshift space halo L-BASICC bispectrum is not suppressed by the FoG due to the lack of the substructure of the dark matter halos. Only the effect of non-linear structure growth and large scale redshift space distortions are left.

For a more clearer discussion, the ratio between the redshift and real space bispectrum  $B_{eq,zs}(k)$  and  $B_{eq,rs}(k)$ , respectively, is plotted in Figure 3.19. The same information is illustrated as in Figure 3.16 but for the L-BASICC dark matter halos. As for dark matter, the goal is to understand how well large scale redshift space distortions are described by the linear Kaiser effect (Heavens et al., 1998; Verde et al., 2002).

The damping, which is visible in Figure 3.16 (dark matter case), is not present in Figure 3.19 (dark matter halos) but a clear trend of overestimating the signal can be seen. The redshift space bispectrum seems to gain signal at larger  $k$ -modes. One explanation would be a more non-linear behavior, which is not taken into account by the linear redshift space distortions. The non-linear Kaiser effect (Scoccimarro, 2004), where the infall of matter is not considered as linear as it is assumed by the linear Kaiser effect (Kaiser, 1987), could

provide such a boost of the amplitude of the redshift space bispectrum. Another explanation would be a scale dependent bias which would lead to different behaviors in real and redshift space. So far, scale dependent bias parameters have not been observed because they should also alter the two-point statistics (Jeong & Komatsu, 2009). A scale dependent bias would have also changed the results of Section 3.1. The non-linear Kaiser effect is the best possible explanation of the ratios in Figure 3.19.

### Reduced bispectrum: dark matter

Until now, only the bispectrum  $B_{eq}(k)$  was examined throughout the whole discussion in this section. The reduced bispectrum  $Q(k_1, k_2, \theta)$  is also a quantity on which such a discussion can be performed to determine the maximum possible  $k$ -mode which can be taken into account for the estimation of the bias parameters. As mentioned in Section 2.6.1,  $Q(k_1, k_2, \theta)$  has the big advantage, that it is in lowest order independent on cosmological parameters and time evolution. For the equilateral triangle configuration Equation (2.102) is simplified to

$$Q_{red,eq}(k) \equiv \frac{B_{eq}(k)}{3(P(k))^2}. \quad (3.2)$$

The reduced equilateral bispectra  $Q_{red,eq}(k)$  for  $z = 0.5$  (left panel) and  $z = 1.0$  (right panel), which are shown in Figure 3.20 are clearly not very sensitive to the time evolution. This behavior can be explained by the division of the bispectrum by the power spectra. The growth factors contained in the bispectra are almost canceled out by the growth factors in the power spectra. In Figure 3.20, the theoretical  $Q_{red,eq}(k)$ s for real and redshift space are represented by a blue and a red straight line, respectively. These values can be calculated easily and are given by  $Q_{red,eq,rs}(k) = 4/7$  and

$$Q_{red,eq,zs}(k) = \frac{2}{105} (30 + 20\beta(z) + 3\beta(z)) + \frac{b_1^3\beta(70 + 42\beta(z) + 3\beta(z)^2)}{1470} - \frac{\beta(z)}{315} (-105 - 63\beta(z) + \beta(z)^3) + \frac{b_1^3 b_2}{30} (30 + 20\beta(z) + 3\beta(z)^3) \quad (3.3)$$

for real and redshift space, respectively. The F-kernel which describes the shape of the bispectrum (this is absolutely true for dark matter particles but not for biased objects, see the discussion in Section 2.6.1) is dependent on the ratios of the involved  $k$ -modes and the angles between them.

For equilateral triangle configurations this ratio is unity and the angle is  $\pi/3$ . Then, the F-kernel is just a single number independent on the size of the considered triangles. Of course, this statement is true for any other triangle configuration as long as the ratios and the angles are kept constant, just the number of the F-kernel changes. If the bispectrum is evaluated for two predefined side lengths  $k_1$  and  $k_2$  where  $\theta_{12}$  is defined as the angle between these  $k$ -modes to ensure the closure of the triangle, different configurations will be mixed in one representation. This will lead to a U-shaped curve as for  $B_{eq}(k)$  (Bernardeau et al., 2002).

The deviations which can be identified in Figure 3.20 start at around  $0.10 h \text{ Mpc}^{-1}$  and

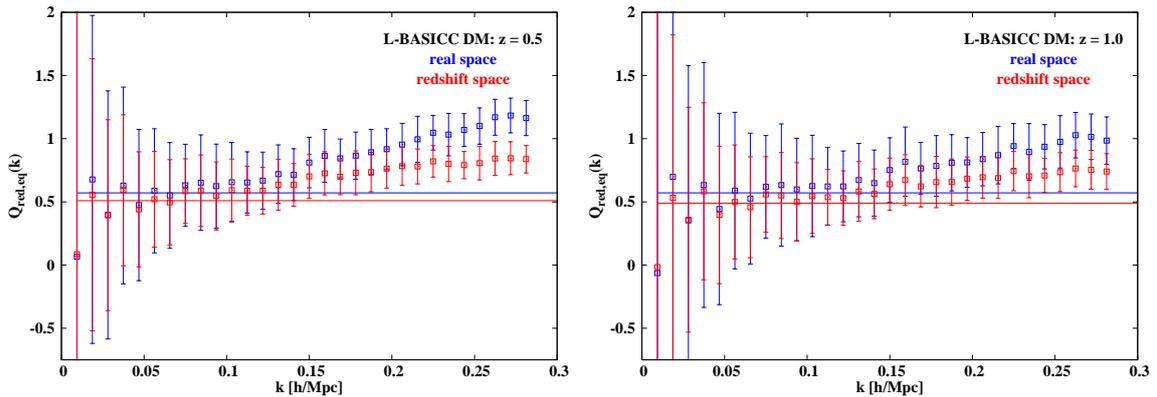


Figure 3.20:  $Q_{eq}(k)$   $z = 0.5$  (left panel) and  $z = 1.0$  (right panel) are shown in blue and red for real and redshift space, respectively. The straight lines are the theoretical  $Q_{eq}(k)$ s and the squares with the error bars are the  $Q_{eq}(k)$ s extracted from the L-BASICC dark matter halo catalogs.

$0.15 h \text{ Mpc}^{-1}$  for  $z = 0.5$  and  $1.0$ , respectively. The same values for the  $k$ -modes up to which the model is able to deliver consistent results with the L-BASICC simulations were found for  $B_{eq}(k)$ . So far, this shows that the modeling is consistent and that the statements made above also hold in this situation.

Non-linearities are responsible for the deviations from the straight line which would be expected by the tree-level ansatz. This also shows that non-linearities are more pronounced in the bispectrum than in the power spectrum.

### Reduced bispectrum: dark matter halos

For biased objects the reduced equilateral bispectrum

$$Q_{red,eq}^h(k) = \frac{Q_{red,eq}(k)}{b_1} + \frac{b_2}{b_1^2}, \quad (3.4)$$

which looks similar to Equation (2.103), is also a featureless function at small  $k$ -modes (at least for  $\lesssim 0.10 h \text{ Mpc}^{-1}$ , see the dark matter discussion in the previous section). The trend, which is visible in Figure 3.20, is not that obvious anymore.

Any signal present in the dark matter halo reduced bispectrum will have a lower significance due to the reduction of the amplitude by the  $b_1$  bias parameter and the increase of the errors due to the low mean density of the halo catalogs. This means that the fitting procedure for  $B_{eq}(k)$  should not be applied to  $Q_{red,eq}(k)$  because at small  $k$ -modes it would be fitting a straight horizontal featureless line with two parameters. This will lead to a huge degeneracy between  $b_1$  and  $b_2$  and lowers the significance of such a fitting process. In addition to the already noticed low significance of the estimated bias parameters from  $B_{eq}(k)$ . Therefore, extracting the bias parameters from  $Q_{eq}(k)$  will be excluded from this thesis. A figure with the reduced equilateral bispectra for the dark matter halos for the

two redshifts would also not make any sense. It would be hardly possible to see at which  $k$ -mode deviations between the estimation and the theory occur.

However, the reduced bispectrum obtained from all configurations will be analyzed in detail in Section 3.2.3. In this case, there will be a shape dependence present due to the mixing of different triangle configurations. At this stage it is known, that the model can be trusted up to a certain  $k$ -mode ( $\lesssim 0.10 h \text{ Mpc}^{-1}$  for  $z = 0.5$  and  $\lesssim 0.15 h \text{ Mpc}^{-1}$  for  $z = 1.0$ ).

### 3.2.3 All configurations

There are two possible methods of estimating the bispectrum. One results in the so-called U-shape, where one specific triangle configuration is of interest. This means, two sides of the triangle, e.g.  $k_1$  and  $k_2$ , are fixed during the estimation process and the third side  $k_3$  is varied with respect to the angle  $\theta_{12}$  between  $k_1$  and  $k_2$ . The second one is to consider all possible configurations up to a predefined  $k$ -mode where the values of the bispectrum are ordered in the following way,  $k_{max} = k_1 \geq k_2 \geq k_3$  under the restriction of the binsize, which is set to be  $2 \times k_f$  throughout the following discussion. This would result in a zick-zack-curve. Taking all possible configurations into account gives a much larger content of information compared to a single U-shape curve, where only one configuration is considered, and hence the bias parameters  $b_1$  and  $b_2$  are more constrained.

Therefore, the zick-zack curve should be chosen for extracting these bias parameters. The proposed order of the triangles has the advantage that if the range of validity of the theoretical model is known, all  $k$ -modes larger than that particular  $k$ -mode will not be taken into account for the estimation of  $b_1$  and  $b_2$ . Indeed, the same procedure is possible for the U-shape representation, but this includes more computational effort and because of throwing-away information, the constraints would be even less restricted.

Therefore, in the following discussions all possible configurations in the above suggested order, except for collapsed configurations ( $k_1 \approx k_2 \gg k_3$ ), are taken into account if not stated otherwise. At the beginning of the dark matter discussion, the influence of the collapsed triangles will shortly be discussed. A more detailed investigation is given in the Appendix A.5, in which the bias parameters  $b_1$  and  $b_2$  will be estimated by taking also the collapsed configurations into account. It will be shown that these estimates for  $b_1$  and  $b_2$  are not correct because the model is not able to predict the bispectrum correctly for these configurations.

#### Bias parameter extraction: dark matter bispectrum

As in Section 3.2.1, the discussion starts with the analysis on the dark matter simulations in order to show, up to which  $k_{max}$  unbiased estimations of the linear bias  $b_1$  and quadratic bias  $b_2$  can be achieved and whether the restrictions made in Section 3.2.1 are still valid. Therefore, the theoretical and the L-BASICC bispectrum will be compared with each other. Later in this section the bias parameters  $b_1$  and  $b_2$  will be extracted as in Section 3.2.1.

*Investigation of the validity of the model*

In this section, the validity of the model will be examined, which means how well the model is able to reproduce the features in the measurement. The measurement and the model will be compared with each other by calculating the ratio between them, with no fitting procedure inbetween. This helps to visualize the range of validity of the model which cannot be achieved by just examining the extracted bias parameters. Of course, this direct comparison between the model and the measurement is only possible for the dark matter simulation. For the dark matter halos the MCMC must be performed first.

The ratios between the measurement and the theory are given in the Figures 3.21, 3.22 and 3.23 for  $z = 0.0, 0.5$  and  $1.0$ , respectively. The lower x-axis in the figures shows the *configuration id*, the position of the triangle configuration, when ordered by the length of the triangle sides. The upper x-axis shows the multiples of  $k_f$  in relation with the *configuration id*, the tics mark the equilateral case. Each tic shows the maximum side length of the triangle up to that *configuration id*. Below such a tic all triangle configurations are considered where one side of the triangle has the length of the  $k$ -mode defined by that tic. The real space ratios can be found on the left panel and the redshift space ratios on the right panel.

In each real space panel (left side) of the ratio figures two curves are shown, in black all triangle configurations are taken into account for calculating the ratio and in blue the collapsed configurations are excluded. In redshift space (right panel), four curves are plotted. The blue and black solid lines show the ratios when the FoGs are taken into account, in red and green they are excluded. The ratios with the collapsed triangles are plotted in black and red whereas for blue and green they are not taken into account. The velocity dispersion of the particles required for the FoG can be calculated theoretically by

$$\sigma_v^2 = \frac{1}{6\pi^2} \int P_{\theta\theta}(k) dk \quad (3.5)$$

and is  $6.071 h^{-1}$  Mpc,  $5.054 h^{-1}$  Mpc and  $3.820 h^{-1}$  Mpc for  $z = 0.0, 0.5, 1.0$ , respectively.

*Discussion of the ratios*

In all figures at all redshifts spikey deviations up to 40% below the unity line are present for the ratios of the L-BASICC estimation and the theoretical prediction, if the collapsed configurations are included for the calculation of the ratios. These deviations will be called negative spikes or peaks in the following discussion. The spikey deviations above the unity line will be called positive spikes or peaks.

These negative peaks are the ratios of the collapsed triangle configurations. Therefore, the collapsed triangles are responsible for the strongest deviation between the model and the L-BASICC estimation independent on the considered  $k$ -mode. This is a clear hint that these configurations cannot be modeled even on the largest scales. Configurations close to collapsed ones can also be related to these spikes because they are in the same bin as the collapsed configurations (binsize is  $2 \times k_f$ ). The order of the triangles is also responsible

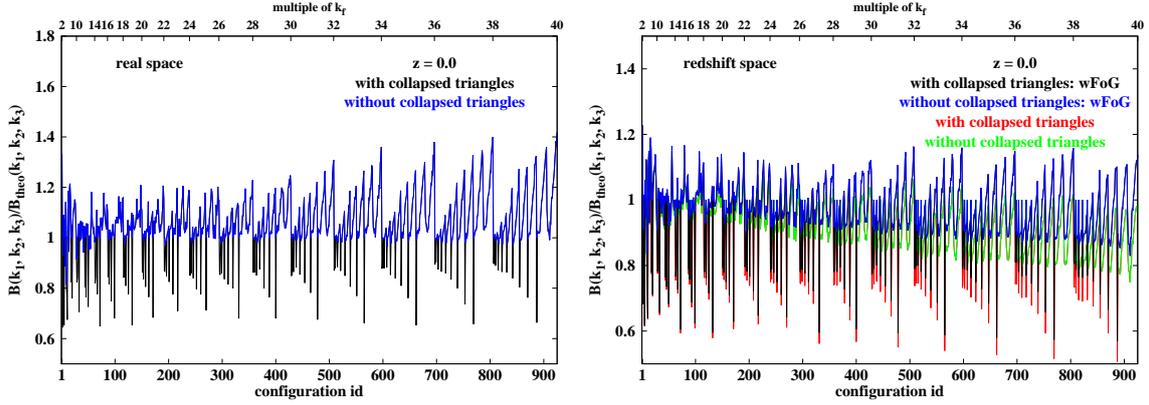


Figure 3.21:  $B(k_1, k_2, k_3)/B_{theo}(k_1, k_2, k_3)$  at  $z = 0.0$  is plotted against the *configuration id* of the triangles (lower x-axis). The upper x-axis displays the multiples of  $k_f$  in relation with the *configuration id*. The ratios in real and redshift space are shown on the left and the right panel, respectively. Whether the collapsed configurations are taken into account or not is illustrated in black and blue for real space. In redshift space, the black and blue solid lines indicate that the FoGs are considered. In red and green the redshift space ratios without the FoGs are shown for the collapsed triangle configurations included or excluded for the calculation of the ratios, respectively.

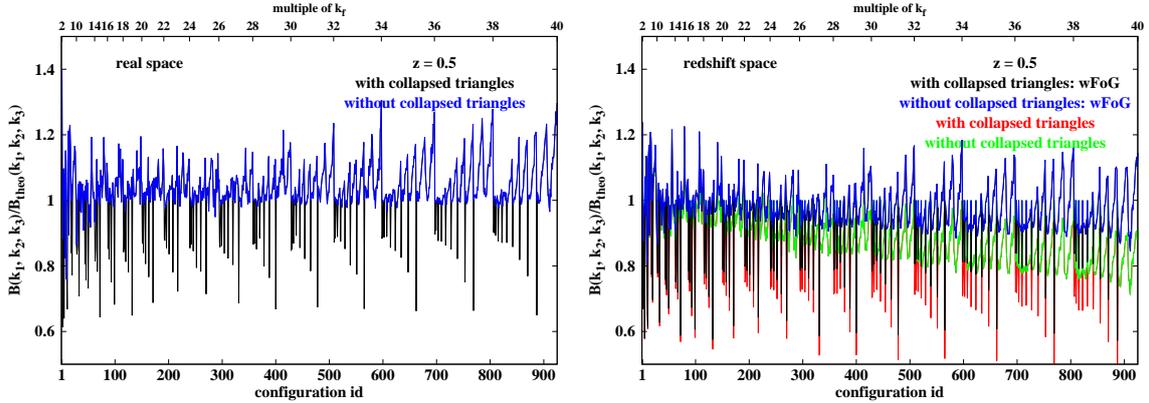


Figure 3.22: Same description as for Figure 3.21, but for  $z = 0.5$

for the positions of the negative spikes. At each of such a spike a new configuration with two fixed sides is considered, the third one is varied until the closure of the triangle cannot be assured anymore.

It can also be noticed, that at a certain *configuration id* positive peaks with a high amplitude are present. This means that the measured bispectrum is larger than the theoretical one. At that *configuration id* it seems that these positive peaks have a periodic behavior. The higher the redshift the larger is the  $k$ -mode from which this effect can be observed independent on the *coordinate frame* (which means real or redshift space). In redshift space, these peaks are also present but at lower *configuration ids* compared to real space at the

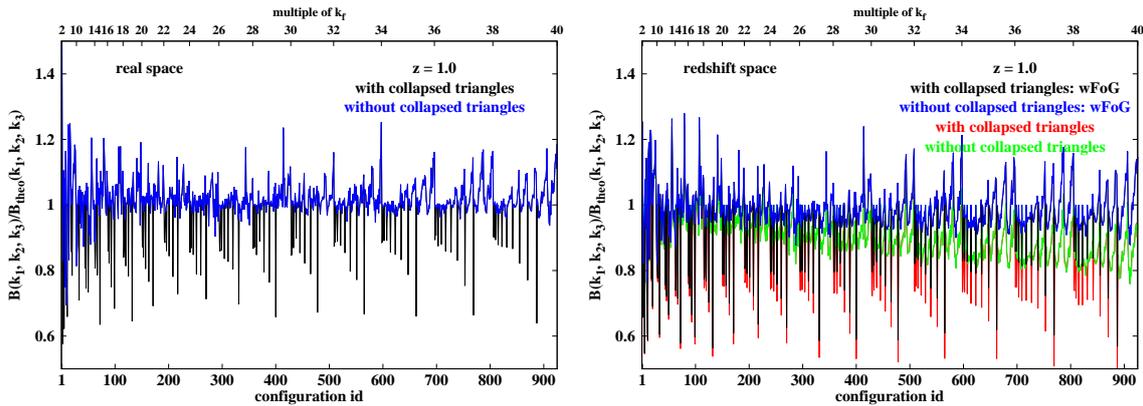


Figure 3.23: Same description as for Figure 3.21, but for  $z = 1.0$

same redshift. Obviously, these peaks influence the estimation of the bias parameters.

At  $z = 1.0$  in real space, these peaks with periodic behavior start at around the *configuration id* of 450, which corresponds to a  $k_{max} = 0.150 h \text{ Mpc}^{-1}$ . Around that  $k$ -mode the trend in  $b_2$  becomes more prominent, see right panel of Figure 3.26.

For redshift space the situation is similar, the positive peaks are present at lower *configuration ids*. Therefore, the bispectrum can only be modeled up to a lower  $k$ -mode compared to real space. The effect of the random motions can also be identified. The theoretical bispectrum has a too high amplitude at large  $k$ -modes compared to the L-BASICC bispectrum. The random motions reduce the clustering signal, as discussed in Section 2.6.2. The theoretical estimation of the velocity dispersion is able to mimic this effect and deviations between the two bispectra are reduced. If the random motions are not taken into account, the model will differ earlier from the measurement. The *configuration id* at which the deviation starts in redshift space at  $z = 1.0$  is around 150, which corresponds to  $k_{max} = 0.094 h \text{ Mpc}^{-1}$ . The random motions help to extent the analysis up to a *configuration id* of 350, a  $k$ -mode of  $0.131 h \text{ Mpc}^{-1}$ .

