



# Durham E-Theses

---

## *The clustering of dark matter, haloes & galaxies*

Angulo de la Fuente, Raúl Esteban

### How to cite:

---

Angulo de la Fuente, Raúl Esteban (2008) *The clustering of dark matter, haloes & galaxies*, Durham theses, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/2516/>

### Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

# The clustering of dark matter, haloes & galaxies

Raúl Esteban Angulo de la Fuente

## Abstract

In this thesis I study the spatial distribution of galaxies, haloes and dark matter particles using a suite of state-of-the-art cosmological N-body simulations of the growth of structure in the Universe. The subjects investigated are conceptually divided into three areas.

One line of research, which is made up of Chapters 2 and 3, is to explore the power and limitations of measurements of the imprint of baryonic acoustic oscillations in the clustering of galaxies. I look at how the appearance of the power spectrum is altered by different effects such as nonlinear evolution or redshift space distortions. In these chapters I also explore the best way to analyse survey data and how well new datasets, from both spectroscopic and photometric surveys, will be able to constrain the dark energy equation of state.

In a second strand, I study dark matter haloes and their substructures. In Chapter 4, I look at the dependence of the clustering strength of dark haloes on the concentration of the sample. I was able to go beyond the traditional 2-point statistics to extend previous analyses to higher order statistics thanks to the development of a novel way to extract the higher order bias parameters. In Chapter 5, I then zoomed into smaller scales to study a number of properties of the population of substructures within dark haloes. In particular, I consider the mass distribution of substructures as well as their radial distribution and orientation. I also demonstrate that mergers between substructures do indeed occur, which result from objects that are dynamically or geometrically linked before accretion.

In the final line of research, presented in Chapters 6 and 7, I develop ideas about how to add more realism to current theoretical predictions for galaxy clustering, and how it would be possible to use low-resolution dark matter simulations to investigate uncertainties in future observations.

# The Clustering of Dark Matter, Haloes & Galaxies

by Raúl Esteban Angulo de la Fuente

A thesis submitted to the University of Durham  
in accordance with the regulations for  
admittance to the Degree of Doctor of Philosophy.

Department of Physics

University of Durham

September 2008

The copyright of this thesis rests with the author or the university to which it was submitted. No quotation from it, or information derived from it may be published without the prior written consent of the author or university, and any information derived from it should be acknowledged.

27 FEB 2009



Copyright (c) 2008 R. E. Angulo.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the licence is available at <http://www.gnu.org/copyleft/fdl.html>

The abstract of this thesis remains copyright (c) University of Durham.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The $\Lambda$ CDM model . . . . .	1
1.1.1	Dark matter . . . . .	2
1.1.2	Inflation . . . . .	3
1.1.3	Cosmological parameters . . . . .	3
1.1.3.1	Dark Energy . . . . .	4
1.1.4	Growth of structures . . . . .	5
1.2	Outline of this thesis . . . . .	7
<b>2</b>	<b>The detectability of baryonic acoustic oscillations in future galaxy surveys</b>	<b>9</b>
2.1	Introduction . . . . .	9
2.2	Method . . . . .	12
2.2.1	N-Body Simulations . . . . .	12
2.2.1.1	The high resolution simulation: the BASICC . . . . .	15
2.2.1.2	The ensemble of low resolution simulations: L-BASICC . . . . .	16
2.2.1.3	Power spectrum estimation and errors . . . . .	17
2.2.2	Modelling the formation and evolution of galaxies . . . . .	21
2.3	The power spectrum of galaxy clustering . . . . .	25
2.3.1	The nonlinear growth of matter fluctuations . . . . .	27
2.3.2	The impact of redshift-space distortions on the power spectrum . . . . .	30
2.3.3	The power spectrum of dark matter halos in real and redshift space . . . . .	32
2.3.4	The power spectrum of galaxies . . . . .	36
2.4	Constraining the Dark Energy Equation of state . . . . .	39
2.5	Results . . . . .	46

2.5.1	The algorithm to extract the scale of the acoustic oscillations in action . . . . .	46
2.5.2	Two tests of the algorithm . . . . .	49
2.5.3	The main results . . . . .	57
2.5.4	Forecasts for future surveys . . . . .	60
2.6	Conclusions . . . . .	62
<b>3</b>	<b>Implications of Photometric Redshift Errors for BAO Detection</b>	<b>67</b>
<b>4</b>	<b>The assembly bias of dark matter haloes to higher orders</b>	<b>73</b>
4.1	Introduction . . . . .	73
4.2	Theoretical background and method . . . . .	75
4.2.1	The counts in cells approach to measuring clustering . . . . .	75
4.2.1.1	Higher order correlations: unsmoothed and continuous density field . . . . .	75
4.2.1.2	Higher order correlations: smoothed and discrete density fields . . . . .	78
4.2.2	Hierarchical amplitudes . . . . .	80
4.2.3	Higher order correlations: biased tracers . . . . .	81
4.2.4	Numerical Simulations . . . . .	83
4.3	Results . . . . .	86
4.3.1	Hierarchical amplitudes for the dark matter . . . . .	86
4.3.2	The hierarchical amplitudes of dark matter haloes . . . . .	89
4.3.3	Cross-correlation estimates of higher order clustering . . . . .	91
4.3.4	The bias parameters of dark matter haloes . . . . .	94
4.4	Summary and Discussion . . . . .	97
<b>5</b>	<b>The Fate of Substructures in Cold Dark Matter Haloes</b>	<b>101</b>
5.1	Introduction . . . . .	101
5.2	Method . . . . .	102
5.2.1	N-body Simulations . . . . .	102
5.2.2	Halo and Subhalo catalogues . . . . .	103
5.3	SubHalo properties . . . . .	106
5.3.1	Subhalo mass function . . . . .	106
5.3.2	Most massive subhaloes . . . . .	109

5.3.3	Radial distribution of subhaloes . . . . .	111
5.3.4	Angular distribution of subhaloes . . . . .	113
5.4	Mergers between subhaloes . . . . .	116
5.4.1	Subhalo merger rate . . . . .	119
5.4.2	Characterization of subhalo-subhalo mergers . . . . .	121
5.4.2.1	Radial distribution of satellite-satellite mergers . . . . .	121
5.4.2.2	Orbits of merging satellites . . . . .	122
5.4.2.3	The mass ratio of subhalo mergers . . . . .	125
5.4.3	Merger probability since accretion . . . . .	128
5.5	Summary and conclusions . . . . .	128
<b>6</b>	<b>Extending the dynamic range of low resolution dark matter simulations</b>	<b>133</b>
6.1	Introduction . . . . .	133
6.2	Method . . . . .	135
6.2.1	Theoretical Motivation . . . . .	135
6.2.2	Implementation . . . . .	136
6.3	Results . . . . .	137
6.3.1	The sub-resolution halo catalogue . . . . .	137
6.3.2	Abundance and clustering . . . . .	138
6.4	Application: Large scale clustering of Luminous Red Galaxies . . . . .	141
6.4.1	The haloes and LRG catalogues . . . . .	143
6.4.1.1	Correlation Function of LRG . . . . .	145
6.5	Conclusions . . . . .	149
<b>7</b>	<b>The construction of a halo light-cone</b>	<b>151</b>
7.1	Introduction . . . . .	151
7.2	Haloes on the light-cone . . . . .	153
7.3	Building a light-cone . . . . .	154
7.3.1	Algorithm . . . . .	155
7.3.2	Light-cones from the Millennium Simulation . . . . .	156
7.4	Results . . . . .	156
7.5	Summary . . . . .	162
<b>8</b>	<b>Conclusions</b>	<b>165</b>
8.1	Galaxy surveys and BAO . . . . .	165