At lower redshifts these peaks are more prominent at lower *configuration ids*. In real space, the *configuration id*, where the positive peaks can clearly be identified, is around 350 ( $k_{max} = 0.131 h \text{ Mpc}^{-1}$ ) and 150 ( $k_{max} = 0.094 h \text{ Mpc}^{-1}$ ) for  $z = 0.5$  and  $0.0$ , respectively. For redshift space without considering the random motions this *configuration id* is around 150 ( $k_{max} = 0.094 h \text{ Mpc}^{-1}$ ) and 100 ( $k_{max} = 0.084 h \text{ Mpc}^{-1}$ ) for  $z = 0.5$  and  $0.0$ , respectively. The situation can be mildly improved by the inclusion of the random motions. The theoretical bispectrum can be trusted up to a *configuration id* of around 200 ( $k_{max} = 0.104 h \text{ Mpc}^{-1}$ ) and 150 ( $k_{max} = 0.094 h \text{ Mpc}^{-1}$ ) for  $z = 0.5$  and  $0.0$ , respectively. This short discussion shows, that the correct estimation of  $b_1$  and  $b_2$  is only possible up to a certain  $k$ -mode, up to which the bispectrum can be modeled correctly. It was decided for the later investigation of this section to consider only well modeled  $k$ -modes. This means all sides of the triangle must be modeled correctly by the modified tree-level ansatz in order to avoid misinterpretations of the found results. It is possible to find well described triangle configurations at larger  $k$ -modes but this is not subject for this first detailed investigation

of the bispectrum. However, in later studies this should be examined in detail.

#### *Explanations of the ratios*

From the discussion above, it can be stated that the theory is not able to account for collapsed triangles. In any analysis, where these configurations are considered, the amplitude of the model is too high. A smaller  $b_1$  would be needed to account for this. These spikes would also introduce a wrong shape dependence, because the amplitude of the correct modeled configurations are then too low and a higher  $b_2$  is required to compensate for this. The discussion in Section 2.6.1 showed, that the  $b_1^2 b_2$ -term in Equation (2.78) has a different shape dependence than the F-kernel term. Therefore, the whole theoretical bispectrum would be calculated wrongly if they were taken into account for the estimation of the bias parameters. The negative spikes are understood by now but there are also positive spikes present.

The equilateral triangles are responsible for the periodic positive peaks with the highest amplitude in such a series. These highest amplitude peaks coincide with the tics of the upper  $x$ -axis which mark the equilateral triangle configurations. As pointed out in Section 3.2.1, that at larger  $k$ -modes ( $\gtrsim 0.15 h \text{ Mpc}^{-1}$  in real space and  $\gtrsim 0.10 h \text{ Mpc}^{-1}$  in redshift space for  $z = 1.0$ , of course these values vary for different redshifts) the model is not able to predict the correct amplitude of the bispectrum anymore. These peaks are the manifestation of the insufficient modeling of non-linear structure growth at these scales. The increasing wing towards such a peak just shows how the remaining sides of the triangle become larger (in Fourier space) and finally result in the equilateral configuration (the highest peak in such a series).

The smaller peaks in such a series can be explained similarly. Two side lengths of the triangle are defined and the third one is getting larger until the condition  $k_1 \geq k_2 = k_3$  is fulfilled. This marks the most non-linear triangle for such a configuration (because  $k_1$  and  $k_2$  are fixed and  $k_3$  is varied). The wing only represents the ratio of two U-shape bispectra in the interval  $0.0 \leq \theta_{12} \leq \pi/3$ . Due to the chosen binning ( $2 \times k_f$ ), the length of such a *period* is dependent on the two fixed side lengths.

As mentioned in the discussion of the equilateral triangle configurations, the  $k$ -mode at which the tree-level ansatz is failing must be found out to know the range of validity of the model. From the discussion above, it can be noticed that there are configurations that are better described by the theory even if the tree-level ansatz is already insufficient for one of the sides of the triangle. However, for this thesis it was decided to consider only modes which are well modeled in order to get a better understanding of the non-linearities which affect the bispectrum. For a later possible application to a real data set an investigation of these better described configurations is useful for obtaining tight constraints on the bias parameters.

This short discussion shows, that the correct estimation of the linear and the quadratic bias is only possible up to a certain  $k$ -mode, up to which the bispectrum can be modeled correctly, as claimed previously in this section. This statement is true for real and redshift space. At larger  $k$ -modes (the exact value is dependent on the considered redshift

and the *coordinate frame* (real and redshift space)) the analysis is not trustworthy anymore.

*Extracting the bias parameters  $b_1$  and  $b_2$  from the dark matter bispectrum*

In this section, the MCMC will be carried out to extract the bias parameters. By performing the analysis at  $z = 0.0$ ,  $z = 0.5$  and  $z = 1.0$  the redshift dependence on the bias parameters estimation will also be investigated. Furthermore, it was also decided to analyze the effect of the random motions on the bispectrum and for the extraction of the bias parameters. Throughout the discussion of the equilateral configurations, this analysis was avoided due to the limited data points (number of different triangles) which were available for that analysis. By considering all possible triangle configurations this limitation does not hold any longer.

$z$		$k_{max}$ [ $h \text{ Mpc}^{-1}$ ] for $b_1$	$k_{max}$ [ $h \text{ Mpc}^{-1}$ ] for $b_2$
0.0	rs	0.10	0.09
	zs	0.08	0.08
	zs with FoG	0.09	0.08
0.5	rs	0.12	0.10
	zs	0.09	0.09
	zs with FoG	0.11	0.10
1.0	rs	0.15	0.14
	zs	0.09	0.11
	zs with FoG	0.13	0.13

Table 3.6: In this table, the  $k_{max}$ -values, up to which an accordant estimation of the bias parameters  $b_1$  and  $b_2$  with the dark matter expectation values is possible, are summarized. The analysis is performed on the dark matter bispectrum in real space called *rs*, redshift space with only coherent infall *zs* and with the FoG included *zs with FoG*. In column 1 the considered redshift can be found, in column 2 the analyzed *coordinate frame* (which means real or redshift space) and in column 3 and 4 the obtained  $k_{max}$ -values for  $b_1$  and  $b_2$ , respectively.

In the Figures 3.24 (for  $z = 0.0$ ), 3.25 (for  $z = 0.5$ ) and 3.26 (for  $z = 1.0$ ), the values for  $b_1$  (left panels) and  $b_2$  (right panels) are plotted against  $k_{max}$ . The  $k_{max}$ -value defines the largest possible side length of a triangle up to which all configurations were taken into account for the fitting process. The values of  $k_{max}$  up to which the linear and the quadratic bias can be estimated in agreement with the expectation values are summarized in Table 3.6. Three different results are contained in these figures depending on the analysis which was carried out, the results obtained from the real space bispectrum  $B_{rs}(k_1, k_2, k_3)$  (see Equation (2.69)) are given in blue, the redshift space bispectrum without any random motions or FoG  $B_{zs}(k_1, k_2, k_3)$  (see Appendix A.1) in red and in black  $B_{zs,mFoG}(k_1, k_2, k_3)$  (see Equation (2.98)) with the random motions or FoG included. In Figure 3.27, the results for

the fitted velocity dispersions  $\sigma_v$ s are shown for  $z = 0.0$ ,  $z = 0.5$  and  $z = 1.0$  in blue, red and black, respectively.

The MCMC was performed with  $0.0 \leq b_1 \leq 5.0$ ,  $-20.0 \leq b_2 \leq 20.0$  and  $0.0 h^{-1} \text{ Mpc} \leq \sigma_v \leq 40.0 h^{-1} \text{ Mpc}$ , for the velocity dispersion when the random motions are included, as flat priors. For the exploration of the parameter space the rms of the Gaussian jumping function was limited by  $\sigma_{b_1} = 0.05$ ,  $\sigma_{b_2} = 0.15$  and  $\sigma_{\sigma_v} = 2.5 h^{-1} \text{ Mpc}$ .

In all figures (which means for all redshifts independent of the *coordinate frame*), two trends can be identified. The linear bias  $b_1$  is underestimated by the model, whereas the quadratic bias  $b_2$  is overestimated. These trends are more pronounced the lower the considered redshift is.

If only one redshift is examined, the deviation of the extracted bias parameters compared

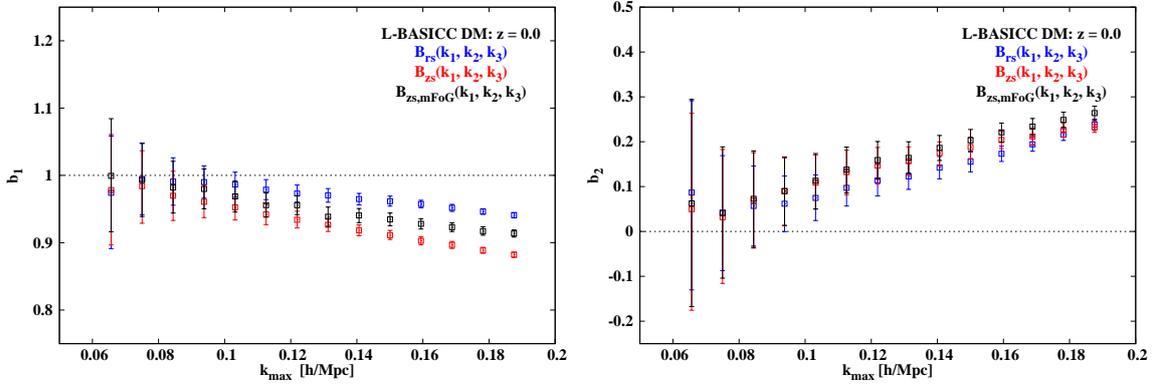


Figure 3.24: Estimations of the  $b_1$  (left panel) and the  $b_2$  (right panel) parameters are plotted against  $k_{max}$ , the maximum  $k$ -mode for the side length of a triangle configuration. The results in real space are given in blue ( $B_{rs}(k_1, k_2, k_3)$ ), redshift space in red ( $B_{zs}(k_1, k_2, k_3)$ ) and black ( $B_{zs,mFoG}(k_1, k_2, k_3)$ ) when the FoG are taken in into account or not, respectively, at  $z = 0.0$ . The black dotted lines in each panel show the dark matter expectation values for the bias parameters ( $b_1 = 1.0$  and  $b_2 = 0.0$ ).

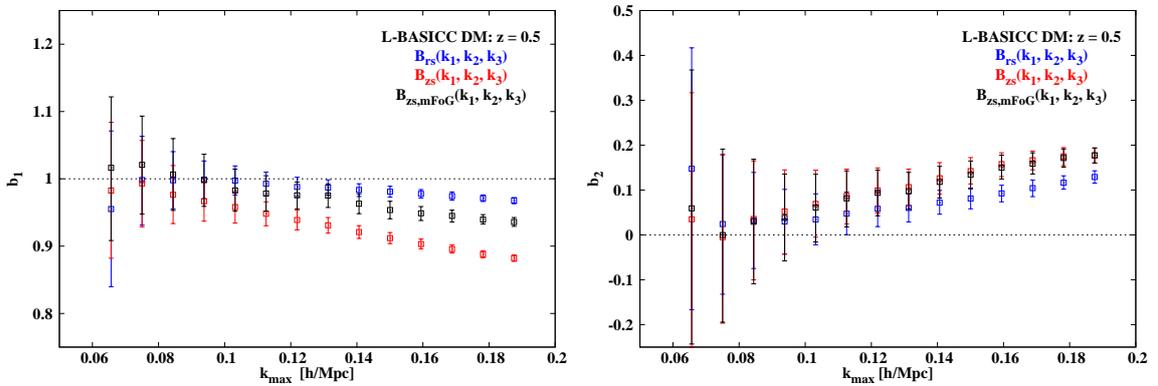


Figure 3.25: Same description as for Figure 3.24, but for  $z = 0.5$

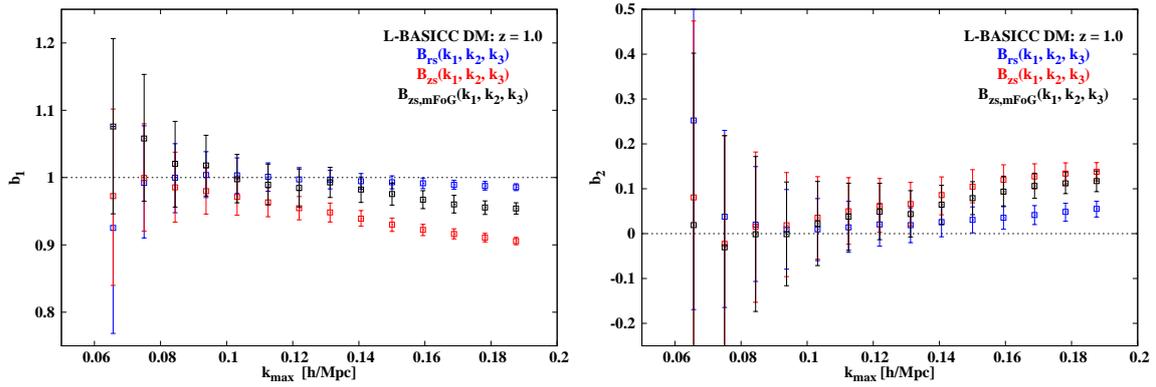


Figure 3.26: Same description as for Figure 3.24, but for  $z = 1.0$

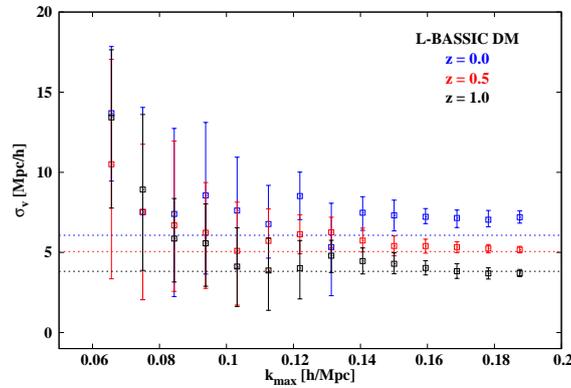


Figure 3.27: The estimated  $\sigma_v$ -values are summarized in this figure. The results obtained for the different redshifts  $z = 0.0, 0.5$  and  $1.0$  are shown in blue, red and black, respectively. The theoretical estimations are indicated by the dotted lines with the corresponding color for the considered redshift.

to the dark matter expectation values will become larger by going to larger  $k$ -modes. This is also true for all investigated *coordinate frames*. A consistent estimation of the linear and the quadratic bias in real space can be achieved up to larger  $k$ -modes compared to redshift space.

In the previous section, it was claimed that by including the random motions, the extraction of  $b_1$  and  $b_2$  in redshift space can be extended to larger  $k$ -modes if these results are compared to the case where only the linear Kaiser effect is modeled. This assumption can be verified by this analysis (see Figures 3.24, 3.25 and 3.26). However, the improvement is very mild and its significance becomes less important the lower the redshift is.

*Discussion of the results of the extracted bias parameters*

The above discussed trends can be explained by

- the non-linear structure growth and
- the redshift space distortions, which are also non-linear (see non-linear Kaiser effect (Scoccimarro, 2004)).

The discussion given here will start in real space. Only the growth of structure is able to influence the clustering signal for the dark matter case. Therefore, the overall explanation for these trends lies in the description of the non-linear structure growth. It was stated that the model is able to deliver meaningful results for  $k$ -modes smaller than  $0.15 h \text{ Mpc}^{-1}$  at least for the  $b_1$ -estimation at  $z = 1.0$ , see Section 3.2.1. This limit is also confirmed by the obtained results in this section. By going to lower redshifts the model is restricted to even smaller  $k$ -modes. This behavior is expected because the structure growth becomes more non-linear at lower redshifts. The model used for the extraction of  $b_1$  and  $b_2$  is restricted to the weakly non-linear regime only. Therefore, the range of validity of the model shrinks when going from higher to lower redshifts what can be noticed by looking at Table 3.6 as well.

In redshift space, the situation is more complicated. Not only non-linearities, induced by gravity, have to be taken into account, but also peculiar velocities and their effect on the volume averaged bispectrum. In the analysis here, it is shown (see the figures above and Table 3.6), that at least on the largest scales the redshift space distortions for the bispectrum can be modeled by the linear Kaiser effect. The modeling of the random motions should mainly consider the small scale effect of peculiar velocities of objects within a massive structure, see Section 2.6.2. A mild improvement is possible when considering the random motions of dark matter particles, see the discussion on the equilateral triangles given in Section 3.2.1. The inclusion of the FoG also alters the shape of the redshift space bispectrum and, therefore, is able to provide an improvement for the estimation of the bias parameters  $b_1$  and  $b_2$ . This can partly be explained by the random motions itself, which are able to mimic some of the features in the bispectrum originated by the non-linear structure growth and correct for the lack of the non-linear Kaiser effect in the model, as discussed in Section 3.2.1. Hence, random motions, which are set as a free fitting parameter, act like an artificial smoothing.

On the largest scales, coherent infall should be dominant, which is modeled by the linear Kaiser effect (Heavens et al., 1998). However, deviations already start at very large scales (see Table 3.6). If this large scale signal is really originated from large scale random motions, they will also be present in the dark matter halo catalog, where the small scale random motions are zero, because there is not any substructure present. A detailed analysis on the dark matter halo catalog will be given in the following section.

There seems to be a degeneracy present between  $b_1$  and  $b_2$  due to their opposite trends at all redshifts independent on the *coordinate frame*. As shown in Section 2.6.1, the first term of Equation (2.78) is only able to change the amplitude of the bispectrum, despite of

the shape dependence of the chosen triangle configuration given by the F-kernel. It was concluded from Figure 2.8, that an additional shape dependence is introduced by the  $b_1^2 b_2$ -term. Therefore, the shape of the theoretical bispectrum can be altered by a combination of  $b_1$  and  $b_2$ , which corresponds to a different weighting of the shape dependencies given by the two terms of Equation (2.78). This will lead to a degeneracy between  $b_1$  and  $b_2$ .

The fact, that  $b_2$  is overestimated at all scales, can only be explained by such an additional shape dependence, which was not taken into account so far. Then, the underestimation of  $b_1$  is required to compensate for the overestimation of  $b_2$ . This can explain to some extent the visible trends in the figures above.

In real space, this can only be a hint towards the insufficient modeling of structure growth, as claimed above. In redshift space, these trends are more pronounced which is an indication that not only the modeling of the growth of structure but also of the redshift space distortions is insufficient at large  $k$ -modes. An extension to higher order perturbation theory could help to get a better theoretical description for these modes but such an investigation is beyond the scope of this thesis.

#### *Discussion of the results of the velocity dispersion*

Figure 3.27 shows the findings for the extracted velocity dispersion  $\sigma_v$  at different redshifts. A trend can be noticed, that the lower the redshift is, the larger are the  $\sigma_v$ -values. This is expected because the particles had more time to organize themselves in massive structures, where their random motions are dependent on the depth of the gravitational potential well.

This is a direct consequence of the virial theorem ( $2\langle T \rangle = -\langle V \rangle$ , for a conservative and central force) which quantifies the time averaged contribution of the kinetic energy  $\langle T \rangle$  and the potential energy  $\langle V \rangle$  to the total energy. It was already concluded that the modeling of the random motions is able to mimic some features of structure growth and also correct for the lack of the non-linear Kaiser effect in the modeling and hence act as an artificial smoothing. This statement could be made by looking at the smallest  $k$ -modes ( $\lesssim 0.10 h \text{ Mpc}^{-1}$ ), where the measured values for  $\sigma_v$  appear to be very large. A more conclusive statement is not possible because of the large errors of the extracted  $\sigma_v$ -values. Around that  $k$ -modes only a limited number of triangle configurations is available for measuring the bispectrum and cosmic variance has a huge influence on the bispectrum. This leads to large errors for the measured bispectrum and for the later extracted fitting parameters as it can be seen in Figure 3.27.

Another explanation for the large  $\sigma_v$ -values could be, that there are really large scale random motions present. However, the errors in Figure 3.27 are too large for more conclusive statements.

By comparing the theoretical results of Equation (3.5) with Figure 3.27, it can be noticed that they are in agreement with each other. Except for  $z = 0.0$  where a  $\sigma_v \approx 7.100 h^{-1} \text{ Mpc}$  was extracted from the dark matter simulations (as a reminder, the theoretical  $\sigma_v = 6.071 h^{-1} \text{ Mpc}$ ). The integral in Equation (3.5) covers all  $k$ -modes, even the ones, on which  $3^{\text{rd}}$  order perturbation theory  $P_{\theta\theta}(k)$  cannot be predicted correctly anymore.

The made assumptions (e.g. no vorticity, Newtonian treatment scales of interest smaller than the Hubble distance) for the derivation of the  $3^{rd}$  order perturbation theory are not valid anymore and the theoretical power spectrum overestimates the power spectrum from e.g. the L-BASICC simulations (of course the situation becomes worse at lower redshifts). Therefore, the low redshift  $\sigma_v$  cannot be computed correctly.

In Section 2.6, it was already concluded to use the modified tree-level ansatz. However, at this stage of the thesis it should be shown in a more extended discussion that this modification really improves the model.

### *Investigation of the tree-level ansatz*

In the previous sections, it was claimed, that the usage of the  $3^{rd}$  order perturbation theory power spectra in Equation (2.69) provides a better description of the bispectrum than the linear power spectra. This means that by modifying the original tree-level ansatz in the suggested manner helps to extend the range of validity of the model. In Section 2.6.1, it was attempted to prove this statement by a short qualitative and quantitative discussion. In this section, this statement will be corroborated by the estimation of the linear and the quadratic bias by means of the original tree-level approach given by Equation (2.78). The results of the modified tree-level ansatz are illustrated in blue and red for real and redshift space, respectively. The results of the original tree-level ansatz are plotted in green and purple for real and redshift space, respectively.