8.2	Dark matter haloes and their substructures . . . . .	166
8.3	Improving current mock catalogues . . . . .	168

# List of Figures

2.1	A test of the choice of starting redshift used in the N-body simulations. . .	13
2.2	The power spectrum of the dark matter in real space measured at $z = 63$	17
2.3	The fractional error in the power spectrum of the dark matter and in the power spectrum of haloes more massive than $1.8 \times 10^{13} h^{-1} M_{\odot}$ . . . . .	20
2.4	the fraction of ‘resolved galaxies’ in the high resolution N-body simula- tion as a function of magnitude and the cumulative luminosity function of galaxies brighter than a given $R$ -band magnitude. . . . .	22
2.5	The growth of the power spectrum of density fluctuations in the dark mat- ter, as measured in real space . . . . .	26
2.6	The nonlinear growth of the power spectrum. . . . .	29
2.7	The ratio of the power spectrum measured for the dark matter in redshift space. . . . .	31
2.8	The power spectrum of dark matter haloes measured in real space compared to a scaled version of the prediction of linear perturbation theory. . . . .	33
2.9	The power spectrum of dark matter haloes measured in redshift space divided by the power spectrum measured in real space for the same sample	34
2.10	The power spectrum of different galaxy samples measured in real space, divided by the square of an effective bias parameter and the appropriately scaled linear perturbation theory power spectrum. . . . .	37
2.11	The ratio of the power spectrum of galaxies measured in redshift space to that in real space, at $z = 0$ and $z = 1$ . . . . .	37
2.12	The relation between the dark energy equation of state parameter, $w$ , and the scale factor, $\alpha$ . . . . .	40
2.13	The power spectra of dark matter particles, dark matter haloes and galaxies at $z = 0$ . . . . .	47

2.14	The ratio of the measured power spectrum divided by a smooth reference spectrum. . . . .	48
2.15	The constraints on the parameters $k_{\text{nl}}$ and $\alpha$ for the power spectra plotted in Fig. 2.13 . . . . .	50
2.16	The best fit value for the scaling parameter $\alpha$ , recovered from the ensemble of low resolution simulations. . . . .	51
2.17	The best-fitting value of the scale factor $\alpha$ as a function of redshift, for different tracers of the density distribution, in real space and redshift space. . . . .	55
2.18	The best-fitting value of the damping scale $k_{\text{nl}}$ as a function of redshift, for different tracers of the density distribution, in real space and redshift space. . . . .	56
2.19	The recovered value of the stretch parameter $\alpha$ for the galaxy samples listed in Table 2.1 . . . . .	57
3.1	The mean and standard deviation of the dark matter power spectrum averaged over an ensemble of 50 N-body simulations at $z = 0.5$ . . . . .	69
3.2	The ratio of the error on the measurement of the BAO scale in photo- $z$ space to that in redshift space (i.e. from a perfect spectroscopic redshift) as a function of the magnitude of the photometric redshift error. . . . .	72
4.1	The ratio $V_{\text{max}}/V_{200}$ as a function of halo mass for gravitationally bound haloes in the BASICC simulation. . . . .	84
4.2	The hierarchical amplitudes measured for the dark matter as a function smoothing scale. . . . .	87
4.3	The fractional scatter, $\sigma(S_{\text{N}})/S_{\text{N}}$ , in the measured hierarchical amplitudes, as estimated from the 50 simulations in the L-BASICC ensemble. . . . .	88
4.4	The hierarchical amplitudes of dark matter haloes, plotted as a function of the peak height corresponding to the halo mass. . . . .	90
4.5	Volume-averaged $i + j$ -point cross-correlation functions, $\bar{\xi}_{i,j}$ , measured for haloes of different mass. . . . .	92
4.6	The fractional error on the four-point cross correlation functions, estimated from the scatter over the L-BASICC runs. . . . .	93
4.7	The bias parameters as a function of halo mass parametrized by $\nu = \delta_c/\sigma(M, z)$ . Each plot shows a different order of bias parameter. In the lower panel of each plot, the residual bias parameters for the 20% of haloes with the highest or lowest values of $V_{\text{max}}/V_{200}$ , a proxy for concentration, are plotted. . . . .	95