The results of this analysis on  $b_1$  and  $b_2$  will be compared with the previous results of this section. In the Figures 3.28, 3.29 and 3.30, the results for  $b_1$  and  $b_2$  are plotted against  $k_{max}$  for  $z = 0.0, 0.5$  and  $1.0$ , respectively. However, this analysis is restricted to real and

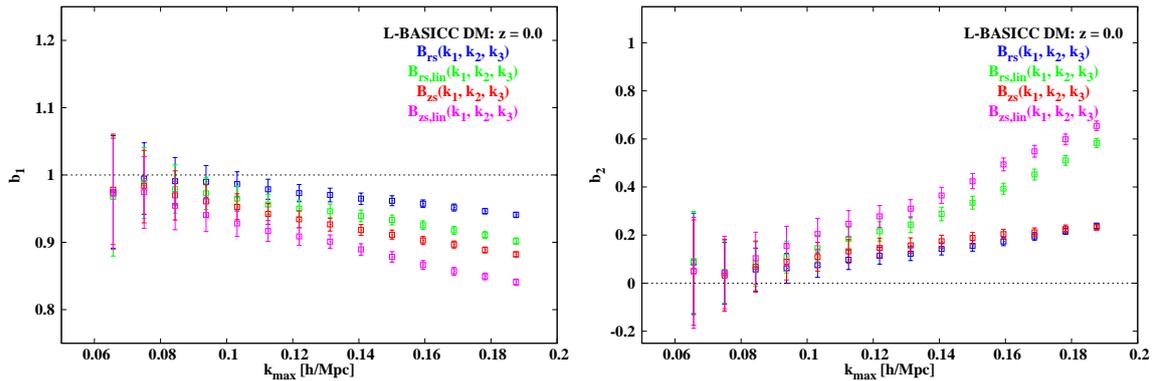
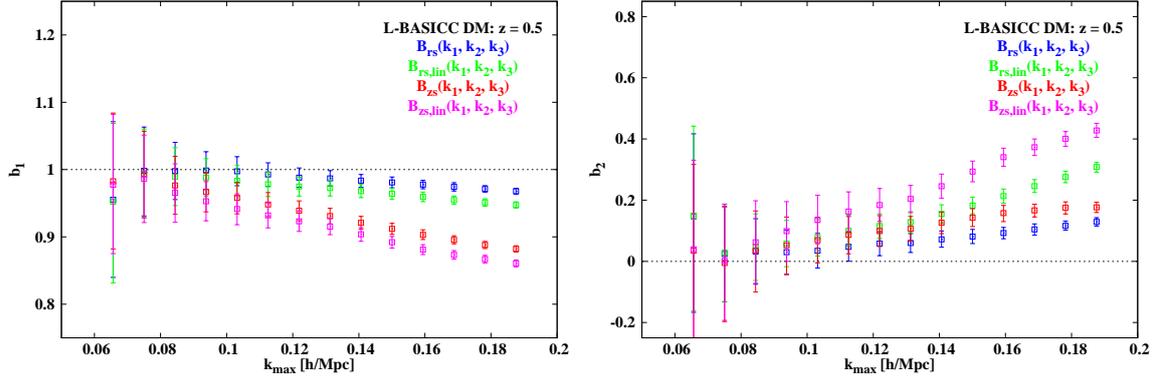
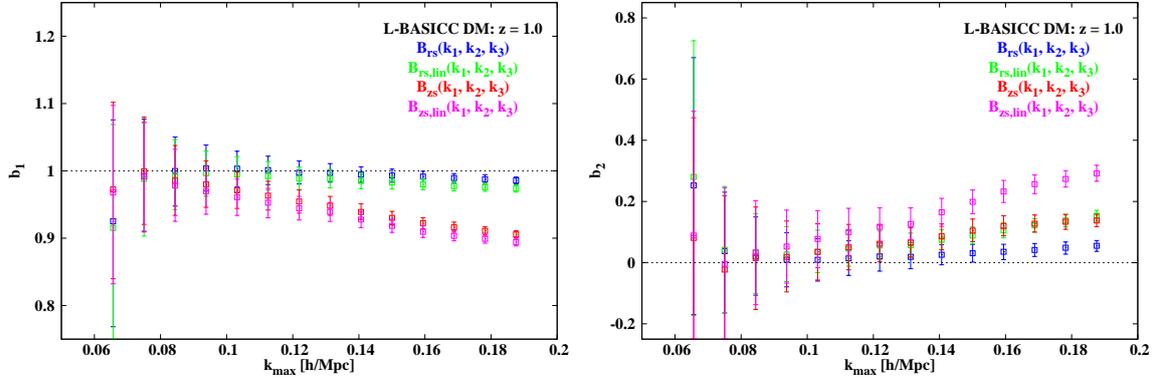


Figure 3.28: Estimations of the  $b_1$  (left panel) and the  $b_2$  (right panel) parameter are plotted against  $k_{max}$  at  $z = 0.0$ . The results in real space for the modified and the original tree-level ansatz are given in blue ( $B_{rs}(k_1, k_2, k_3)$ ) and green ( $B_{rs,lin}(k_1, k_2, k_3)$ ), respectively. The redshift space estimations are shown in red ( $B_{zs}(k_1, k_2, k_3)$ ) and purple ( $B_{zs,lin}(k_1, k_2, k_3)$ , original tree-level ansatz). The FoGs were not considered in this analysis.

Figure 3.29: Same description as for Figure 3.28, but for  $z = 0.5$ Figure 3.30: Same description as for Figure 3.28, but for  $z = 1.0$ 

redshift space where only the linear Kaiser effect is taken into account because considering the FoG only leads to a mild improvement of the extracted bias parameters.

It can clearly be concluded that  $b_1$  and  $b_2$  can be estimated in agreement with the dark matter expectation values for all considered redshifts for both models on the smallest  $k$ -modes. The deviations between the two prescriptions of the bispectrum already start at small  $k$ -modes independent on the redshift. All trends reported in the discussions above are also present for the analysis of the original tree-level ansatz but are more pronounced than in the modified case. From the direct comparison of the two bispectrum modeling recipes, it can be concluded that the modified tree-level ansatz is able to obtain consistent results with the dark matter expectation values for  $b_1$  and  $b_2$  at  $k$ -modes where the original one already fails.

In real space, a reliable estimation of  $b_1$  is possible up to a  $k_{max} \approx 0.09 h \text{ Mpc}^{-1}$ ,  $0.11 h \text{ Mpc}^{-1}$  and  $0.13 h \text{ Mpc}^{-1}$  for  $z = 0.0$ ,  $0.5$  and  $1.0$ , respectively. Approximately the same results for  $k_{max}$  can be found for the  $b_2$  parameter. It can also be noticed from the figures, that the situation for  $B_{zs,lin}(k_1, k_2, k_3)$  is worse. A consistent estimation of  $b_1$  for  $B_{zs,lin}(k_1, k_2, k_3)$  compared the dark matter expectation value for  $b_1$  is only possible up

to a  $k_{max} \approx 0.07 h \text{ Mpc}^{-1}$ ,  $0.08 h \text{ Mpc}^{-1}$  and  $0.09 h \text{ Mpc}^{-1}$  for  $z = 0.0, 0.5$  and  $1.0$ , respectively. The same values are upper limits for the  $b_2$  estimation. Except for  $z = 1.0$  where  $b_2$  is in agreement with the expectation value up to a  $k_{max} \approx 0.10 h \text{ Mpc}^{-1}$ , mainly due to the large errors. The trend to overestimate  $b_2$  is already present at smaller modes as for the results of  $B_{rs}(k_1, k_2, k_3)$  and  $B_{zs}(k_1, k_2, k_3)$ .

#### *Discussion of the results of the original tree-level ansatz*

It is concluded that the modified tree-level ansatz improves the estimation of the bias parameters compared to original tree-level ansatz. From this analysis, it can also be concluded that non-linear structure growth can be handled in a better way by the modified tree-level approach. These statements are valid for real and redshift space. It can be noticed by looking at the figures above, that the reported trends are more pronounced in redshift space than in real space.

The inclusion of the random motions in redshift space is excluded from this analysis because it was already found out that only a mild improvement will be gained, if they are taken into account (see the discussion on the dark matter bispectrum). It is expected that the overall trends, which were measured here, would remain unchanged. Therefore, redshift space was just modeled by considering the linear Kaiser effect.

It can be concluded that the modified tree-level ansatz is better suited for estimating  $b_1$  and  $b_2$  than the original one. In the next paragraph, the reduced bispectrum will be examined.

#### **Bias parameter extraction: reduced dark matter bispectrum**

The following discussion will be focused on the reduced bispectrum  $Q(k_1, k_2, k_3)$  extracted from the L-BASICC dark matter  $N$ -body simulations. It will be started with the comparison between the L-BASICC and the theoretical reduced bispectrum for which the ratios are calculated. These ratios are plotted in the Figures 3.31, 3.32 and 3.33 with the same color coding as in Figure 3.21 for  $z = 0.0, 0.5$  and  $1.0$ , respectively. The red solid line at unity indicates the level of perfect match between the L-BASICC and the theoretical bispectrum. The random motions are not taken into account due to their small impact on the bispectrum.

It can clearly be noticed that the reduced bispectrum cannot be modeled correctly independent on the redshift and the considered  $k$ -modes. Only a very limited range of agreement between measurement and theory can be identified well below a *configuration id* of 100 ( $k_{max} \approx 0.084 h^{-1} \text{ Mpc}$ ). After that *configuration id* the model predicts a too high signal for the reduced bispectrum compared to the measurement. The only exceptions are the collapsed triangle configurations which are still present as negative peaks in the figures above. Extracting bias parameters by means of the reduced bispectrum does not make any sense and is excluded from this thesis.

*Discussion of the results of the reduced dark matter bispectrum*

The insufficient modeling of the non-linear mode-coupling and the non-linear Kaiser effect are responsible for these trends. For understanding the above found trends it is necessary to remind oneself how the reduced bispectrum is defined. The reduced bispectrum is the ratio between a sum of squared power spectra, where the enumerator is weighted by the considered triangle configuration via the F-kernel, see Equation (2.66). To zeroth order the amplitude of  $Q(k_1, k_2, k_3)$  is given by the bias parameters and its shape by the chosen triangle configuration.

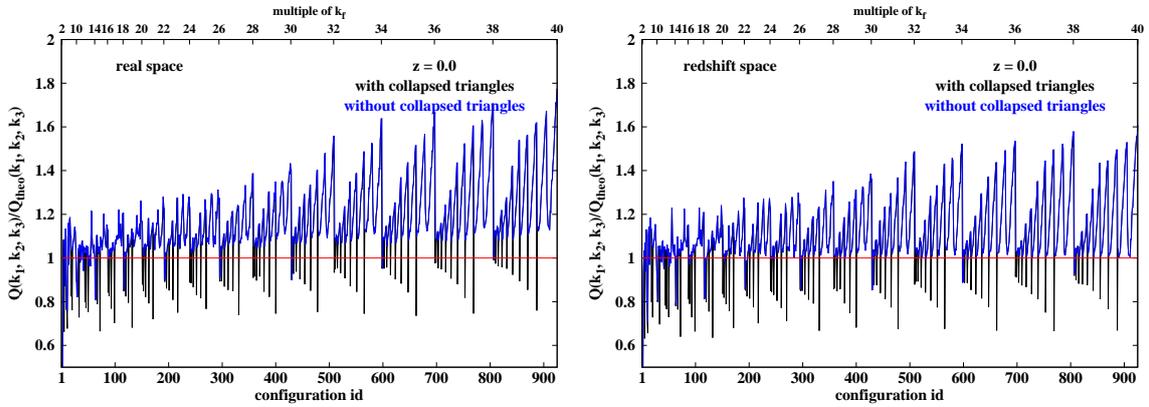


Figure 3.31:  $Q(k_1, k_2, k_3)/Q_{theo}(k_1, k_2, k_3)$  at  $z = 0.0$  is plotted against the *configuration id* of the triangles (lower x-axis). The upper x-axis displays the multiples of  $k_f$  in relation with the *configuration id*. The ratios in real and redshift space are shown on the left and the right panel, respectively. Whether the collapsed configurations are taken into account or not is illustrated in black and blue. The redshift space reduced bispectrum is only modeled by the linear Kaiser effect. The red vertical line represents the level of perfect match between measurement and theory.

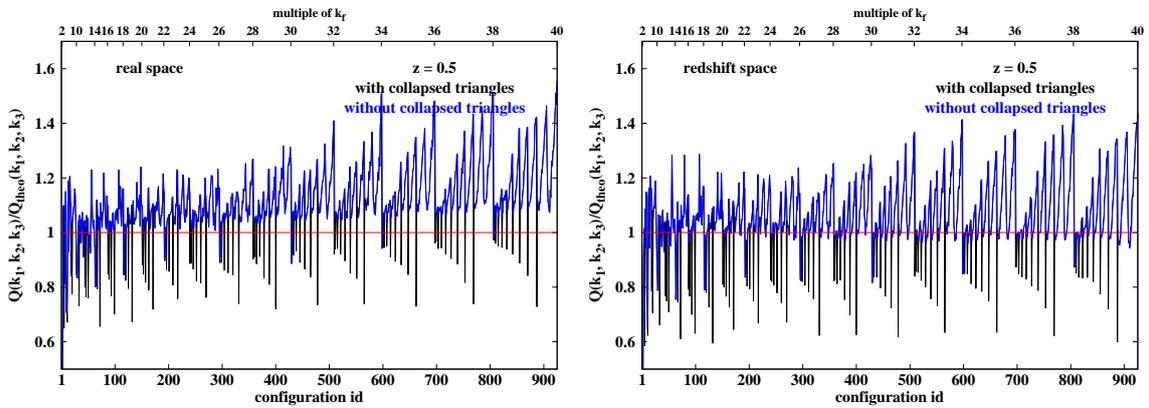


Figure 3.32: Same description as for Figure 3.31 but for  $z = 0.5$

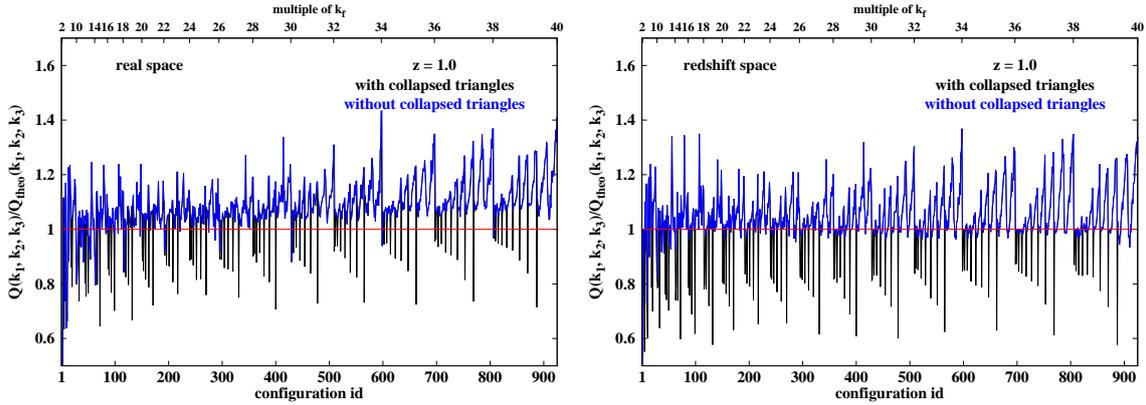


Figure 3.33: Same description as for Figure 3.31 but for  $z = 1.0$

The deviations from the unity line for the reduced bispectrum are more pronounced than for the bispectrum itself. For the reduced bispectrum the deviations range from 0.6 to 1.7 whereas for the bispectrum a range from 0.6 to 1.4 is covered. This is not surprising because the uncertainties in the modeling of the bispectrum and the power spectrum add up according to the error propagation law. The smaller the considered redshift is, the larger is the difference pronounced between the estimation and the expectation.

In real space, this is a hint towards the insufficiently modeled non-linear structure growth and with the lack of the non-linear Kaiser effect also in the modeling of the redshift space bispectrum. It can be concluded that as long as these effects are not modeled correctly, the reduced bispectrum cannot be utilized for the extraction of  $b_1$  and  $b_2$ .

The range of validity of the theoretical model was discussed in detail throughout this section by means of the dark matter bispectrum. As a next step the dark matter halo bispectrum catalogs will be examined in order to estimate the linear and the quadratic bias.

### Bias parameter extraction: dark matter halo bispectrum

As for the dark matter investigation, before discussing the results of the extracted bias parameters from the dark matter halo L-BASICC measurement is compared to the theoretical prediction. The best-fit bias parameters extracted from the MCMC by means of the bispectrum were used for this investigation.

The Figures 3.34, 3.35 and 3.36 show the ratios of the measurement and the prediction for  $z = 0.0, 0.5$  and  $1.0$ , respectively. The ratios from the real space bispectra and the redshift space bispectra are plotted on the left and the right panel, respectively. The black color indicates that the collapsed triangle configurations are considered for the evaluation of the ratios and in blue that they are disregarded. The perfect match between measurement and model is shown by the red solid line. Because dark matter halos are biased objects, bias parameters are required for the calculation of the theoretical predictions. The best-fit parameters were chosen for the  $k$ -range where a consistent extraction of the linear and the quadratic bias in real and redshift space can be obtained. The FoGs are not considered

in this discussion because in the previous investigation it was already concluded that they improve only mildly the estimation of these bias parameters (and therefore the bispectrum). It can be noticed that for each redshift there is a range of  $k$ -modes in which

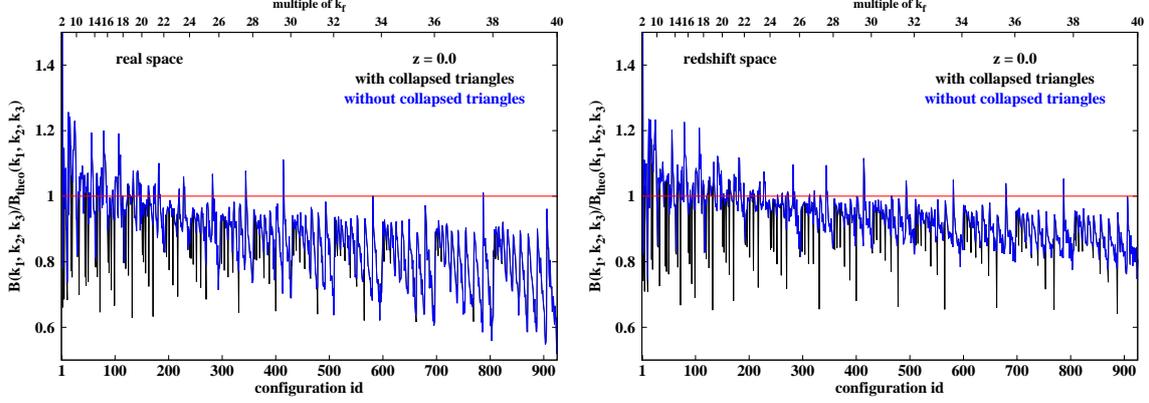


Figure 3.34: Same description as for Figure 3.21, but for the dark matter halo bispectrum

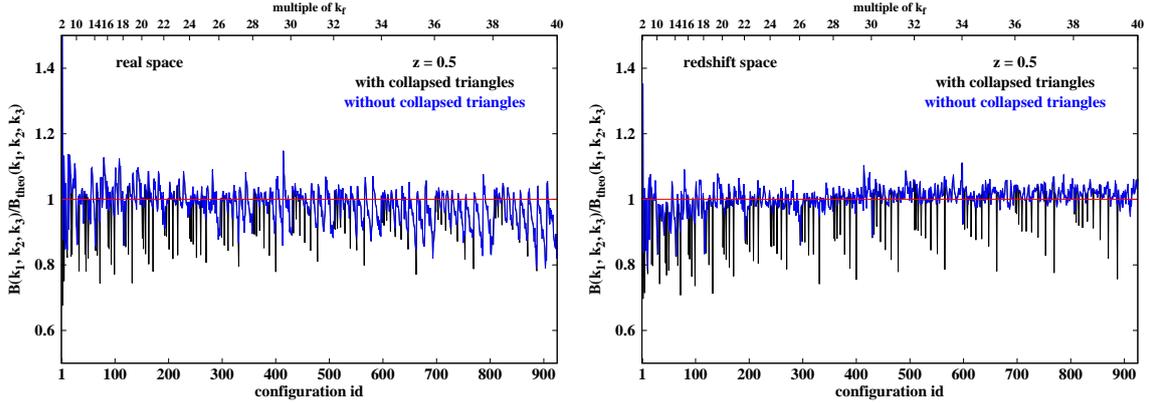


Figure 3.35: Same description as for Figure 3.34, but for  $z = 0.5$

the measurement and the theoretical prediction are in agreement with each other. This range shrinks when going from higher to lower redshifts. This can be expected from the discussion in the previous section.

At  $z = 1.0$  (independent on the *coordinate frame*), it looks as if the theory is not able to predict the measurement correctly. This can be explained by the low signal which can be extracted from the halo catalogs at that redshift. Only a few triangle configurations can be found due to the low number density for such a large volume (for comparison:  $z = 1.0$ :  $\bar{n} \approx 5.82 \times 10^{-5} (h/\text{Mpc})^3$ ;  $z = 0.5$ :  $\bar{n} \approx 1.25 \times 10^{-4} (h/\text{Mpc})^3$ ;  $z = 0.0$ :  $\bar{n} \approx 5.82 \times 10^{-5} (h/\text{Mpc})^3$ ). For simplicity it was avoided to plot the errors.