5.1	Differential number of substructures per host halo as a function of their mass relative to that of the host halo. . . . .	105
5.2	The distribution of the fractional mass of the 1st, 2nd and 3rd largest substructures in haloes of different mass at $z = 0$ . . . . .	110
5.3	The number density of subhaloes relative to the mean within $r_{\text{host}}$ , as a function of the distance to the centre of their host halo, in units of the radius of the host halo. . . . .	112
5.4	The probability density distribution of the cosine of the angle between the angular momentum vector of the host halo and the vector joining its centre with that of the subhalo. . . . .	114
5.5	The mean number of satellite mergers per subhalo and per unit of time relative to the age of the universe, as a function of the mass of the progenitor of the less massive object involved in the merger. . . . .	117
5.6	The number density of subhalo-subhalo mergers relative to the mean density of subhaloes within $r_{200}$ as a function of the distance to the centre of the host halo. . . . .	118
5.7	Probability distribution of the cosine of the separation angle $\theta$ between the progenitors of two substructures that are going to merge. . . . .	123
5.8	Three representative examples extracted from the MS for each of the two most common configurations between two satellite subhaloes that merge. . . . .	124
5.9	Contour plot showing the logarithm of the number of satellite-satellite mergers as a function of the masses of the merging subhaloes at $z = 0.5$ . . . . .	126
5.10	The fraction of substructures that have experienced a merger with another substructure since the time of accretion into the current host halo. . . . .	127
5.11	The mean number of satellite-central subhalo mergers per subhalo and per unit of time as a function of the subhalo mass. . . . .	131
6.1	The mean number of haloes per logarithmic mass bin, as a function of their mass. . . . .	139
6.2	The linear bias parameter as a function of halo mass. . . . .	140
6.3	The linear bias parameter for sub-resolution haloes measured in redshift space divided by that measured in real space. . . . .	142
6.4	Number of LRGs as a function of the host halo mass. . . . .	144
6.5	The mean and variance of the LRG correlation function from our catalogues at $z = 0.5$ . . . . .	146

6.6	The variance in the 2-pt correlation function measured from 50 LRGs catalogues in real and redshift space at $z = 0.5$ . . . . .	147
6.7	The normalized covariance matrix from 50 LRG mock catalogues. . . . .	148
7.1	The projected halo density in the light-cone in a slice of depth 50 Mpc. . .	157
7.2	The mass function of haloes measured in the light-cone over the redshift range $0 < z < 1$ . . . . .	158
7.3	The two-point correlation function for haloes in the light-cone. . . . .	159
7.4	The linear bias parameter as a function of scale for haloes in the light-cone.	161

# List of Tables

2.1	The values of some of the basic parameters used in the simulations. . . . .	14
2.2	The results of applying the general fitting procedure described in §2.4 to power spectra measured for different galaxy catalogues at $z = 0$ . . . . .	53
2.3	Same as 2.2 but at $z = 1$ . . . . .	54
5.1	The best-fit parameters to the mass function of subhaloes residing in haloes of different mass at $z = 0$ , using Eq. 5.2 . . . . .	107

## Declaration

The work described in this thesis was undertaken between 2005 and 2008 while the author was a research student under the supervision of Professor Carlos Frenk, Dr. Carlton Baugh & Dr. Cedric Lacey in the Department of Physics at the University of Durham. This work has not been submitted for any other degree at the University of Durham or any other University.

Chapter 2 of this thesis has been published in the form of a paper,

- Angulo, R. E., Baugh, C. M., Frenk, C. S and Lacey, C. G.,  
“The detectability of baryonic acoustic oscillations in future galaxy surveys”,  
2008, MNRAS, 383, 755A.

The simulations presented in Chapter 2 have been used in the following paper:

- Sanchez, A. G., Baugh, C. M. and Angulo, R. E.,  
“What is the best way to measure baryonic acoustic oscillations?”,  
2008, MNRAS, 390, 1470S.

Chapter 3 is to be published as part of the following paper:

- Cai Y-C., Angulo R. E., Baugh C. M., Cole S., Frenk C. S and Jenkins A.,  
“Mock galaxy catalogues and their application to Pan-STARRS”.  
2008, astro-ph/0810.2300

Chapter 4 has been published in the form of a paper,

- Angulo, R. E., Baugh, C. M. and Lacey, C. G.  
“The assembly bias of dark matter haloes to higher orders”.  
2008, MNRAS, 387, 921A.