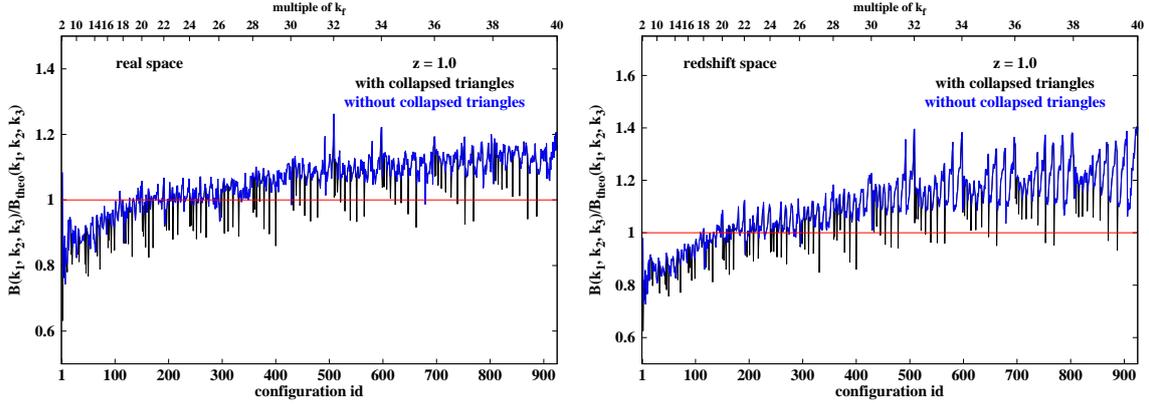


Figure 3.36: Same description as for Figure 3.34, but for  $z = 1.0$

*Extraction of the bias parameters  $b_1$  and  $b_2$  from the dark matter halo bispectrum*

In this section, the bias parameters  $b_1$  and  $b_2$  will be extracted from the dark matter halo L-BASICC catalogs by means of the bispectrum. The results of the MCMC are plotted in the Figures 3.37, 3.38 and 3.39 for  $z = 0.0$ ,  $z = 0.5$  and  $z = 1.0$ , respectively, and are summarized in Table 3.7. The figures of this section follow the notation of the previous section but without the inclusion of the FoG; on the left panel the results for  $b_1$  against  $k_{max}$  and on the right panel the results for  $b_2$  against  $k_{max}$  are shown.

As it was noticed for the dark matter case, on the smallest  $k$ -modes (on large scales, see Table 3.7) consistent results between real and redshift space can be achieved. At all redshifts, the extracted values for the linear bias  $b_1$  in real space are larger than in redshift space. For the quadratic bias  $b_2$  an opposite trend can be identified. These two behaviors were present in the dark matter case, too.

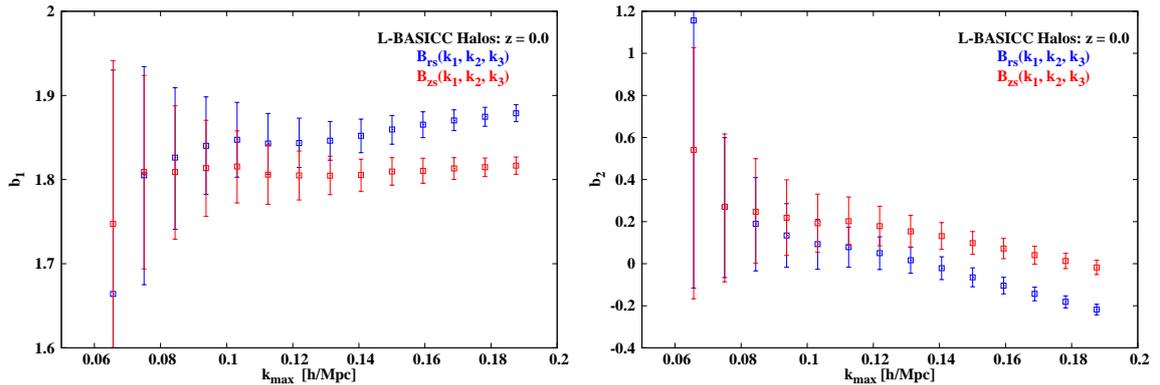
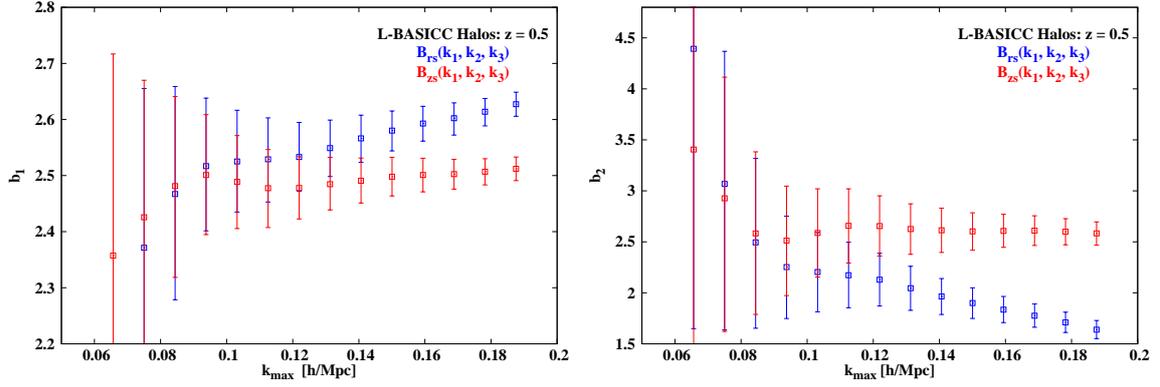
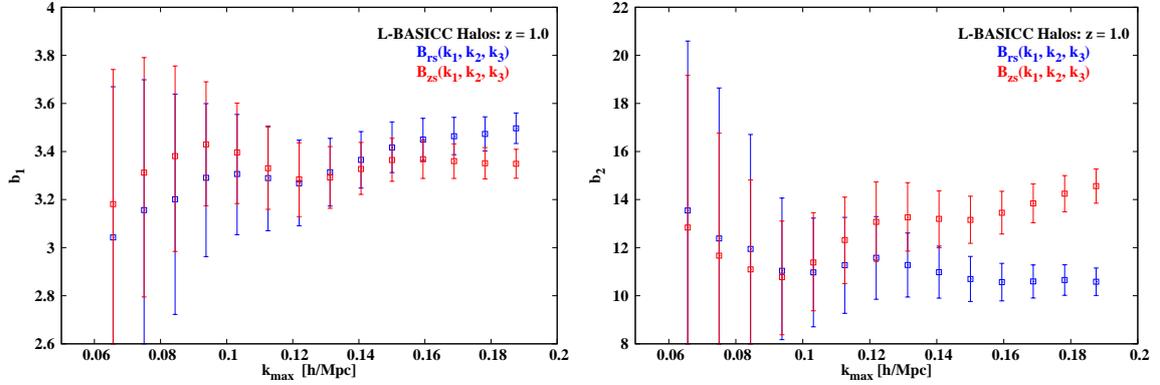


Figure 3.37: Estimations of the  $b_1$  (left panel) and the  $b_2$  (right panel) parameters are plotted against  $k_{max}$ , the maximum  $k$ -mode for the side length of a triangle configuration. The results in real space are given in blue ( $B_{rs}(k_1, k_2, k_3)$ ) and redshift space modeled with the linear Kaiser effect in red ( $B_{zs}(k_1, k_2, k_3)$ ), respectively, for  $z = 0.0$ .

Figure 3.38: Same description as for Figure 3.37, but for  $z = 0.5$ Figure 3.39: Same description as for Figure 3.37, but for  $z = 1.0$ 

$z_{obs}$	$(b_1, k_{max} [h \text{ Mpc}^{-1}])$	$(b_2, k_{max} [h \text{ Mpc}^{-1}])$
0.0	$(1.83 \pm 0.08, 0.10)$	$(0.18 \pm 0.24, 0.09)$
0.5	$(2.50 \pm 0.12, 0.11)$	$(2.54 \pm 0.94, 0.09)$
1.0	$(3.35 \pm 0.24, 0.15)$	$(11.04 \pm 3.05, 0.11)$

Table 3.7: The bias parameters  $b_1$  and  $b_2$  extracted from the dark matter halo catalogs with the  $k_{max}$ -values up to which a consistent estimation in real and redshift space can be achieved are given in column 2 and 3. The investigated redshift can be found in column 1.

The inclusion of the random motions does not change the discussed trends above. The improvement which can be achieved by their consideration is even less pronounced than it was already the case for the dark matter particles. At  $z = 0.5$  and  $1.0$  no improvement can be identified at all considered  $k$ -modes. Except for  $z = 0.0$  where the model slightly benefits from the inclusion of the FoGs at least for the extraction of  $b_1$  for  $k$ -modes larger than  $0.14 h \text{ Mpc}^{-1}$ . The  $k$ -modes up to which the real and redshift space results are not in agreement with each other do not change from the found results here. Such a behavior

cannot be identified for  $b_2$ .

Because of these mild effects on the bispectrum it was decided to exclude the results of the extracted bias parameters with the FoGs included for simplicity.

#### *Discussion of the results of the extracted bias parameters*

The trends found in the last paragraph can be explained by three possible effects

- the non-linear structure growth,
- the non-linear Kaiser effect and
- the lack of structure of the dark matter halos.

The first two points were discussed previously in detail for the dark matter case and for equilateral triangles. The last possible effect can be explained by comparing the dark matter results with the dark matter halo results. The dark matter halos do not contain any structure within the halo. They are just points within the catalogs with their positions and peculiar velocities. However, the dark matter particles follow the potential of the halos to which they are bound.

At a certain  $k$ -mode (depending on the depth of the potential of the halo), the chance of finding a triangle where two dark matter particles are located in one halo increases compared to the largest configurations where each dark matter particle is located in a different halo. At the same time, non-linear structure growth becomes more important because more non-linear configurations are considered. These configurations have already an effect on large scales.

The discussion becomes clearer in redshift space. By comparing the boost-factors for dark matter particles and dark matter halos, see Figures 3.16 and 3.19, respectively, different behaviors can be identified for  $k$ -modes larger than  $0.10 h \text{ Mpc}^{-1}$ . In the dark matter case a damping is present after that  $k$ -mode but for the halos the opposite is true. These two different behaviors show how the bispectrum is affected by the internal structure of halos which is resolved by dark matter particles because they follow the halo potential. As long as a chosen triangle configuration is large enough, the tree-level ansatz is sufficient for modeling the mode-coupling of the involved  $k$ -modes. When going to larger  $k$ -modes with one side length the non-linear structure growth becomes more important and the tree-level ansatz is a too simplified description for the mode-coupling.

In the halo catalogs, all considered corners for building triangles are located in a different halo. Therefore, the modeling can be sufficiently described by the tree-level ansatz which represents exactly this case. The non-linear structure growth is just given by the clustering of the halos.

This explains why the estimations of  $b_1$  and  $b_2$  can consistently be estimated in real and redshift space up to  $k$ -modes around  $0.11 h \text{ Mpc}^{-1}$ . This is almost the same limit as for the dark matter estimation at  $z = 1.0$  for redshift space. From the discussion in Section 3.2.1, this would not be expected for the dark matter halos at high redshifts. The reason is that

they are more non-linearly evolved structures compared to lower redshifts. The non-linear growth of structure and the lack of the structure within the halos are two counter-playing effects and can explain the results in this section.

The linear and the quadratic bias were also extracted for two different mass bins (mass of separation:  $M = 2.64 \times 10^{13} h^{-1} M_{\odot}$ ). The results confirmed the above conclusions. The high mass bin contains the most non-linearly evolved structures of the L-BASICC simulations. A consistent estimation between real and redshift space is only possible for smaller  $k$ -modes compared to the results of this paragraph. The contrary is true for the low mass bin because only the most linear structures of the simulations were considered.

#### *Investigation of the non-linear Kaiser effect*

In Figure 3.40, the ratios between the real and redshift space, or the (Kaiser) boost-factors, are shown at  $z = 0.0$  (top-left panel),  $0.5$  (top-right panel) and  $1.0$  (bottom-middle panel) in the upper sub-panel. All triangles configurations were considered for the following discussion. The boost-factors from the L-BASICC simulations and the model are plotted in black and red, respectively. In the lower sub-panel the ratio between the boost-factors are shown where the dotted black indicates when the two ratios would match perfectly. The L-BASICC boost-factors (black color) were directly extracted from the simulations without any theoretical assumptions for their evaluation. Therefore, for this case the non-linearities are treated correctly. Of course, this statement cannot be hold for the theoretical prediction (red color).

It can be noticed that the L-BASICC boost-factors increase when larger  $k$ -modes are regarded compared to the theoretical predictions. This behavior is strongly pronounced at  $z = 0.0$  and  $z = 1.0$  but not that strong at  $z = 0.5$ . This boost must have its origin in redshift space.

#### *Discussion of the results of the non-linear Kaiser effect*

In the discussion of the equilateral triangles, this increase of the signal was already noticed and identified as the non-linear Kaiser effect. The results from Section 3.2.2 (discussion of the equilateral triangles for dark matter halos) can be confirmed by the investigation of this paragraph. In the previous section, the breakdown of the redshift space bias parameters estimation at smaller  $k_{max}$ -values compared to real space was explained by the non-linear Kaiser effect. The non-linear Kaiser effect would not be present on the smallest  $k$ -modes (largest scales), but would become more dominant when going to larger  $k$ -modes. This statement can also be identified in Figure 3.40 because at small  $k$ -modes the ratio of the two boost-factors fluctuates around unity.

The identification of the smallest halos is not dependent on the redshift. The mass scale for the smallest halo is fixed, see Section 2.3, from which the mass function is built up. At higher redshifts, this corresponds to a more non-linear structure compared to lower redshifts. Peculiar velocities are not as linear as they were modeled by the linear Kaiser effect. The additional signal is responsible for the boost which can be seen in Figure 3.40

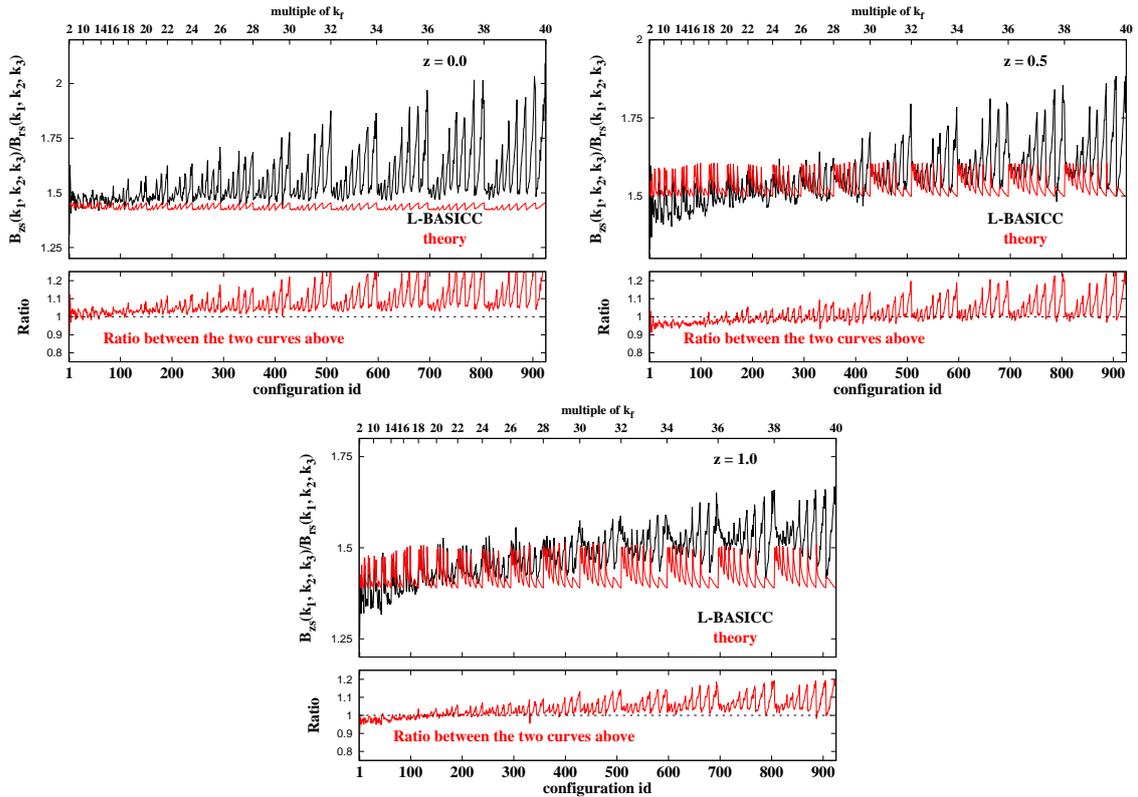


Figure 3.40: The ratio between redshift and real space bispectrum (boost-factor) are plotted against the *configuration id* (lower x-axis) for  $z = 0.0, 0.5$  and  $1.0$  in the top-left, top-right and bottom-middle panel, respectively. The upper x-axis displays the multiples of  $k_f$  in relation with the *configuration id*. Each panel is divided into two sub-panels. The upper sub-panel shows the boost-factor obtained from the L-BASICC simulations in black and the theory in red. In the lower sub-panel the ratio between these two boost-factors from the upper sub-panel is plotted in red.

for the L-BASICC boost-factor and not in the theoretical prediction. The enhanced signal from the non-linear Kaiser effect cannot be suppressed by the FoG because dark matter halos do not have any substructure. Therefore, the situation is contrary to the dark matter case.

### 3.2.4 Application to galaxy mocks: Lyman Alpha Emitters

In this section, the extraction of the linear and the quadratic bias of galaxies will be tested at high redshifts. For this task a not yet published  $N$ -body simulation populated with Lyman Alpha Emitters (LAEs), which are galaxies in the early Universe, will be used. In the first part of this section, the LAEs will be introduced. The second part is focused on the simulation and in the third part the results of  $b_1$  and  $b_2$  estimation will be presented.

### Definitions and explanations of the LAEs

LAEs are galaxies in the early Universe, mostly found at optical to near IR wavelengths at  $z > 2.0$  and up to  $z \approx 7.0$  with a strong Lyman- $\alpha$ -emission line ( $\lambda = 121.6$  nm). The Lyman- $\alpha$ -line is the transition of the electron within a neutral hydrogen atom from the second lowest energy level to the lowest energy level (ground state) and indicates ongoing star formation. The LAEs serve as a tool to understand the high redshift Universe (Nilsson & Meisenheimer, 2009). Galaxies with such a spectral feature are mostly young with a young stellar population and a large gas fraction.

The **H**obby-**E**berly **T**elescope **D**ark **E**nergy **eX**periment (HETDEX) is mainly focused on the understanding of the nature of dark energy (Hill et al., 2008). The LAEs were selected to be tracers of the high redshift matter density field because at such high redshifts they are the best choice to perform spectroscopy (Ciardullo et al., 2012). Deviations from the  $\Lambda$ CDM cosmology will be visible, if dark energy is not close to the cosmological constant at such high redshifts. Combining these results with low redshift surveys like the SDSS (York et al., 2000) will hopefully constraining the cosmological parameters tighter, especially the parameter of the equation of state of dark energy.

At the time the project will be finished, which will be around 2017, it is expected to have detected 0.8 million LAEs within  $1.9 < z < 3.0$  and additionally more than one million OII emitting galaxies at  $z < 0.5$ , which are not of interest in this section. In the following section,  $b_1$  and  $b_2$  will be extracted from two mock catalogs, which were designed for the HETDEX project, by means of the bispectrum.

### The GIPCC simulation

The LAEs were selected to be the tracers of the matter density field for HETDEX. For a better understanding of the clustering signal and galaxy evolution, two N-body simulations with a box size of  $L_{Box} = 1.0 h^{-1}$  Gpc were constructed at  $z = 2.2$  and  $3.0$ . They are called GIPCC, which stands for GigaParsec simulation run at the Institute for Computational Cosmology (Baugh et al., 2012, in preparation). A flat  $\Lambda$ CDM cosmology with  $\Omega_M = 0.26$ ,  $H_0 = 71.5 \frac{km}{s \text{ Mpc}}$  and a  $\sigma_8 = 0.8$  was chosen for the simulation. By adopting GALFORM, a semi-analytic galaxy formation code, developed by Cole et al. (2000) 2 871 540 and 1 584 333 LAEs at  $z = 2.2$  and  $3.0$ , respectively, were artificially formed in the GIPCC simulation.

Because only one realization for each redshift is available, the calculation of the errors is not feasible and must be performed theoretically. For the bispectrum estimator given in Section 2.5, this is possible and an estimate of the variance (and therefore for the errors) can be evaluated under the assumption that the Fourier components are Gaussian distributed (Scoccimarro et al., 1998; Sefusatti et al., 2006; Sefusatti & Komatsu, 2007). Then, the variance of the bispectrum  $var(B_{zs}(\vec{k}_1, \vec{k}_2, \vec{k}_3))$  is estimated by

$$var(B_{zs}(\vec{k}_1, \vec{k}_2, \vec{k}_3)) = s_B \frac{k_f^3}{V_B} P_{est}(k_1) P_{est}(k_2) P_{est}(k_3) \quad (3.6)$$

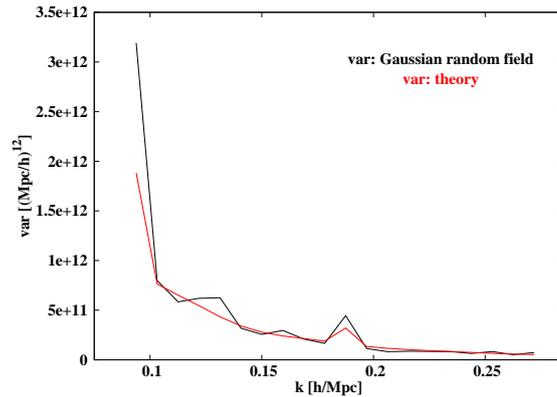


Figure 3.41: The variance calculated from the Gaussian random field (black line) is compared to the theoretical estimation (red line) given by Equation (3.6).

with  $s_B = 1, 2, 6$  for general, isosceles and equilateral triangles, respectively, and  $V_B$  from Equation (2.33) is defined as

$$V_B \simeq 8\pi^2 k_1 k_2 k_3 \delta k^3 . \quad (3.7)$$

The quantity  $s_B$  indicates the contribution of different triangle configurations to the variance of the bispectrum (Sefusatti et al., 2006).

In Figure 3.6, Equation (3.6) is plotted in red and is compared to the estimation of the variance from the Gaussian density field (black solid line) which was used in Section 2.4. It can be noticed, that the two curves reproduce the same trends and it can also be assumed, that the variance of the bispectrum can be estimated correctly by Equation (3.6) and will be used for the following analysis.

The spike at around  $0.18 h \text{ Mpc}^{-1}$  is originated from the isosceles configuration (the  $s_B$ -factor changes from unity for general configurations to two for isosceles configurations). Another isosceles configuration is present at  $k \approx 0.09 h \text{ Mpc}^{-1}$  but only the falling wing can be seen.

### Results of the LAEs

Before discussing the results, the best-fit bias parameters of  $b_1$  and  $b_2$  will be used to compare the theoretical bispectrum to the GIPCC bispectrum. As for the dark matter case, the ratio of them will be evaluated and are shown in the Figures 3.42 and 3.44 for  $z = 2.2$  and  $z = 3.0$ , respectively. Similar to previous discussions here, the black solid line shows the ratio with the included collapsed triangles and for the blue solid line they are excluded. The real and redshift space ratios are plotted in the left and the right panel. The random motions are not shown for the ratios because their mild effect on the bispectrum can hardly be noticed in such a plot.

The results for  $b_1$  and  $b_2$  are shown in Figure 3.43 and 3.45 for  $z = 2.2$  and  $3.0$ , respectively, where again the estimation of  $b_1$  can be found on the left panel and  $b_2$  on the right panel. Real space results are shown in blue, redshift space results in red or black depend-

ing on the exclusion or inclusion of the random motions. In Figure 3.46, the extracted values of the velocity dispersions  $\sigma_v$  are plotted, where  $z = 2.2$  is given in black and  $z = 3.0$  in red. In Table 3.8, the results of this discussion will be summarized for both redshifts.

### Discussion of the LAEs bispectrum at $z = 2.2$

The comparison between the GIPCC bispectrum and the theoretical bispectrum is shown in Figure 3.42. It can be noticed that over all considered  $k$ -modes the ratios fluctuate around unity with deviations up to fifty percent. The distinction of the equilateral and the collapsed triangle configurations cannot be made by just looking at the peaks of the ratios (positive and negative peaks, respectively). The fluctuations are such large that this distinction requires the x-axis information. The reason for this behavior is the dominant effect of cosmic variance on the considered scales (because only one realization per redshift is available).

For  $z = 2.2$ , the linear bias can be estimated to be  $b_1 \approx 2.10$  up to  $k_{max} \approx 0.19 h \text{ Mpc}^{-1}$ , which is the largest side length of the triangle considered in this analysis. Only small de-

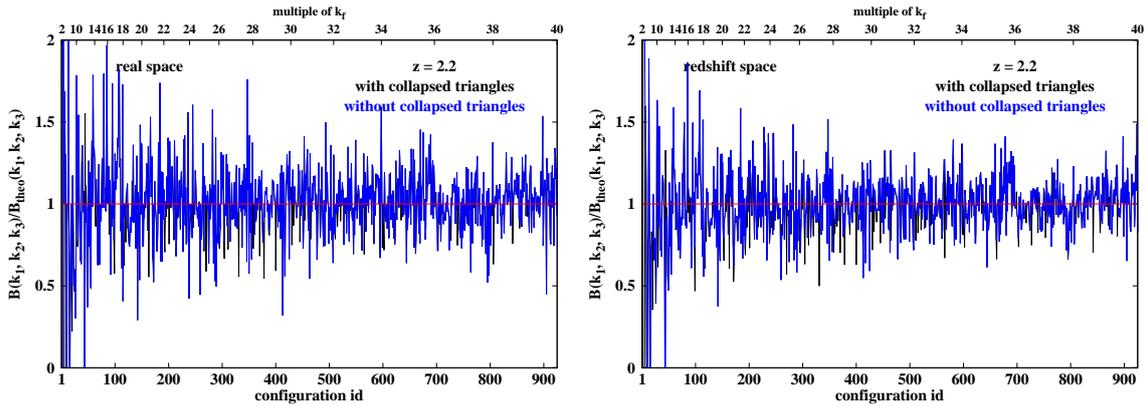


Figure 3.42: Same description as for Figure 3.34, but for the LAEs bispectrum at  $z = 2.2$

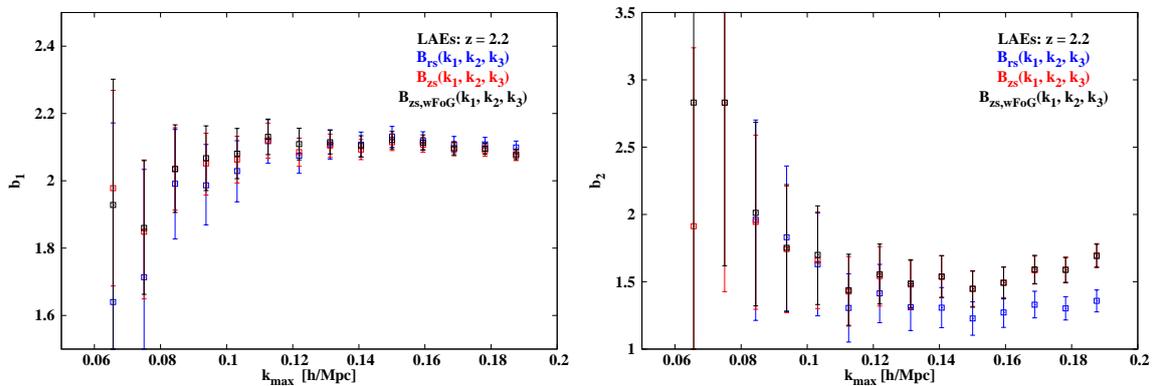


Figure 3.43: Same description as for Figure 3.37, but for  $B_{LAEs}(k_1, k_2, k_3)$  at  $z = 2.2$

viations can be noticed between real and redshift space. Including random motions let the results unchanged. Independent on the *coordinate frame* it seems that smaller  $b_1$ -values are extracted on largest scales, but the errors on these estimations are so huge, that a clear conclusion is not feasible.

The situation for  $b_2$  is more complicated than for  $b_1$ . A conservative estimation of the  $b_2$  parameter is only possible up to  $k_{max} \approx 0.12 h \text{ Mpc}^{-1}$ . It looks as if larger  $b_2$ -values are estimated at large and small scales whereas smaller  $b_2$ -values can be extracted from the scales between, see Figure 3.43. A similar trend seems to be present on large scales for extracting  $b_1$  but cannot be quantified due to large errors. On smaller scales, this is a hint towards non-linear structure growth. However, real space seems to be less affected than redshift space, but this behavior was already discussed in the sections above.

#### *Discussion of the LAEs bispectrum at $z = 3.0$*

The situation for  $z = 3.0$  is comparable to the situation at  $z = 2.2$ . The GIPCC and the theoretical bispectrum are in agreement with each other as it can be seen by the ratio of these two quantities which is plotted in Figure 3.44. The description of this figure follows Figure 3.42 but for  $z = 3.0$ . In real and redshift space, an increase of the ratio ( $> 1$ ) can be noticed when going to larger  $k$ -modes.

In Figure 3.45, a small bump can be identified for the  $b_1$  estimation which is not present at  $z = 2.2$ . From around  $k_{max} \approx 0.14 h \text{ Mpc}^{-1}$  to  $0.17 h \text{ Mpc}^{-1}$  this bump seems to be a local maximum. If this bump was not be present, the linear bias could be estimated to be  $b_1 \approx 2.80$ , as shown in Table 3.8.

The reasons for this effect are not known but the bump is present in real and redshift space. This bump could be explained by the presence of a large structure within the catalog. However, it is very unlikely that the same feature is not visible in the examination of  $b_1$  at  $z = 2.2$  because gravity has only a limited amount of time to act between the two considered redshifts, only 890 million years, and it is expected that such a structure would accrete more mass. Therefore, this bump would be more pronounced in the low redshift

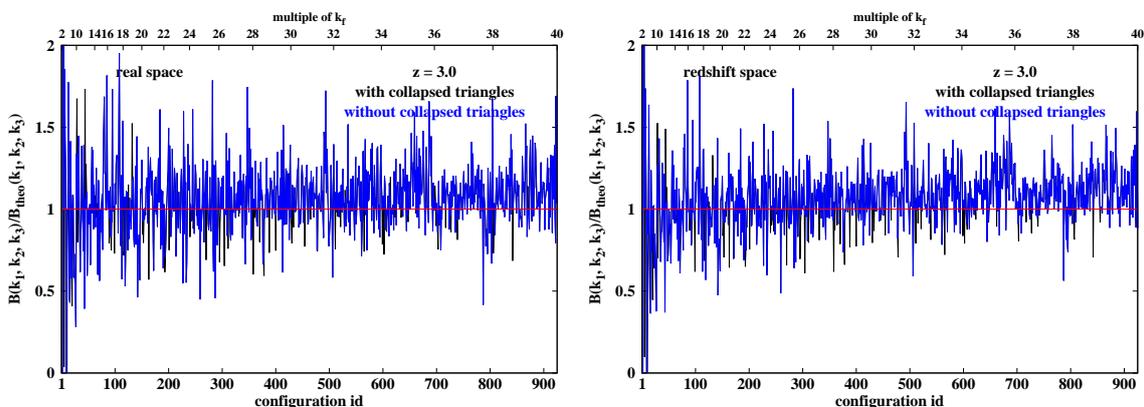


Figure 3.44: Same description as for Figure 3.42, but for  $z = 3.0$

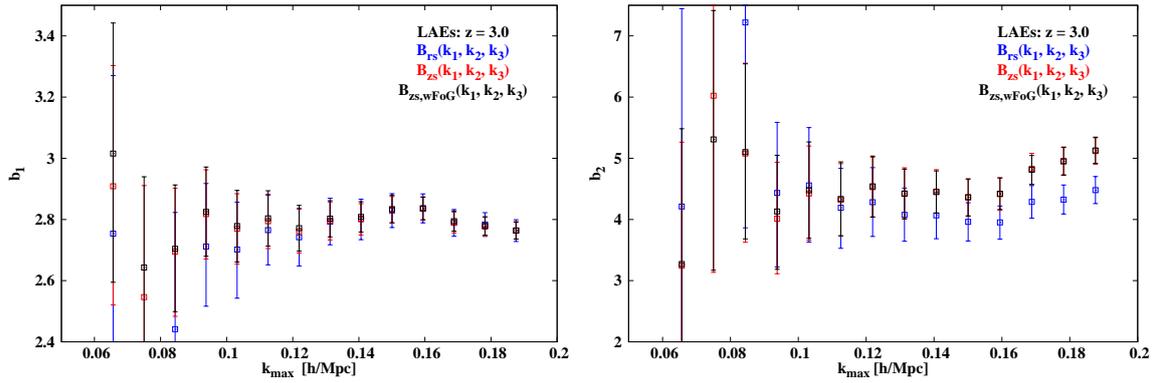


Figure 3.45: Same description as for Figure 3.43, but for  $z = 3.0$

realization.

Around that  $k$ -modes the estimates of  $b_2$  seems to be constant around 4.30 and at  $k_{max} \approx 0.16 h \text{ Mpc}^{-1}$  an increase of the extracted  $b_2$ -values can be noticed. However, a consistent estimation in real and redshift space is only possible up to  $k_{max} \approx 0.12 h \text{ Mpc}^{-1}$ .

$z_{obs}$	$(b_1, k_{max} [h \text{ Mpc}^{-1}])$	$(b_2, k_{max} [h \text{ Mpc}^{-1}])$
2.2	$(2.10 \pm 0.03, 0.19)$	$(1.37 \pm 0.36, 0.12)$
3.0	$(2.80 \pm 0.12, 0.13)$	$(4.30 \pm 0.88, 0.12)$

Table 3.8: Same description as for Table 3.7, but for the LAEs and at  $z = 2.2$  and 3.0

### *Bispectrum vs. power spectrum: Comparison of the extracted bias parameters*

In Section 2.6.1, the work of Jeong & Komatsu (2009) was introduced in which the bias parameters  $b_1$  and  $b_2$  were extracted from the galaxy power spectrum. The proposed method was also performed on the GIPCC catalogs (priv. comm. Jeong, 2011). The results of this analysis are given in Table 3.9. Unfortunately no errors were calculated for these estimates but it is known from Jeong & Komatsu (2009) that  $b_2$  is weakly constrained by this method (Jeong & Komatsu, 2009).

$z_{obs}$	$b_1$	$b_2$
2.2	2.20	1.54
3.0	2.90	2.90

Table 3.9: The extracted bias parameters  $b_1$  and  $b_2$  from LAEs power spectrum  $P_{LAEs}(k)$  are given in column 2 and 3, respectively, for the two considered redshifts which are listed in column 1.

The bias parameters extracted from the LAEs bispectrum  $B_{LAEs}(k_1, k_2, k_3)$  are in agreement with the results from  $P_{LAEs}(k)$ . The estimates for  $b_1$  at  $z = 2.2$  and  $b_2$  at  $z = 3.0$  are exceptions because they seem not to be consistent with the bispectrum results. However, without an estimate on the errors of these quantities a conclusive statement is not possible.

#### *Discussion of the results of the extracted velocity dispersion*

As expected from the previous discussions, the random motions do not have a noticeable effect. The estimation of  $\sigma_v$  in Figure 3.46 shows only a very slight difference between  $z = 2.2$  and  $z = 3.0$ . From the findings of this section it can be stated, that  $\sigma_v$  is only weakly dependent on the redshift. On the smallest  $k$ -modes, no difference between the two redshifts can be noticed. At  $k_{max} \approx 0.08 h \text{ Mpc}^{-1}$  to  $0.12 h \text{ Mpc}^{-1}$   $\sigma_v$  is estimated to be 1.8 for both redshifts. On smaller scales  $\sigma_v$  seems to be slightly larger at  $z = 2.2$  than at  $z = 3.0$ , as expected. However, the extracted velocity dispersions are very small, almost in agreement with zero, which explains their insignificant effect on the bias parameters estimates in redshift space.

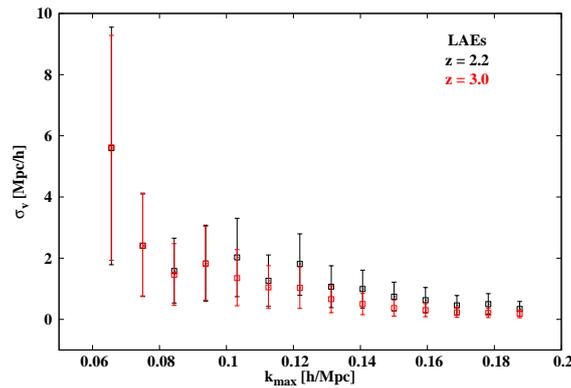


Figure 3.46: The estimated  $\sigma_v$ -values are summarized in this figure. The results obtained for the redshift  $z = 2.2$  and  $z = 3.0$  are represented in black and red, respectively.

#### *Discussion of the non-linear Kaiser effect*

In Figure 3.47, the ratios between the redshift space bispectrum  $B_{zs}(k_1, k_2, k_3)$  and the real space bispectrum  $B_{rs}(k_1, k_2, k_3)$ , also called (Kaiser) boost-factor, for the GIPCC simulations and the theory are plotted. The boost-factors for  $z = 2.2$  and  $z = 3.0$  are plotted on the left and the right panel, respectively. Each of these panels is divided into two sub-panels. In the upper sub-panel the GIPCC boost-factor is illustrated in black and the theoretical prediction in red. In the lower sub-panel the ratio between the GIPCC and the theoretical boost-factor is shown. It can be noticed that there are not any trends present as it was for the L-BASICC dark matter halo ratios. The expected increase of the signal when going from small to large  $k$ -modes is not present as for the L-BASICC dark matter

halos. It seems that the linear Kaiser effect is able to model the redshift space distortions at these redshifts on the considered  $k$ -modes. Therefore, the non-linear Kaiser effect is not required anymore on the examined  $k$ -modes. This would also explain the agreement between real and redshift space for the estimates of  $b_1$  and  $b_2$ .

It can be concluded that the estimation of these bias parameters can be performed at high redshifts and for LAEs as tracers for the underlying matter density field. Therefore, in a future analysis extracting  $b_1$  and  $b_2$  from such a data set can be carried out.

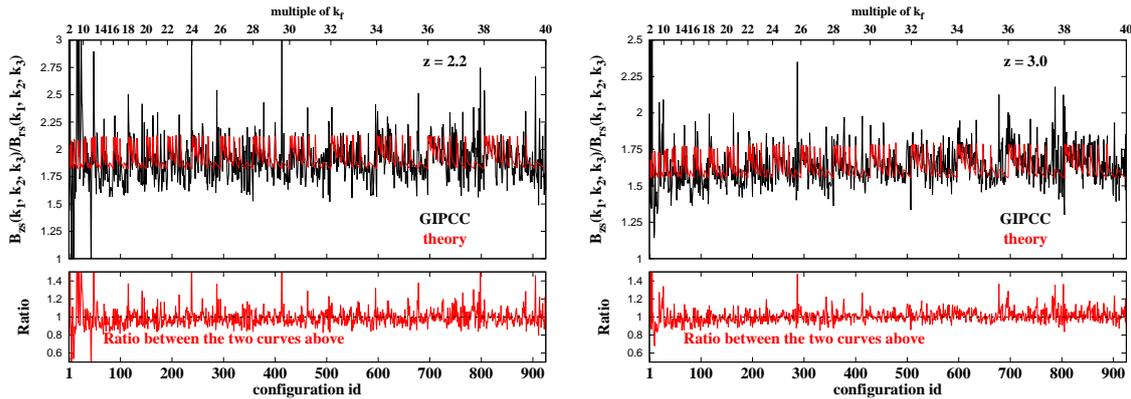


Figure 3.47: Ratio between the redshift space bispectrum  $B_{zs}(k_1, k_2, k_3)$  and the real space bispectrum  $B_{rs}(k_1, k_2, k_3)$  (also called boost-factor) for the LAEs at  $z = 2.2$  and  $3.0$  are plotted on the left and right panel, respectively. Each panel contains two sub-panels. The upper sub-panel shows the GIPCC ratio in black and the theoretical prediction in red. In the lower sub-panel the ratio between the two boost-factors is shown where the dotted black line indicates perfect match between the two quantities.

### 3.2.5 Other work in the literature

As mentioned previously, the three-point statistics is not as thoroughly explored as the two-point clustering. Nevertheless, over the last ten years some approaches were proposed for modeling the bispectrum. In this section, only the most recent publications will be named. In most cases a direct comparison is not possible in redshift space because the bispectrum was only investigated in real space.

In Scoccimarro et al. (1998) and Sefusatti et al. (2011), it was shown that at least in real space and at high redshifts the 1-loop corrections, corresponding to  $6^{th}$  order perturbation theory, could provide a better description of non-linear structure growth. However, these publications probed their models only on large scales ( $k < 0.15 h \text{ Mpc}^{-1}$  for  $z = 1.0$ ), on which the tree-level approach can still be used reliably and redshift space was not investigated. For a consistent treatment of the redshift space bispectrum the full  $6^{th}$  order perturbation theory should be considered in the derivation. Until now, such an approach was not examined.

In Pan et al. (2007), scale transformations were applied to the matter tree-level bispectrum.

They extended this method from the power spectrum to the bispectrum and concluded that even at low redshifts the bispectrum can be well modeled. Redshift space was not examined in their work. Therefore, it is not clear if the linear Kaiser effect can be as easily incorporated as proposed in this thesis. They also claim that their ansatz is comparable to 1-loop corrections at least at high redshifts and refer to the similarity of the reduced bispectrum. However, it was concluded in this thesis, that the reduced bispectrum cannot be modeled correctly.