Chapter 5 is to be published in the form of a paper,

- Angulo, R. E. Lacey, C. G, Baugh, C. M. and Frenk C. S. ,  
“The fate of substructures in cold dark matter haloes.”  
2008, astro-ph/0810.2177

## Acknowledgements

Firstly, I must thank Carlton Baugh, Carlos Frenk and Cedric Lacey who made my PhD at Durham an extremely enjoyable experience. I cannot imagine better supervisors. The work presented in this thesis, is undoubtedly a result of their endless teachings, advice, motivation, encouragement, support and especially patience during these three years. In addition, their integrity and enthusiasm have been an enormous inspiration and example to me.

My passion for cosmology would not have been the same without the challenging and motivating interaction I had with my advisors and several other people. Special thanks to Richard Bower, Shaun Cole, Lian Gao, Adrian Jenkins, Darren Reed, Ariel Sanchez and Tom Theuns. I also cannot forget to thank those who first opened my eyes to the world of science and astronomy: Felipe Barrientos, Marcio Catelan, Nelson Padilla and Boriz Quiroz.

I would like to thank to Lydia Heck for invaluable technical and computer support without which this work could have not been carried out. Thanks also to Lindsay and Dorothy for all the times that you helped me with numerous issues.

I am grateful to my friends at the ICC for uncountable entertaining moments and pleasant conversations about cosmology, life and the world. They are: Phil Bett, Yanchuan Cai, Rob Crain, Greg Davies, Nikos Fanidakis, Juan E. Gonzalez, Alvaro Orsi, Milan Raičević and Tim Rawle.

Finally, Heather, I am truly grateful to you for your constant love, help and support. Your company made the final stage of my PhD the most enjoyable period of my life.

This thesis was supported by a PPARC/British Petroleum sponsored Dorothy Hodgkin postgraduate award.



# Chapter 1

## *Introduction*

The publication of the theory of relativity by Einstein in 1916 revolutionized our understanding of the Universe and led to the birth of cosmology as a predictive and quantitative science. Subsequent advances along both observational and theoretical fronts have delivered a comprehensive and robust insight into the properties of the Universe. Nevertheless, there are key pieces missing from the picture challenging fundamental aspects of cosmology. The subject of this thesis is to provide further knowledge that could help to solve some of these problems.

In the first section of this introductory chapter (§1.1), we describe the simplest, most favoured description for the Universe - the Cosmological Constant-Cold Dark Matter (or  $\Lambda$ CDM) model. Note that since there are many excellent textbooks which carefully set out the main ideas of modern cosmology (e.g. Peebles, 1980, 1993; Padmanabhan, 1993; Peacock, 1999), we will simply provide a very focused and brief introduction to put into context the chapters that make up this thesis. It is not our objective to provide a detailed introduction to the field.

In the second section (§1.2), we describe the specific contents of each of the subsequent chapters of this thesis.

### 1.1 The $\Lambda$ CDM model

The current paradigm for the origin and evolution of structure in the Universe arguably rests on two pillars. The first is the presence of a class of as of yet unidentified particles, the “Cold Dark Matter” (CDM) that provides the link between the early, smooth Universe and the highly inhomogeneous one that we observe today. The second, “Inflation” explains the flat, homogeneous and isotropic appearance of the Universe as well as generating the initial seeds for the formation of structure.



### 1.1.1 Dark matter

Although the nature of the dark matter remains uncertain, its abundance and properties have been long supported by independent astrophysical observations. For instance, the influence of dark matter is evident in the rotation curves of galaxies, in the distortion of the light travelling from distant galaxies and in the large scale distribution of galaxies and clusters. The importance of dark matter lies in the fact that it interacts only through gravity. Reaching about 25 percent of the total energy budget of the Universe, dark matter is the dominant matter component and hence plays a crucial role in the formation of structure. The collisionless character of the dark matter is a enormous advantage for understanding the final matter distribution, since the much more complicated and poorly understood gas physics laws can be neglected in the evolution of the large scale structure. In this way, gravitational instability alone provides the framework for the growth of density perturbations in the Universe.