In Smith et al. (2008), the bispectrum was derived for the halo model (Cooray & Sheth, 2002) in real and redshift space. Their investigation was mainly focused on the modeling for the U-shape representation itself. No parameters were fitted. They found that on considerably large scales ( $k \gtrsim 0.10 h \text{ Mpc}^{-1}$ ) the 2-halo term starts to contribute already a non-vanishing signal to the reduced bispectrum. For the configuration  $k_1 = 0.10 h \text{ Mpc}^{-1}$  and  $k_2 = 0.20 h \text{ Mpc}^{-1}$  this can be up to 10%. It can be assumed that the same is true for  $B(k_1, k_2, \theta_{12})$  because of the calculation of the reduced bispectrum. The reduced signal from the 2-halo term due to the division by the power spectra is not present for the calculation of the bispectrum and therefore this signal is not reduced. The model of Smith et al. (2008) is in agreement with the simulation at least on the large scales in real space but in redshift space the same problems occur as discussed in this work. Their redshift modeling includes linear Kaiser effect and FoG as in this thesis but no explanation for the mismatch was given there.

Improving the theory of mode-coupling is one possibility to extend the valid range of a model. In Simpson et al. (2011), another approach was chosen. The most non-linear part, the highest density peaks, of the data will be clipped away and these high density peaks are set to a predefined maximum value. The resulting bispectrum is not dominated by these extreme structures anymore and can reliably be used up to  $0.7 h \text{ Mpc}^{-1}$  at  $z = 0.0$  for the Millenium simulation (Springel et al., 2005) but only in real space. Their approach is comparable to the situation when only low mass halos of the L-BASICC catalogs were utilized for estimating  $b_1$  and  $b_2$ . In this case, no information of the massive halos would be used. If the density field is clipped, there will be still some information of the unclipped density field present. As for the previous investigations, redshift space was not examined in Simpson et al. (2011). They know that their method extracts not the real bias parameters because of the clipping. They claim that this change of the bias is not necessarily important because the extracted matter power spectrum is close to the linear one. Therefore, an application to a real data set, where one is interested in the real bias parameters, is still not possible.

Similar results of the reduced bispectrum which were discussed in this section were also found in Pollack et al. (2012) where they claimed that the locality assumption of Fry & Gaztañaga (1993) is violated. Of course, this would lead to a different biasing in the two- and three-point statistics. They smoothed the density field and extracted the linear and the quadratic bias from the bispectrum and directly from the density field. They stated that consistent results of these bias parameters with the original tree-level ansatz can be achieved with a smoothing scale of  $R \approx 20 h^{-1} \text{ Mpc}$ . This smoothing scale is not known a priori and set as a free fit parameter. So far, their investigation was only performed in

real space. They also found that higher order corrections are required for recovering the bias parameters.

In Sefusatti et al. (2006), an attempt to extract cosmological parameters from the bispectrum was examined. They modeled deviations between the mock catalogs generated by PTHALOS (Scoccimarro & Sheth, 2002) and the 2<sup>nd</sup> order Lagrangian perturbation theory (Scoccimarro, 2000) simulations with the original tree-level approach. With this method they are able to use the bispectrum up to  $0.3 h \text{ Mpc}^{-1}$  in real and redshift space. However, this is not a full modeling because it requires the presence of  $N$ -body simulations to which the measurement can be compared. It was concluded that the Likelihood-functions are almost the same as for the power spectrum. Only a combined analysis on the two clustering analyses will be able to give tighter constraints on cosmological parameters.

In this thesis, the modeling of the dark matter bispectrum from first principles was not only examined in real space but also in redshift space. In contrast to the above discussed publications this is mostly not the case. An additional contribution of the analysis in this thesis is not only the extended investigation of the bispectrum in redshift space but also the wide range of examined triangle configurations (all configurations up  $k_{max} \lesssim 0.19 h \text{ Mpc}^{-1}$ ).



# Chapter 4

## Summary and Conclusions

This chapter is divided into three parts. In Section 4.1, the results obtained from two- and three-point statistics will be summarized. Possible improvements on the modeling of the anisotropic two-point correlation function  $\xi(r_p, \pi)$  and the bispectrum  $B(k_1, k_2, k_3)$  will be discussed in Section 4.2 in order to extend the range of validity of the model. This thesis will be concluded with Section 4.3 where possible future projects for real data sets will be proposed.

### 4.1 Summary

In this section, the results from Chapter 3 will be summarized. In Section 4.1.1, the results on the constraints of the equation of state parameter of dark energy  $w_{DE}$  and the linear bias  $b$ , from the anisotropic two-point correlation function  $\xi(r_p, \pi)$  will be presented. This review is from Schlagenhauer et al. (2012). In Section 4.1.2, the estimates of linear bias  $b_1$  and quadratic bias  $b_2$  from the bispectrum  $B(k_1, k_2, k_3)$  will be recapitulated. In order to distinguish between the two estimations of the linear bias, they will be named differently. The linear bias  $b$  for  $\xi(r_p, \pi)$  is defined by Equation (2.75) whereas the linear bias  $b_1$  for the bispectrum is given by the first parameter of the series expansion given in Equation (2.77). From these two definitions it is clear that the  $b$ -variable is an effective quantity which contains contributions from all  $b_n$ -parameters of the series expansion.

#### 4.1.1 Anisotropic two-point correlation function: Estimation of $w_{DE}$ and $b$

In this thesis, a model of  $\xi(r_p, \pi)$  was developed and tested, and was used to investigate the influence of photometric redshift errors on the measurement of  $w_{DE}$ .  $\xi(r_p, \pi)$  was modeled using  $3^{rd}$  order perturbation theory (Jain & Bertschinger, 1994) to account for the non-linear nature of the growth of structure and the non-linear Kaiser effect (Scoccimarro, 2004). Redshift errors can be included in the model by convolving it with the pairwise redshift error distribution, which can easily be computed from the (known) photometric

redshift errors.

In order to test the validity of the model for  $\xi(r_p, \pi)$ , it was fit to the mean measured correlation function of dark matter halos in a suite of 50 large-volume, medium-resolution  $N$ -body simulations (the L-BASICCS II later just L-BASICC, (Angulo et al., 2008; Sánchez et al., 2008)). Both in real and redshift space the fit yields unbiased values of the dark energy equation of state parameter  $w_{DE}$  and the linear bias  $b$ . With approximately 300 000 halos per box, in real space  $w_{DE}$  and  $b$  can be determined with an accuracy of about 12% and 7%, respectively. In redshift space, these constraints become slightly weaker,  $w_{DE}$  can be measured with an accuracy of approximately 15%, and the relative error of  $b$  becomes  $\sim 8\%$ .

If only the shape of  $\xi(r_p, \pi)$  is used to infer  $w_{DE}$  and  $b$ , the errors on both will increase as there is significant information contained in the amplitude. The relative error of the bias increases more than the relative error of  $w_{DE}$ , since the value of the bias is mainly encoded in the amplitude (and less in the quadrupole and hexadecapole contribution to the redshift space distortions), whereas the equation of state parameter of dark energy influences both shape and amplitude likewise.

In order to investigate the effect of redshift errors on the measurement, a small offset was added to one of the coordinates of the dark matter halos, which was drawn randomly from a Gaussian error distribution. The model was convolved with the corresponding pairwise redshift error distribution in the direction along the line-of-sight ( $\pi$ ). Redshift errors smear out the clustering signal and diminish its amplitude; at the same time the convolution leads to a mixing and increase of the noise of the measurement in single pixels, because intrinsic errors are also distributed along the line-of-sight. On the constraints on cosmological parameters, the impact of this is two-fold: Since the signal of the **B**aryonic **A**coustic **O**scillations (BAOs) (as the main feature of the otherwise smooth correlation function) becomes weaker in the observed range of scales, its predictive power decreases – in the case of very large redshift errors ( $\sigma_z > 0.06$ ) the signal is smeared out over such a large range of scales that it completely disappears in the noise. However, since much higher accuracies can be achieved in realistic ongoing or near-future photometric surveys such as e.g. Pan-STARRS (see Saglia et al. (2012)), this is not a cause for concern.

Integrating  $\xi(r_p, \pi)$  to obtain  $w(r_p)$ , as originally proposed by Peebles (1980) as a means to overcome redshift space distortions, does not help to improve the constraints, as in real space the BAO is a *ring* in the  $\xi(r_p, \pi)$  plane, and, when integrated, is distributed over  $0 \leq r_p \lesssim 120 h^{-1} \text{Mpc}$ . Since it is impossible to integrate  $\xi(r_p, \pi)$  to  $\pi = \infty$ , the resulting amplitude and shape of  $w(r_p)$  depends on the choice of integration limits as well as the underlying cosmology, which adds a further complication. Secondly, the noise itself increases in the presence of redshift errors, which creates an additional difficulty. Due to the decreased signal-to-noise of the two-point correlation function, the accuracy of the constraints on  $w_{DE}$  and  $b$  decreases.

In order to beat down systematics coming from cosmic variance (which is still large, even on BAO scales), it is desirable (and important) to observe the largest volumes possible at one particular redshift. Also, in order to measure a possible variation in the equation of state with look-back time, observations have to be carried out at higher redshifts as

well. At this moment in time both are still only feasible with photometric redshifts. The anisotropic two-point correlation function  $\xi(r_p, \pi)$ , which can be used to infer cosmological parameters like  $w_{DE}$ , is well suited to incorporate photometric redshifts. The model of  $\xi(r_p, \pi)$  which was developed in this thesis will be able to provide *unbiased* constraints on  $w_{DE}$  and  $b$  for photometric redshift surveys. The maximum redshift error for which this model will work certainly depends on the exact shape of the redshift error distribution, the volume, and number density of the survey to which it is applied.

### 4.1.2 Bispectrum: Bias parameter estimation

In this thesis, a model for the three-point clustering in Fourier space, the so-called bispectrum, was developed and tested. The bispectrum was modeled including a modification of the tree-level ansatz in which the usually used linear power spectra, as in Bernardeau et al. (2002), were replaced by the corresponding  $3^{rd}$  order perturbation theory power spectra in order to account for the non-linear growth of structure. In the later discussion, this model will be called modified tree-level ansatz. This modification was also applied to redshift space which was modeled including the linear Kaiser effect for three-point statistics (Heavens et al., 1998). The incorporation of redshift errors as for  $\xi(r_p, \pi)$  was not considered due to the loss of information by the smearing of the clustering signal along the line-of-sight (see Section 2.7.3).

The validity of the bispectrum modeling was tested by extracting the linear bias  $b_1$  and the quadratic bias  $b_2$  from the dark matter L-BASICC simulations by fitting the model to the mean of this suite of 50 large-volume, medium-resolution numerical  $N$ -body simulations. Throughout this investigation it was concluded that the modified tree-level ansatz is indeed better suited to describe the non-linear growth of structure compared to the original tree-level ansatz.

In a first analysis, only the equilateral triangle configurations were examined from the dark matter simulations in real and redshift space. The big advantage of this configuration is that all sides of the triangles have the same size. The breakdown of the model can be related directly to one specific  $k$ -mode by investigating the deviations between the model and the measurement. For more general configurations, where all sides of the triangle have different sizes, such a clear relation is no longer possible. It was concluded that for  $z = 0.5$  and  $z = 1.0$  a correct modeling of the bispectrum from the modified tree-level ansatz can be achieved up to  $0.10 h \text{ Mpc}^{-1}$  and  $0.15 h \text{ Mpc}^{-1}$ , respectively. The  $z = 0.0$  case was excluded from this analysis because of the low gain of information for that specific configuration.

As a next step the bias parameters  $b_1$  and  $b_2$  were extracted from the dark matter halo L-BASICC catalogs by means of the bispectrum for equilateral configurations. It was concluded that the content of information available from this analysis was too little in order to obtain well determined bias parameters.

Therefore, the extraction of  $b_1$  and  $b_2$  was performed for all triangle configurations at  $z = 0.0, 0.5$  and  $1.0$ , except for collapsed configurations. For this analysis it was decided to consider only configurations for which each side ( $k$ -mode) of the triangle can be modeled

correctly. The analysis was carried out on the dark matter as well as on the halo catalogs of the L-BASICC simulations. The limitations of the model, that is up to which  $k$ -modes the bispectrum can be correctly modeled, as obtained from the equilateral configurations, were confirmed by the investigation of all triangle configurations. It was found that within the weakly non-linear regime the estimated bias parameters are in agreement with the expected value for the dark matter case. The bias parameters can also be extracted consistently in real and redshift space for the dark matter halos within the weakly non-linear regime.

Beyond the weakly non-linear regime at larger  $k$ -modes the theoretical bispectrum cannot be modeled correctly anymore. The observed deviations between the model and the measurement have two different origins.

First, the growth of structure cannot be described correctly by the modified tree-level ansatz anymore. This leads to an underestimation of the theoretical bispectrum compared to the measurement. If the ratio between these two bispectra is computed, the deviations of the ratios will vary within a range from 0.6 to 1.4. The largest deviations below and above unity can be identified as collapsed and equilateral configurations, respectively.

Second, the redshift space distortions are not as linear as is assumed by the linear Kaiser effect. For the dark matter case, this additional increase of the bispectrum is damped by the peculiar velocities of particles in bound structures which leads in fact to a reduction of the bispectrum. For the extraction of the bias parameters  $b_1$  and  $b_2$  the velocity dispersion of the particles  $\sigma_v$  was introduced as an additional free fit parameter to account for this effect. It was concluded that only a mild improvement can be achieved by taking them into account.

The dark matter halos do not contain any substructure because they are just point-like objects in the catalogs. Therefore, no FoG effect is present there and a damping as described for the dark matter case cannot be identified. In fact, a contrary trend can be noticed. By examining the ratios between the real and redshift space bispectra from the L-BASICC simulations and the theoretical predictions, a larger signal than expected can be observed. It was concluded that this is a manifestation of the non-linear Kaiser effect. On top of this, the modeling of the redshift space bispectrum benefited even less from the inclusion of the large scale random motions than for the dark matter case.

It was also noticed that the bias parameters  $b_1$  and  $b_2$  are degenerate. They try to minimize the differences between the model and the measurement which leads to an under- as well as an overestimation of  $b_1$  and  $b_2$ , respectively. This happens because these bias parameters contribute to different terms of tree-level ansatz which affect the shape of the bispectrum differently. These different shape dependencies are responsible for the extraction of  $b_1$  and  $b_2$  by means of the bispectrum as well as for the introduction of the above mentioned degeneracy in order to account for the effect of the insufficiently modeled non-linear structure growth and peculiar velocities beyond the weakly non-linear regime.

The last investigation of the bispectrum was an application to the Lyman- $\alpha$  emitting galaxies in the high redshift universe. The model was tested against the GPICC simulations (Baugh et al., 2012, in preparation) in which Lyman- $\alpha$  emitting galaxies were artificially created by GALFORM (Cole et al., 2000) for  $z = 2.2$  and  $z = 3.0$ . It was concluded that for such a data set the proposed model in this thesis is able to extract consistent bias

parameters in real and redshift space up to  $0.12 h \text{ Mpc}^{-1}$ . For  $z = 2.2$  the extracted bias parameters are  $b_1 = 2.10$  and  $b_2 = 1.37$  and for  $z = 3.0$  the results are  $b_1 = 2.80$  and  $b_2 = 4.30$ .

## 4.2 Improvements of the model for the bispectrum

During the discussions in Chapter 3, it was stated that a more accurate description of the non-linear structure growth and the redshift space distortions would be required to extend the model to larger  $k$ -modes. On the largest scales the model introduced in this thesis is able to extract the bias parameters  $b_1$  and  $b_2$  correctly. However, including smaller scales could place tighter constraints on these estimates. The main ingredient of the theoretical model is  $2^{\text{nd}}$  order perturbation theory for the bispectrum. It was concluded that for high redshifts and on large scales the modified tree-level ansatz is sufficient to extract  $b_1$  and  $b_2$ . The next logical step to extend the range of validity of the model is a more accurate theoretical description of the non-linearities. Some promising approaches are available but still not tested in detail. Moreover, their calculation is very time consuming due to the computation of the mode-coupling integrals (Scoccimarro et al., 1998). Also mentioned previously, for the calculation of the analytic model some assumptions were required, e.g. neglecting the pressure and vorticity. It is also clear that for  $z < 1.0$  some of these assumptions are not valid and more realistic descriptions are needed. There are two obvious improvements for the analytic model which should be investigated:

- complete  $6^{\text{th}}$  order perturbation theory should be used for the bispectrum instead of  $2^{\text{nd}}$  order perturbation theory with the replacement of the linear by the  $3^{\text{rd}}$  order perturbation theory power spectrum in Equation (2.78)
- the non-linear Kaiser effect should be derived for the bispectrum instead of using the linear Kaiser effect for the redshift bispectrum (Heavens et al., 1998)

The suggested improvement for the calculation of redshift space distortions was already implemented for  $\xi(r_p, \pi)$ . In order to avoid dealing with a perturbative series, the clustering statistics can be derived using renormalized perturbation theory. The results of the power spectrum are given in Crocce & Scoccimarro (2006) and the 1-loop corrections were already adopted to extract the cosmological parameters (Sánchez et al., 2006; Montesano et al., 2011; Sánchez et al., 2012).

First attempts to test the renormalized perturbation theory for the bispectrum were described in Bernardeau et al. (2008). Unfortunately, they only tested the model against dark matter simulations in real space for one specific triangle configuration at large scales,  $k_1 = 0.068 h \text{ Mpc}^{-1}$ ,  $k_2 = 2 \times k_1$  and varying  $k_3$  at  $z = 0.0$  for the reduced bispectrum. It must be stated that the match between the  $N$ -body simulation, which they used, and the proposed model is promising. An improvement compared to the tree-level approach is obvious. It can be expected that models at higher redshifts benefit even more from this approach, at least in real space. However, the model is tested only for one specific triangle

configuration. In this thesis, it was noticed that even for  $k$ -modes where the modified tree-level ansatz is not able to describe the coupling between different modes properly, it is possible to find well described triangle configurations. On top of that, redshift space was not investigated, which is known from this thesis must also be treated very carefully. Even if it is very time consuming to obtain a better modeling for the bispectrum, the power spectrum will benefit from this effort.

### 4.3 Application to data sets

The main parts of this thesis were feasibility studies: First, on the estimation of the equation of state parameter of dark energy  $w_{DE}$  and the linear bias  $b$  by means of  $\xi(r_p, \pi)$  and, second, on the bias parameters  $b_1$  and  $b_2$  by means of the bispectrum. The next step should be to improve these models and then to carry out all the analyses introduced and discussed in this thesis using real data sets. Both models, for the two- and three-point statistics, are applicable to the large scale structures traced by dark matter halos, as concluded in Chapter 3.

In Section 4.3.1 possible applications to real data sets for the anisotropic two-point correlation function will be proposed. Similar projects will be suggested for the bispectrum in Section 4.3.2.

#### 4.3.1 Anisotropic two-point correlation function

Spectroscopic surveys like the **B**aryonic **O**scillations **S**pectroscopic **S**urvey (BOSS) (Eisenstein et al., 2011; Dawson et al., 2012) are required for an application of the three-point statistics. BOSS is one of the four surveys of the SDSS III project, which targets luminous red galaxies to measure the BAOs mainly using two-point statistics in order to infer cosmological parameters. Expertise has been gained with two-point statistics, e.g. properly taking the survey geometry and selection function into account. A study of  $\xi(r_p, \pi)$  would benefit directly from this expertise and it is expected that cosmological parameters should be better constrained due to the large content of information. Measurements of three-point statistics suffer from the same problems for which a careful handling is required.

In this work, the anisotropic two-point correlation function  $\xi(r_p, \pi)$  was analyzed with respect to the extraction of  $w_{DE}$  and  $b$  from photometric redshift surveys. DES (Tucker et al., 2010) and PAU (Benítez et al., 2009) are still under construction and their data are not yet available. An investigation for the photometric redshifts of Pan-STARRS was already performed (Saglia et al., 2012). It was concluded that the redshift error distribution can be assumed to be Gaussian with a rms of  $\sigma \approx 0.05$ . Therefore, it should be possible to apply  $\xi_{zerr}(r_p, \pi)$  analysis to Pan-STARRS data in order to determine unbiased results on  $w_{DE}$  and  $b$ , given that  $\sigma_z \lesssim 0.06$  (the limiting accuracy requirement, see Section 3.1.2).