Alternative models where dark matter is not required have also been proposed. One of the most popular is a phenomenological modification of Newton's law of gravity on large scales called MOND (Bekenstein, 2004). Initially in such scenarios, only the rotation curves of galaxies were naturally explained, but recent extensions have been introduced to the theories to explain several other observations. In spite of this, these ideas have not been convincing in explaining gravitational lensing in places where there is no visible matter (Clowe et al., 2006), or in reproducing the current large scale structure given the observations of the early Universe. Nevertheless, there is still debate about the exact nature of gravity on cosmological scales (Angus et al., 2006).

Although there is a compelling set of astrophysical evidence suggesting that dark matter is non-baryonic, collisionless, dissipationless and possesses small random initial velocities, it has not been detected directly. However, this situation may soon be about to change. There are several experiments designed to distinguish and detect different dark matter candidates proposed by super-symmetric theories (see Bertone et al., 2005, for a review of current candidates and evidence). On one hand, cryogenic particles detectors such as DAMA, Xenon10, CDMS or CRESST are carrying out direct dark matter searches. On the other hand, the "Large Hadron Collider" or the "International Linear Collider" could reach the energy levels required to unveil or rule out different proposed particles. On a different front, a parallel quest is focused on the sky, where annihilation between dark matter particles could be detected, by satellites such as PAMELA, in the form of  $\gamma$ -

rays from the centre of clusters, the places in the Universe where the dark matter density is expected to be the highest.

### 1.1.2 Inflation

The second pillar of the current paradigm is an “inflationary period” during which the Universe underwent an exponential expansion (Guth, 1981). This imprints distinctive and key properties on the Universe.

The first of these features is the presence of macroscopic anisotropies which result from the amplification of initial quantum fluctuations. The perturbations are assumed to be Gaussian and adiabatic with a nearly scale-invariant power spectrum (Starobinskij, 1982). The Fourier amplitudes follow a power law with index approximately equal to 1 ( however recent measurements prefer a value slightly lower than unity e.g. Sánchez et al., 2006). Later, the primordial power spectrum is modified during the early Universe due to a scale dependent growth rate, which results from perturbations entering the horizon at different times. The exact change of the shape depends on the type of dark matter and the relative densities of the components in the Universe. In particular, if the dark matter is cold, perturbations that entered the horizon prior to matter radiation equality cannot grow due to the collapse time scale being much longer than the time scale for the expansion of the Universe (Meszaros, 1974). This produces a damping in the power spectrum at small scales. On the other hand, baryonic perturbations of different sizes arrive at different phases of the cycle produced by their coupling with radiation (Peebles and Yu, 1970). This imprints oscillations on the primordial spectrum. The second main outcome of inflation was that any hypothetical initial non-zero spatial curvature was smoothed, leaving the universe with a energy density equal to the critical density associated with a flat geometry. Although it is impossible to observe directly the inflationary period, the consequences of this epoch have been successfully measured in the cosmic microwave background radiation (CMB) by many experiments, most notably by the satellite missions COBE (Cosmic Background Explorer, Smoot et al., 1992) and WMAP(Wilkinson Microwave Anisotropy Probe, Spergel et al., 2003).

### 1.1.3 Cosmological parameters

The  $\Lambda$ CDM Universe is highly quantifiable thanks to that the main parameters of the model manifest themselves in a multitude of observables. For this reason, along with the ideas outlined above, the accurate census of the main cosmological parameters has been

of central importance in the establishment of the paradigm.

Given the variety of constraints on the main cosmological parameters, it is remarkable that modern observations have consistently pointed towards the same region in parameter space (although there are degeneracies between the parameters and controversy regarding the exact number of variables that are needed to describe the cosmological model). Currently, the current best estimates for the amount of baryons, dark matter and dark energy expressed in terms of the critical density for a flat geometry are respectively  $\Omega_b = 0.0462 \pm 0.0015$ ,  $\Omega_c = 0.233 \pm 0.013$ ,  $\Omega_\Lambda = 0.721 \pm 0.015$  (Komatsu et al., 2008). These numbers are derived from the measurements of the power spectrum of the anisotropies in the CMB radiation combined with the large scale structure and Supernovae type Ia.