### 4.3.2 Bispectrum

Using a mock catalog, it was shown in Section 3.2.4 that the model introduced in this thesis will be able to extract the linear bias  $b_1$  and the quadratic bias  $b_2$  consistently in real and redshift space at high redshifts using Lyman- $\alpha$  emitting galaxies as tracers of the underlying matter density field. These galaxies can be found at optical to near IR wavelengths at  $z > 2.0$  up to  $z \approx 7.0$  with a strong Lyman- $\alpha$ -emission line (Nilsson & Meisenheimer, 2009). The **Hobby-Eberly Telescope Dark Energy eXperiment** (HETDEX) uses these galaxies as tracers of the high redshift matter density field to understand the nature of dark energy (Hill et al., 2008). The bias parameters of these galaxies should be known in order to obtain the true  $w_{DE}$  by means of the power spectrum for comparably large  $k$ -modes. This will also help to extract tight constraints on other cosmological parameters. In this thesis, it was concluded that the bispectrum can be reliably modeled up to  $0.12 h \text{ Mpc}^{-1}$  for high redshift Lyman- $\alpha$  emitting galaxies. Therefore, the bispectrum is perfectly suited for the extraction of  $b_1$  and  $b_2$  for a galaxy survey like HETDEX which in the end will hopefully help to distinguish between different models which try to explain the late time accelerated expansionary phase of the Universe.



# Appendix A

## Appendix

### A.1 Analytic solution for the bispectrum in redshift space

The averaging over the angles  $\gamma_1$  and  $\gamma_2$  in Equation (2.98) in order to obtain the spherically averaged bispectrum in redshift space was performed analytically by Mathematica 8.0. The solution is split in three different parts, depending on the mode coupling, and is given by

$$B_{zs}(k_1, k_2, \theta_{12}) = H(k_1, k_2, \theta_{12}) + H(k_1, k_2, \theta_{13}) + H(k_2, k_3, \theta_{23}) \quad (\text{A.1})$$

where the  $\theta_{ij}$  are the angles between the corresponding  $k$ -modes. The  $H(k_i, k_j, \theta_{ij})$  functions describe the effects of the redshift space distortions of the mode coupling for two  $k$ -modes. Their functional form is very messy and can be expressed as follows

$$\begin{aligned} H(k_1, k_2, \theta_{12}) = & \frac{b_1^3}{105k_1k_2} (15 + 10\beta(z) + 2\beta^2(z) + \beta^2(z)\cos(2\theta_{12})) \\ & (7\cos(\theta_{12})k_1^2 + 2(6 + \cos(2\theta_{12}))k_1k_2 + 7\cos(\theta_{12})k_2^2) \\ & + \frac{b_1^3b_2}{15} (15 + 10\beta(z) + 2\beta^2(z) + \beta^2(z)\cos(2\theta_{12})) \\ & + \frac{\beta(z)b_1^3}{735k_3^2} \left( 6 + 8\cos^2(\theta_{12}) + 7\cos(\theta_{12}) \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right) \right) \\ & ((70 + 3\beta(z)(28 + 9\beta(z)))\cos(\theta_{12})k_1k_2 + 3\beta^2(z)\cos(3\theta_{12})k_1k_2 \\ & (35 + \beta(z)(35 + 9\beta(z)))(k_1^2 + k_2^2) + \beta(z)(7 + 6\beta(z))\cos(2\theta_{12}) \\ & (k_1^2 + k_2^2)) + \frac{\beta(z)b_1^4}{315k_1k_2} ((3(35 + 63\beta(z) + 42\beta^2(z) + 10\beta^3(z))\cos(\theta_{12}) \\ & \beta^2(z)(9 + 5\beta(z))\cos(3\theta_{12}))k_1^2 + 2(105 + 147\beta(z) + 81\beta^2(z) \\ & + 18\beta^3(z) + 2\beta(z)(21 + 27\beta(z) + 8\beta^2(z))\cos(2\theta_{12}) + \beta^3(z)\cos(4\theta_{12})) \\ & k_1k_2 + (3(35 + 63\beta(z) + 42\beta^2(z) + 10\beta^3(z))\cos(\theta_{12}) + \beta^2(z) \\ & (9 + 5\beta(z))\cos(3\theta_{12}))k_2^2) \end{aligned} \quad (\text{A.2})$$

$$\begin{aligned}
H(k_1, k_3, \theta_{13}) = & \frac{b_1^3}{105k_1k_3^3} (\beta(z) (5 + 3\beta(z)) k_1^2 + 2\beta(z) (5 + 3\beta(z)) \cos(\theta_{12})k_1k_2 \\
& + \beta(z) (5 + 2\beta(z) + \beta(z)\cos(2\theta_{12})) k_2^2 + 5 (3 + \beta(z)) k_3^2) (7\cos(\theta_{13})k_1^2 \\
& + 2 (6 + \cos(2\theta_{13})) k_1k_3 + 7\cos(\theta_{13})k_3^2) + \frac{b_1^3b_2}{15k_3^2} (\beta(z) (5 + 3\beta(z)) k_1^2 \\
& + 2\beta(z) (5 + 3\beta(z)) \cos(2\theta_{12})k_1k_2 + \beta(z) (5 + 2\beta(z) + \beta(z)\cos(2\theta_{12})) \\
& k_2^2 + 5 (3 + \beta(z)) k_3^2) + \frac{\beta(z)b_1^3}{735k_1k_3^3} (\beta(z) (14 + 9\beta(z) + (7 + 6\beta(z)) \\
& \cos(2\theta_{12})) k_1^2 + 6\beta(z)\cos(2\theta_{12}) (7 + 4\beta(z) + \beta(z)\cos(2\theta_{12})) k_1k_2 \\
& + 3\beta(z) (7 + 3\beta(z) + 2\beta(z)\cos(2\theta_{12})) k_2^2 + 7 (5 + 2\beta(z) + \beta(z)(2\theta_{12})) \\
& k_3^2) (7\cos(\theta_{13})k_1^2 + 2 (5 + 2\cos(2\theta_{13})) k_1k_3 + 7\cos(\theta_{13})k_3^2) - \frac{\beta(z)b_1^4k_2}{315k_1k_3^4} \\
& (-9\beta(z) (7 + 5\beta(z)) \cos(\theta_{12})k_1^4\beta(z) (-126 - 54\beta(z) + 20\beta^2(z) \\
& + 3 (-21 - 12\beta(z) + 5\beta^2(z)) \cos(2\theta_{12})) k_1^3k_2 + 3\cos(\theta_{12})k_1^2 \\
& (\beta(z) (-63 + 25\beta^2(z) + 10\beta^2(z)\cos(2\theta_{12})) k_2^2 + 5 (-7 + 3\beta^2(z)) k_3^2) \\
& + 3k_1k_2 (\beta(z) (-21 + 18\beta(z) + 18\beta^2(z) + 4\beta(z) (3 + 4\beta(z)) \\
& + \beta^2(z)\cos(4\theta_{12})) k_2^2 + (-35 + 28\beta(z) + 27\beta^2(z) + 2\beta(z) (7 + 9\beta(z)) \\
& \cos(2\theta_{12})) k_3^2) + \cos(\theta_{12}) (5\beta^2(z) (9 + 5\beta(z) + 2\beta(z)\cos(2\theta_{12})) k_2^4 \\
& + 18\beta(z) (7 + 4\beta(z) + \beta(z)\cos(2\theta_{12})) k_2^2k_3^2 + 21 (5 + 3\beta(z)) k_3^2) ,
\end{aligned} \tag{A.3}$$

and

$$\begin{aligned}
H(k_2, k_3, \theta_{23}) = & \frac{b_1^3}{105k_2k_3^3} \left( \beta(z) (5 + 2\beta(z) + \beta(z)\cos(2\theta_{12})) k_1^2 + 2\beta(z) (5 + 3\beta(z)) \right. \\
& \cos(\theta_{12})k_1k_2 + \beta(z) (5 + 3\beta(z)) k_2^2 + 5 (3 + \beta(z)) k_3^2 \left. (7\cos(\theta_{23})k_2^2 \right. \\
& + 2 (6 + \cos(2\theta_{23})) k_2k_3 + 7\cos(\theta_{23})k_3^2) + \frac{b_1^3b_2}{15k_3^2} \\
& (\beta(z) (5 + 2\beta(z) + \beta(z)\cos(2\theta_{12})) k_1^2 + 2\beta(z) (5 + 3\beta(z)) \cos(\theta_{12}) \\
& k_1k_2 + \beta(z) (5 + 3\beta(z)) k_2^2 + 5 (3 + \beta(z)) k_3^2) + \frac{\beta(z)b_1^3}{735k_3^2} (3\beta(z) \\
& (7 + 3\beta(z) + 2\beta(z)\cos(2\theta_{12})) k_1^2 + 6\beta(z)\cos(\theta_{12}) (7 + 4\beta(z)) \\
& + \beta(z)\cos(2\theta_{12})) k_1k_2 + \beta(z) (14 + 9\beta(z) + (7 + 6\beta(z)) \cos(2\theta_{12})) \\
& k_2^2 + 7 (5 + 2\beta(z) + \beta(z)\cos(2\theta_{12})) k_3^2) (6 + 8\cos^2(2\theta_{23}) \\
& + 7\cos(\theta_{23}) \left( \frac{k_2}{k_3} + \frac{k_3}{k_2} \right)) - \frac{\beta(z)b_1^4k_1}{315k_2k_3^4} (5\beta^2(z)\cos(\theta_{12}) (9 + 5\beta(z)) \\
& + 2\beta(z)\cos(2\theta_{12})) k_1^4 + 3\beta(z) (3 (-7 + 6\beta(z) + 6\beta^2(z)) + 4\beta(z) \\
& (3 + 4\beta(z)) \cos(2\theta_{12})\beta^2(z)\cos(4\theta_{12})) k_1^3k_2 - 3\cos(\theta_{12}) (k_2^2 - k_3^2) \\
& (3\beta(z) (7 + 5\beta(z)) k_2^2 + 7 (5 + 3\beta(z)) k_3^2) + 3\beta(z)\cos(\theta_{12})k_1^2 \\
& ((-63 + 25\beta^2(z) + 10\beta^2(z)\cos(2\theta_{12})) k_2^2 + 6 (7 + 4\beta(z))\beta(z)\cos(2\theta_{12})) \\
& k_3^2) + k_1k_2 (\beta(z) (-126 - 54\beta(z) + 20\beta^2(z) + 3 (-21 - 12\beta(z) \\
& + 5\beta^2(z)) \cos(2\theta_{12})) k_2^2 + 3 (-35 + 28\beta(z) + 27\beta^2(z) + 2\beta(z) \\
& (7 + 9\beta(z)) \cos(2\theta_{12})) k_3^2) .
\end{aligned} \tag{A.4}$$

## A.2 Derivation of the non-linear redshift space distortions for $\xi(r_p, \pi)$

The derivation of the non-linear Kaiser for the anisotropic two-point correlation function  $\xi(r_p, \pi)$  will be shown here, because until now this cannot be found in any other publication. An important ingredient for the later calculation are the relations between the two-point statistics in configuration and Fourier space, which is given by

$$P_l(k) = i^l 4\pi \int_0^\infty r^2 dr j_l(kr) \xi_l(r) \tag{A.5}$$

$$\xi_l(r) = \frac{4\pi}{i^l} \int_0^\infty \frac{k^2 dk}{(2\pi)^3} P_l(k) j_l(kr) \tag{A.6}$$

where  $i$  is the imaginary unit,  $j_l(kr)$  are the spherical Bessel functions, which will be given in Section A.4, and by the index  $l$  the multipole is indicated. The multipoles are required

to write the two-dimensional clustering quantities as

$$P_l(\vec{k}) = \sum_l \mathcal{P}_l(\mu) P_l(k) \quad (\text{A.7})$$

$$\xi_l(\vec{r}) = \sum_l \mathcal{P}_l(\mu) \xi_l(r) . \quad (\text{A.8})$$

So far, in the discussion no discrimination between real and redshift space was made. In real space, only the  $l = 0$  term, also called monopole, does not vanish after the integration of Equation (A.6). The computation in Fourier space can be easier carried out for the multipole terms. The following integration must be performed

$$P_l(k) = \frac{(2l+1)}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi \sin(\theta) d\theta \mathcal{P}_l(\mu) P(k, \mu) \quad (\text{A.9})$$

where  $\mathcal{P}_l$  are the Legendre polynomials, see Section A.3. In Scoccimarro (2004), the non-linear Kaiser effect on large scales for the power spectrum was derived as

$$P(k, \mu) = b^2 \left( P_{\delta\delta}(k) + 2\beta\mu^2 P_{\delta\theta}(k) + \beta^2\mu^4 P_{\theta\theta}(k) \right) . \quad (\text{A.10})$$

After some simple mathematical exercises, the monopole ( $l = 0$ ) can then be written as

$$\begin{aligned} P_0(k) &= \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi \sin(\theta) d\theta \mathcal{P}_0(\mu) P(k, \mu) = \\ &= b^2 \left( P_{\delta\delta}(k) + \frac{2}{3}\beta P_{\delta\theta}(k) + \frac{1}{5}\beta P_{\theta\theta}(k) \right) , \end{aligned} \quad (\text{A.11})$$

the quadrupole ( $l = 2$ ) as

$$P_2(k) = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi \sin(\theta) d\theta \mathcal{P}_2(\mu) P(k, \mu) = b^2 \left( \frac{4}{3}\beta P_{\delta\theta}(k) + \frac{4}{7}\beta^2 P_{\theta\theta}(k) \right) \quad (\text{A.12})$$

and the hexadecapole ( $l = 4$ ) as

$$P_4(k) = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi \sin(\theta) d\theta \mathcal{P}_4(\mu) P(k, \mu) = b^2 \left( \frac{8}{35}\beta^2 P_{\theta\theta}(k) \right) . \quad (\text{A.13})$$

All multipoles with  $l > 4$  and the odd ones vanish. The results from above can be inserted in Equation (A.6) to get the configuration space quantities and the multipoles are given by

$$\xi_{0,zs}(r) = \frac{b^2}{2\pi} \int_0^\infty j_0(kr) P_0(k) k^2 dk = b^2 \left( \xi_{\delta\delta}(r) + \frac{2}{3}\beta \xi_{\delta\theta}(r) + \frac{1}{5}\beta^2 \xi_{\theta\theta}(r) \right) \quad (\text{A.14})$$

$$\begin{aligned} \xi_{2,zs}(r) &= \frac{b^2}{2\pi} \int_0^\infty j_2(kr) P_2(k) k^2 dk = \\ &= b^2 \left( \frac{4}{3}\beta \left( \xi_{\delta\theta}(r) - \bar{\xi}_{\delta\theta}(r) \right) + \frac{4}{7}\beta \left( \xi_{\theta\theta}(r) - \bar{\xi}_{\theta\theta}(r) \right) \right) \end{aligned} \quad (\text{A.15})$$

$$\xi_{4,zs}(r) = \frac{b^2}{2\pi} \int_0^\infty j_4(kr) P_4(k) k^2 dk = b^2 \left( \frac{8}{35}\beta^2 \left( \xi_{\theta\theta}(r) + \frac{5}{2}\bar{\xi}_{\theta\theta}(r) - \frac{7}{2}\bar{\bar{\xi}}_{\theta\theta}(r) \right) \right) \quad (\text{A.16})$$

where

$$\bar{\xi}_{\delta\theta}(r) \equiv \frac{3}{r^3} \int_0^s ds s^2 \xi_{\delta\theta}(s) \quad (\text{A.17})$$

$$\bar{\bar{\xi}}_{\delta\theta}(r) \equiv \frac{5}{r^5} \int_0^s ds s^4 \xi_{\delta\theta}(s) \quad (\text{A.18})$$

$$\bar{\bar{\xi}}_{\theta\theta}(r) \equiv \frac{5}{r^5} \int_0^s ds s^4 \xi_{\theta\theta}(s) . \quad (\text{A.19})$$

According to Equation (A.8)  $\xi_{zs}(r_p, \pi)$  for the non-linear Kaiser effect can be written as

$$\xi_{zs}(r_p, \pi) = (P_0 \xi_{0,zs})(r_p, \pi) + (P_2 \xi_{2,zs})(r_p, \pi) + (P_4 \xi_{4,zs})(r_p, \pi). \quad (\text{A.20})$$

### A.3 Legendre polynomials

The ordinary differential equation for  $x \in [-1, 1]$

$$\frac{d}{dx} \left[ (1-x^2) \frac{dP_l(x)}{dx} \right] + l(l+1) P_l(x) = 0 \quad (\text{A.21})$$

where  $l \in \mathbb{N}_0$  and can be solved by the Legendre polynomials. The general solution can be written as

$$P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l \quad (\text{A.22})$$

and the first five Legendre polynomials are given by

$$P_0(x) = 1 \quad (\text{A.23})$$

$$P_1(x) = x \quad (\text{A.24})$$

$$P_2(x) = \frac{(3x^2 - 1)}{2} \quad (\text{A.25})$$

$$P_3(x) = \frac{(5x^3 - 3x)}{2} \quad (\text{A.26})$$

$$P_4(x) = \frac{(35x^4 - 30x^2 + 3)}{8} \quad (\text{A.27})$$

but only the even ones (up to  $l = 4$ ) are of interest for the derivation of the non-linear Kaiser effect for  $\xi(r_p, \pi)$  in Section A.2.

### A.4 Spherical Bessel functions

The ordinary differential equation, named after Friedrich Wilhelm Bessel, which is given by

$$\left( x^2 \frac{d^2}{dx^2} + 2x \frac{d}{dx} + (x^2 - l(l+1)) \right) j_l(x) = 0 \quad (\text{A.28})$$

can be solved by the spherical Bessel functions

$$j_l(x) = x^l \left( -\frac{1}{x} \frac{d}{dx} \right)^l \frac{\sin(x)}{x}. \quad (\text{A.29})$$

The first five functions can be computed as

$$j_0(x) = \frac{\sin(x)}{x} \quad (\text{A.30})$$

$$j_1(x) = \frac{\sin(x)}{x^2} - \frac{\cos(x)}{x} \quad (\text{A.31})$$

$$j_2(x) = \frac{3\sin(x)}{x^3} - \frac{\sin(x)}{x} - \frac{3\cos(x)}{x^2} \quad (\text{A.32})$$

$$j_3(x) = \frac{\cos(x)}{x} \left( 1 - \frac{15}{x^2} \right) - \frac{\sin(x)}{x^2} \left( 10 - \frac{105}{x^2} \right) \quad (\text{A.33})$$

$$j_4(x) = \frac{\sin(x)}{x} \left( 1 - \frac{(45 - \frac{105}{x^2})}{x^2} \right) + \frac{\cos(x)}{x^2} \left( 10 - \frac{105}{x^2} \right) \quad (\text{A.34})$$

where only the odd  $l$ -values up to  $l = 4$  are required for the derivation of the non-linear Kaiser effect for  $\xi(r_p, \pi)$ .

## A.5 Bias parameter estimation with collapsed triangles

In this short section, the results will be investigated if the collapsed triangle configurations are taken into account for the bias parameters estimation of  $b_1$  and  $b_2$ . In the Figures A.1, A.2 and A.3, the results of  $b_1$  and  $b_2$  derived from the dark matter bispectrum are shown for  $z = 0.0, 0.5$  and  $1.0$ , respectively. As in Section 3, the extracted linear bias  $b_1$  and quadratic bias  $b_2$  are plotted on the left and right panel, respectively.

It can be noticed that for all redshifts  $b_1$  and  $b_2$  cannot be extracted consistently with the dark matter expectation value which is indicated by the black dotted vertical line in all figures. The linear bias  $b_1$  is always underestimated with a significance of at least 1.5  $\sigma$ -confidence level. The estimated  $b_2$ -values are always overestimated but with at least the same significance which was found for  $b_1$ . For these results it does not matter whether the analysis is performed in real or redshift space and it was decided to exclude the case with included random motions from this discussion because only a mild improvement can be achieved, see Section 3.2.3. It should be clear by now, that the collapsed triangle configurations must be excluded for a correct bias parameter estimation, as it was stated in the beginning of Section 3.2.

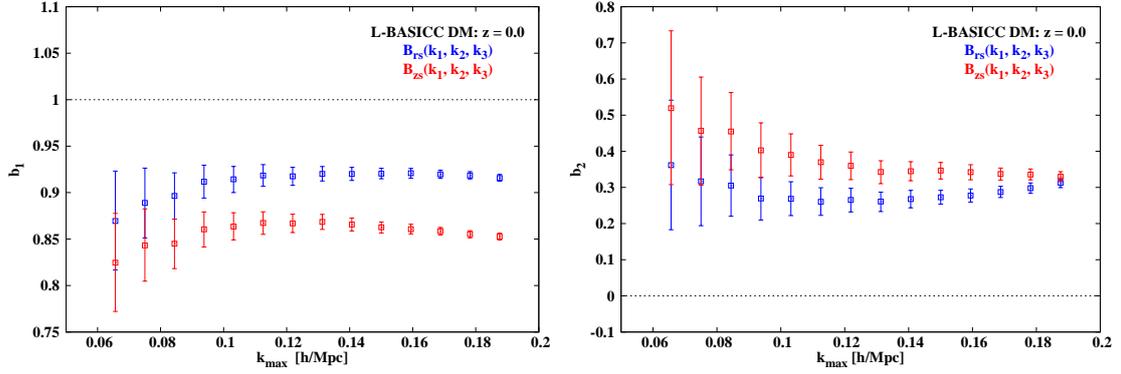


Figure A.1: Estimation of  $b_1$  (left panel) and  $b_2$  (right panel) against  $k_{\max}$ , the maximum  $k$ -mode for the side length of a triangle configuration. The real space estimation is given in blue and redshift space in red.