Perhaps the most striking result derived from these measurements is a flat geometry *and* a joint matter and radiation content much smaller than the critical density (note that this was already suggested more than 20 years ago). This implies that most of the energy budget (about 75%) in the Universe is in the mysterious form of dark energy. In one of the simplest possibilities, the dark energy corresponds to a cosmological constant,  $\Lambda$ , which generates negative pressure and, if dominant, can produce enough force to drive the Universe into an accelerated expansion. Indeed, during the first half of its life the Universe was dominated by matter and it continuously expanded but at a progressively lower rate. However, as the matter density drops, the dark energy overtakes and initiates a second period of acceleration in the history of the Universe. This incredible scenario was first directly suggested by the Hubble diagrams built from observations of Supernovae Type Ia (Riess et al., 1998; Perlmutter et al., 1999) but subsequent independent measurements confirmed the idea and now an accelerating Universe has become widely accepted.

### 1.1.3.1 Dark Energy

Clearly, along with accurate measurements of the main components and properties of our Universe a huge question arose: What is the nature of the dark energy which is driving the current accelerating expansion of the Universe? This enigma has been one of the biggest problems in theoretical physics during the last 10 years. Moreover, it represents perhaps the most drastic challenge for the  $\Lambda$ CDM model. In spite of the difficulty of the question, the answer may start to become apparent very soon.

The next generation of cosmological experiments will have as a primary science goal the measurement of the properties of this mysterious force. If the current forecasts are

correct, then the constraints on the properties of the dark energy will be very tight, which will distinguish between the candidates for this substance, and hence, we could start to uncover the nature of what fills our Universe.

There are many promising approaches to reveal the properties of the dark energy, including gravitational lensing (Huterer et al., 2006), the abundance of clusters (Haiman et al., 2001), and Supernovae Ia measurements (Wood-Vasey et al., 2007). But one of the most robust estimations relies on the use of the imprint of baryonic acoustic oscillations (BAO) on the pattern of galaxy clustering. This feature can be measured very precisely in galaxy surveys, relying only on the measurement of galaxy positions. By combining the physical size with the observed size of the BAO, we have a standard ruler in the Universe with which it is possible to constrain the properties of the dark energy accurately. Nevertheless, the level of accuracy demanded for competitive and useful constraints requires an understanding of the systematic errors in this approach since they can be larger than the random uncertainties expected from future surveys, degrading the measurements and constraints. In the first two Chapters of this thesis we explore the detectability of BAO in future galaxy surveys. In order to be able to exploit the forthcoming data fully, realistic modelling of galaxy clustering and a understanding of the observational systematics are essential. Hence in these chapters we pay special attention to these aspects.

Of course, with the ever increasing accuracy required by the measurements and the imminent arrival of the largest galaxy surveys in the history of astronomy, the level of understanding of uncertainty in our modelling of galaxy clustering will play a crucial role in the constraints we are able to derive. In the final chapters of this thesis, we present some tools to help in this direction. In particular, in Chapter 6 we introduce a new technique which allows us to extend the dynamical range probed by a numerical simulation. In Chapter 7, I illustrate how an extra level of sophistication can be introduced in the prediction of numerical simulations by constructing a light-cone.

#### 1.1.4 Growth of structures

With the initial conditions and the main components of the Universe set, its evolution and final state are, in principle, also determined. As the radiation temperature decreased as a consequence of the expansion of the Universe, the baryonic matter recombined and then decoupled from the photons. At this moment the overdensities in the distribution of baryons and dark matter were free to grow due gravitational instability. Denser regions become denser while underdense regions become more underdense. The study of the evo-

lution of these perturbations is probably one of the most interesting and elegant branches of cosmology. As long as the perturbations are small, their dynamic and growth can be computed analytically by expanding the equations of motion to either linear or higher orders (e.g. Bernardeau et al., 2002). However, once the perturbations are in the highly nonlinear regime their evolution can be only accurately followed by numerical simulations (see the review by Bertschinger, 1998).