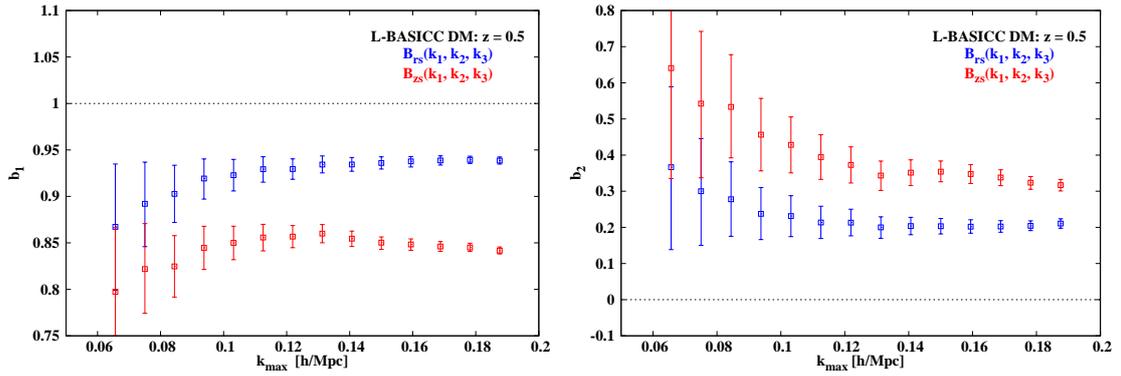


Figure A.2: Same description as for Figure A.1

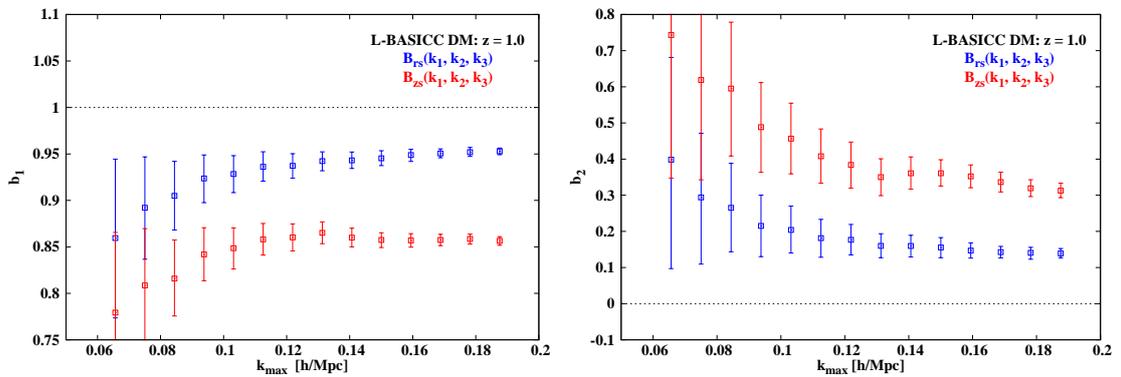


Figure A.3: Same description as for Figure A.1



# Bibliography

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, 182, 543
- Adelman-McCarthy, J. K., et al. 2008, *ApJS*, 175, 297
- Anderson, L., Aubourg, E., Bailey, S., et al. 2012, [arXiv:1203.6594](https://arxiv.org/abs/1203.6594)
- Angulo, R. E., Baugh, C. M., Frenk, C. S., & Lacey, C. G. 2008, *MNRAS*, 383, 755
- Angulo, R. E., Baugh, C. M., & Lacey, C. G. 2008, *MNRAS*, 387, 921
- Albrecht, A. et al., 2006, preprint ([arXiv:astro-ph/0609591](https://arxiv.org/abs/astro-ph/0609591))
- Alcock, C., & Paczynski, B. 1979, *nat*, 281, 358
- Barriga, J., & Gaztañaga, E. 2002, *MNRAS*, 333, 443
- Baugh, C. et al., in preperation
- Baum, W. A. 1962, *IAU Symposium*, 15, 390
- Bel, J., & Marinoni, C. 2012, [arXiv:1205.3200](https://arxiv.org/abs/1205.3200)
- Benítez, N. et al., 2009, *ApJ*, 691, 241
- Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, *PhysRep*, 367, 1
- Bernardeau, F., Crocce, M., & Scoccimarro, R. 2008, *PRD*, 78, 103521
- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics: Second Edition*, by James Binney and Scott Tremaine. ISBN 978-0-691-13026-2 (HB). Published by Princeton University Press, Princeton, NJ USA, 2008.,
- Blake, C., & Glazebrook, K. 2003, *ApJ*, 594, 665
- Blake, C., & Bridle, S. 2005, *MNRAS*, 363, 1329
- Bode, P., & Bertschinger, E. 1995, *ArXiv Astrophysics e-prints*, [arXiv:astro-ph/9504040](https://arxiv.org/abs/astro-ph/9504040)
- Bosma, A. 1978, Ph.D. Thesis

- Bouchet, F. R., Juszkiewicz, R., Colombi, S., & Pellat, R. 1992, *ApJL*, 394, L5
- Bouchet, F. R., Colombi, S., Hivon, E., & Juszkiewicz, R. 1995, *AAP*, 296, 575
- Cabré, A., & Gaztañaga, E. 2009, *MNRAS*, 396, 1119
- Cai, Y.-C., Angulo, R. E., Baugh, C. M., Cole, S., Frenk, C. S., & Jenkins, A. 2009, *MNRAS*, 395, 1185
- Caldwell, R. R., Kamionkowski, M., & Weinberg, N. N. 2003, *Physical Review Letters*, 91, 071301
- Carlberg, R. G., Yee, H. K. C., Morris, S. L., Lin, H., Hall, P. B., Patton, D., Sawicki, M., & Shepherd, C. W. 2000, *ApJ*, 542, 57
- Chambers, K. C., & Pan-STARRS Team 2004, *Bulletin of the American Astronomical Society*, 36, 828
- Chuang, C.-H., Wang, Y., & Hemantha, M. D. P. 2010, arXiv:1008.4822
- Ciardullo, R., Gronwall, C., Wolf, C., et al. 2012, *ApJ*, 744, 110
- Cole, S., Fisher, K. B., & Weinberg, D. H., 1994, *MNRAS*, 267, 785
- Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, *MNRAS*, 319, 168
- Cole, S., et al. 2005, *MNRAS*, 362, 505
- Copeland, E. J., Sami, M., & Tsujikawa, S. 2006, *International Journal of Modern Physics D*, 15, 1753
- Copeland, E. J. 2007, *Particles, Strings, and Cosmology-PASCOS 2007*, 957, 21
- Colless, M., Dalton, G., Maddox, S., et al. 2001, *MNRAS*, 328, 1039
- Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, *ApJL*, 648, L109
- Cooray, A., & Sheth, R. 2002, *PhysRep*, 372, 1
- Crocce, M., & Scoccimarro, R. 2006, *PRD*, 73, 063519
- Cunha, C. E. et al. 2009, *MNRAS*, 396, 2379
- Cui, W., Liu, L., Yang, X., et al. 2008, *ApJ*, 687, 738
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2012, arXiv:1208.0022
- Davis, M. & Peebles, P. J. E. 1983, *ApJ*, 267, 465

- Dodelson, S. 2003, *Modern cosmology* / Scott Dodelson. Amsterdam (Netherlands): Academic Press. ISBN 0-12-219141-2, 2003, XIII + 440 p.,
- Einstein, A. 1915, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften* (Berlin), Seite 844-847., 844
- Eisenstein, D. J., et al. 2001, *AJ*, 122, 2267
- Eisenstein, D. J., et al. 2005, *ApJ*, 633, 560
- Eisenstein, D. J., Seo, H.-J., & White, M. 2007, *ApJ*, 664, 660
- Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, *AJ*, 142, 72
- Feldman, H. A., Kaiser, N., & Peacock, J. A. 1994, *ApJ*, 426, 23
- Frige, M. & Johnson, S. G., *The Design and Implementation of FFTW3* Proceedings of the IEEE 93 (2), 216-231 (2005)
- Fry, J. N. 1984, *ApJ*, 279, 499
- Fry, J. N., & Gaztañaga, E. 1993, *ApJ*, 413, 447
- Fry, J. N. 1996, *ApJL*, 461, L65
- Gaztañaga, E., Cabré, A., & Hui, L. 2008, *ArXiv e-prints*, 807, arXiv:0807.3551
- Gaztañaga, E., Cabré, A., Castander, F., Crocce, M., & Fosalba, P. 2009, *MNRAS*, 399, 801
- Goroff, M. H., Grinstein, B., Rey, S.-J., & Wise, M. B. 1986, *ApJ*, 311, 6
- Guo, H., & Jing, Y. P. 2009, *ApJ*, 698, 479
- Hamilton, A. J. S. 1992, *ApJL*, 385, L5
- Hamilton, A. J. S. 1993, *ApJ*, 417, 19
- Heavens, A. F., Matarrese, S., & Verde, L. 1998, *MNRAS*, 301, 797
- Hill, G. J., Gebhardt, K., Komatsu, E., & MacQueen, P. J. 2004, *The New Cosmology: Conference on Strings and Cosmology*, 743, 224
- Hill, G. J., Gebhardt, K., Komatsu, E., et al. 2008, *Panoramic Views of Galaxy Formation and Evolution*, 399, 115
- Hinshaw, G., Spergel, D. N., Verde, L., et al. 2003, *ApJS*, 148, 135
- Hockney, R. W., & Eastwood, J. W. 1981, *Computer Simulation Using Particles*, New York: McGraw-Hill, 1981,

- Hui, L., & Parfrey, K. P. 2007, ArXiv e-prints, 712, arXiv:0712.1162  
<http://camb.info/sources/>  
<http://cfa-www.harvard.edu/~mzaldarr/CMBFAST/cmbfast.html>
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
- Friedmann, A. 1922, Zeitschrift für Physik, 10, 377
- Hu, W., Sugiyama, N., & Silk, J. 1995, ArXiv Astrophysics e-prints, arXiv:astro-ph/9504057
- Hu, W., & White, M. 1996, ApJ, 471, 30
- Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168
- Jackson, J. C. 1972, MNRAS, 156, 1P
- Jain, B., & Bertschinger, E. 1994, ApJ, 431, 495
- Jennings, E., Baugh, C. M., Pascoli, S., 2011, ApJL, 727, 9
- Jeong, D., & Komatsu, E. 2006, ApJ, 651, 619
- Jeong, D., & Komatsu, E. 2009, ApJ, 691, 569
- Jing, Y. P., Mo, H. J., & Boerner, G. 1998, ApJ, 494, 1
- Jing, Y. P. 2005, ApJ, 620, 559
- Kaiser, N. 1984, ApJL, 284, L9
- Kaiser, N. 1987, MNRAS, 227, 1
- Kayo, I., Suto, Y., Nichol, R. C., et al. 2004, PASJ, 56, 415
- Kazin, E. A., et al. 2010, ApJ, 710, 1444
- Kazin, E. A., Blanton, M. R., Scoccimarro, R., McBride, C. K., & Berlind, A. A. 2010, ApJ, 719, 1032
- Kazin, E. A., Sánchez, A. G., Blanton, M. R., 2011, arXiv:1105.2037
- Kerscher, M., Szapudi, I., & Szalay, A. S. 2000, ApJL, 535, L13  
<http://gyudon.as.utexas.edu/komatsu/CRL/index.html>
- Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330
- Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18

- Landy, S. D., & Szalay, A. S. 1993, *ApJ*, 412, 64
- Lange, A., de Bernardis, P., de Petris, M., et al. 1995, *SSR*, 74, 145
- Lewis, A., & Bridle, S. 2002, *PRD*, 66, 103511
- Lewis, A., Challinor, A., & Lasenby, A. 2000, *ApJ*, 538, 473
- Linder, E. V. 2007, ArXiv e-prints, 709, arXiv:0709.1113
- Lourakis, M. I. A. 2005, <http://ics.forth.gr/lourkis/levmar/levmar.pdf>
- Ma, C.-P., & Bertschinger, E. 1995, *ApJ*, 455, 7
- Matsubara, T. 2008, *PRD*, 77, 063530
- McDonald, P. 2006, *PRD*, 74, 103512
- Matsubara, T. 2004, *ApJ*, 615, 573
- Matsubara, T., & Suto, Y. 1996, *ApJL*, 470, L1
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E., Equation of State Calculations by Fast Computing Machines, *The Journal of Chemical Physics*, Vol.21,No.6.(June,1953).pp.1087-1092
- Mo, H. J., Jing, Y. P., & White, S. D. M. 1997, *MNRAS*, 284, 189
- Mo, H. J., & White, S. D. M. 1996, *MNRAS*, 282, 347
- Montesano, F., Sanchez, A. G., & Phleps, S. 2011, arXiv:1107.4097
- Mukhanov, V. 2005, *Physical foundations of cosmology*, by V. Mukhanov. Cambridge, UK: Cambridge University Press, 2005
- Myers, A. D., White, M. & Ball, N. M., 2009, *MNRAS*, 399, 2279
- Nishimichi, T., Ohmuro, H., Nakamichi, M., et al. 2007, *PASJ*, 59, 1049
- Nilsson, K. K., & Meisenheimer, K. 2009, *NAR*, 53, 37
- Nolta, M. R., Dunkley, J., Hill, R. S., et al. 2009, *ApJS*, 180, 296
- Norberg, P., Baugh, C. M., Gaztañaga, E. & Croton, D. J., 2009, *MNRAS*, 396, 19
- Okumura, T., Matsubara, T., Eisenstein, D. J., Kayo, I., Hikage, C., Szalay, A. S., & Schneider, D. P. 2008, *ApJ*, 676, 889
- Padmanabhan, N., White, M., Norberg, P., & Porciani, C. 2008, ArXiv e-prints, 802, arXiv:0802.2105

- Padmanabhan, N., & White, M. 2008, PhD, 77, 123540
- Pan, J., & Szapudi, I. 2005, MNRAS, 362, 1363
- Pan, J., Coles, P., & Szapudi, I. 2007, MNRAS, 382, 1460
- Peacock, J. A., Schneider, P., Efstathiou, G., Ellis, J. R., Leibundgut, B., Lilly, S. J., & Mellier, Y. 2006, ESA-ESO Working Group on "Fundamental Cosmology", Edited by J.A. Peacock et al. ESA, 2006.,
- Peebles, P. J. E. 1980, Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p.,
- Perlmutter, S., et al. 1999, ApJ, 517, 565
- Phleps, S., Wolf, C., Peacock, J. A., Meisenheimer, K., & van Kampen, E. 2007, AAP, 468, 113
- Pollack, J. E., Smith, R. E., & Porciani, C. 2012, MNRAS, 420, 3469
- Press et al. 1996, Numerical Recipes for Fortran 90
- Räth, C., Bunk, W., Huber, M. B., et al. 2002, MNRAS, 337, 413
- Räth, C., Banday, A. J., Rossmannith, G., et al. 2011, MNRAS, 415, 2205
- Riess, A. G., et al. 1998, AJ, 116, 1009
- Ross, N. P., et al. 2007, MNRAS, 381, 573
- Ross, N. P., Shanks, T., & Cruz da Ângela, J. 2007, Cosmic Frontiers, 379, 68
- Rossmannith, G., Räth, C., Banday, A. J., & Morfill, G. 2009, MNRAS, 399, 1921
- Rubin, V. C., Ford, W. K. J., & Thonnard, N. 1980, ApJ, 238, 471
- Saglia, R. P. et al. 2012, ApJ, 746, 128
- Sánchez, A. G., Baugh, C. M., Percival, W. J., Peacock, J. A., Padilla, N. D., Cole, S., Frenk, C. S., & Norberg, P. 2006, MNRAS, 366, 189
- Sánchez, A. G., Baugh, C. M., & Angulo, R. 2008, MNRAS, 390, 1470
- Sánchez, A. G., Scoccola, C. G., Ross, A. J., et al. 2012, arXiv:1203.6616
- Schlagenhafer, H. A., Phleps, S., & Sanchez, A. G. 2012, arXiv:1206.2640
- Scoccimarro, R., Colombi, S., Fry, J. N., et al. 1998, ApJ, 496, 586
- Scoccimarro, R. 2000, ApJ, 544, 597

- Scoccimarro, R., & Sheth, R. K. 2002, MNRAS, 329, 629
- Scoccimarro, R. 2004, PRD, 70, 083007
- Sefusatti E., *Thesis*, Probing fundamental Physics with Large-Scale Structure, 05/2005
- Sefusatti, E., & Scoccimarro, R. 2005, PRD, 71, 063001
- Sefusatti, E. Crocce, M., Pueblas, S., & Scoccimarro, R. 2006, PRD, 74, 023522
- Sefusatti, E., & Komatsu, E. 2007, PRD, 76, 083004
- Sefusatti, E. 2009, PRD, 80, 123002
- Sefusatti, E., Crocce, M., & Desjacques, V. 2011, arXiv:1111.6966
- Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119
- Smith, R. E., et al. 2003, MNRAS, 341, 1311
- Smith, R. E., Sheth, R. K., & Scoccimarro, R. 2008, PRD, 78, 023523
- Simpson, F., James, J. B., Heavens, A. F., & Heymans, C. 2011, Physical Review Letters, 107, 261301
- Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, ApJS, 148, 175
- Spergel, D. N., et al. 2007, ApJS, 170, 377
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, NAT, 435, 629
- Swanson, M. E. C., Tegmark, M., Blanton, M., & Zehavi, I. 2008, MNRAS, 385, 1635
- Szalay, A. S., Matsubara, T., & Landy, S. D. 1998, ApJL, 498, L1
- Szapudi, S., & Szalay, A. S. 1998, ApJL, 494, L41
- Tucker, D. L., et al. 2010, Bulletin of the American Astronomical Society, 42, #470.09
- Verde, L., Heavens, A. F., Percival, W. J., & Matarrese, S. 2002, arXiv:astro-ph/0212311
- Verde, L., Peiris, H. V., Spergel, D. N., et al. 2003, ApJS, 148, 195
- Wake, D. A., et al. 2006, MNRAS, 372, 537
- Wake, D. A., et al. 2008, ArXiv e-prints, 802, arXiv:0802.4288
- White, S. D. M., Navarro, J. F., Evrard, A. E., & Frenk, C. S. 1993, NAT, 366, 429
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579
- Zehavi, I., et al. 2005, ApJ, 621, 22



# Acknowledgements

This last section is dedicated to the people whom I want to express my gratitude for their not only scientific help in various ways.

First of all, Ralf Bender who gave me the opportunity to work on the field of large scale structure, especially on the three point clustering. I want to thank Stefanie Phleps, who was responsible for my daily supervision, helped me a lot by the accomplishment of the scientific issues which arose throughout the four years of my thesis, Ariel Sánchez for his comments on  $\xi(r_p, \pi)$  and for being a member of my thesis committee as well as Donghui Jeong and Eiichiro Komatsu for their precious help for measuring the bispectrum.

I also want to thank all other members of the OPINAS group on the MPE in Garching for the friendly welcome they were offering to me and especially my officemates Jennifer Connelly, Max Fabricius, Francesco Montesano and Michael Opitsch.

It would not be possible for me to finish this work without the support with words and deeds of my parents, Barbara and Alois Schlagenhauser, and my sister, Antje. I think at some moments during this four years they felt even more desperate about the outcome of this thesis than I did.

And of course I want to mention my friends who gave me distraction when I needed it even when I was opposite opinion. I want to thank Daniel Gothe for the time he spent on watching all the social critical movies, playing console games and who showed me the beauty of icehockey. I will never forget our trip to Ingolstadt with his wife Alex and Sandra Pohl who was a total stranger to us at that time. The result of this trip was the foundation of the Fan-Club "Rauten Roosters".

On many, perhaps too many, evenings Daniel Semsch and I discussed new tactics for an online space adventure game. For sure the time could have been invested in a much more productive way, but the distraction I experienced was worth it.

I want to thank Thomas Pointl, an old friend, in both meanings, for his weekend plannings during my homestays.

And last but not least I want to mention my friends from the STA, Christian Prokopp, Heiko Müller and Frank Heinitz, for many memorable weekends (to be honest it could have been more weekends) in Erlangen and Erzhausen.



# Curriculum vitae

Holger Alois Schlagenhauer

Date of birth: 31.12.1980 in Nabburg

Education: 1987 - 1991: Grundschule Diendorf  
(Primary school Diendorf)

1991 - 1993: Teilhauptschule Diendorf  
(Secondary modern school Diendorf)

1993 - 1997: Realschule Nabburg  
(Junior high school Nabburg)

1997 - 1999: Fachoberschule Schwandorf  
(Technical school Schwandorf)

1999 - 2001: Siemens Technik Akademie in Erlangen  
(Siemens technical academy Erlangen)

2001 - 2002: Fachhochschule München  
(Munich university of applied sciences)

2002 - 2008: Ludwig Maximilians Universität München  
(Ludwig Maximilian University of Munich)

2008 - 2012: Promotion am Max-Planck-Institut für extraterrestrische Physik  
(PhD at the Max-Planck-Institute for extraterrestrial Physics)



# Erklärung

Ich erkläre hiermit, dass ich meine Doktorarbeit mit dem Titel

**Two- and three-point clustering statistics**

selbständig verfasst, sowie keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

München, den 27.09.2012

---

Holger Alois  
Schlagenhafer