Once the density enclosed in a particular region exceeds some critical value, it decouples from the global expansion of the Universe and collapses, forming a dark matter halo. The process occurs hierarchically in the CDM models: small structures collapse first, then, they increase their mass by both smooth accretion of dark matter particles or by merging with other collapsed structures (Lacey and Cole, 1993). Note that, if the dark matter is hot, any early small perturbation is smeared out due to free-streaming and the first objects that form are big clusters which then fragment to form smaller haloes.

In the simultaneous process, attracted by the substantial gravitational force exerted by dark matter haloes, baryons (gas) fall, heat up by shocks and get trapped within dark haloes. The gas eventually cools by radiative processes, condenses and fragments to form stars and galaxies (White and Rees, 1978). Although the main processes of galaxy formation are relatively well established, it is extremely difficult to follow them in detail. For instance, star formation, the multi-phase nature of the interstellar medium, the return of energy back into the medium by supernovae or AGN, are processes that cannot be described nor followed from first principles and therefore, after semi-analytic techniques are developed to predict the properties of galaxies and consequently obtain the ultimate aim of any theory; comparison with observations (Baugh, 2006).

It is therefore clear that, the abundance, clustering and merger histories as well as internal properties of dark matter haloes are closely linked to fundamental properties of the galaxies they host. Given this relationship, it is also clear that that the study of such properties is necessary to understand the observable Universe. In particular, one of the most important properties in the study of the spatial distribution of galaxies is the clustering strength of the haloes that host such galaxies. The standard model (for a review see Zentner, 2007), the amplitude of the clustering of haloes is only a function of their mass, with massive haloes being more clustered than their less massive counterpart. In Chapter 4, motivated by previous work where the dependence of the linear bias parameter on the internal properties of haloes was investigated, we revisit this prediction and test it using higher order statistics. In principle, this effect could also be observed in galaxy

clustering (Croton et al., 2007) offering a further test for the ideas that underpin our current vision of the growth of the structure in Universe.

The hierarchical formation of structures leaves a signature of CDM models inside dark matter haloes: The presence of numerous self-bound structures corresponding to remnants of the previously accreted haloes (e.g. Klypin et al., 1999b; Moore et al., 1999). This offers a direct test of the  $\Lambda$ CDM model. Unfortunately, the relationship between subhaloes and observed satellite galaxies is complicated, usually depending on the details of the adopted galaxy formation model, which therefore makes a direct comparison infeasible. In spite of this, the properties of the subhalo population still provides the framework for current comparisons between theory and observations. In the fifth Chapter of this thesis, using some of the largest cosmological simulations currently available, we explore the properties of subhaloes. Moreover, gravitational lensing could directly detect these objects (Kochanek and Dalal, 2004) in which case a robust theoretical prediction is required.

The final distribution of galaxies on cosmological scales is far from homogeneous. In fact, it displays a very distinctive pattern as seen by the latest generation of galaxy surveys, the 2-degree Field Galaxy Redshift Survey (2dFGR, Colless et al., 2001) and the Sloan Digital Sky Survey (SDSS, York et al., 2000). Surprisingly, the vast richness of cosmic structures ranging from filaments, voids and clusters of very dissimilar size, is simply a consequence the gravitational amplification of the initial fluctuations seeded by inflation (Davis et al., 1985; Springel et al., 2005).

## 1.2 Outline of this thesis

In this thesis, I will work mainly in three areas related to the spatial distribution of structures on cosmological scales. In one line of research, we focus on the large scale distribution of galaxies and on the detectability of BAO in power spectrum measurements. In a second line, we investigate the clustering of dark matter haloes and its substructures. In particular, we first look at the higher order clustering of haloes and then we concentrate on the properties of the substructures. In the third line of research I explore how to add more realism to current theoretical predictions and how it would be possible to investigate uncertainties in future measurement by using low resolution dark matter simulations. I continue by describing the contents of each chapter in more detail.

In Chapter 2, we combine a specially designed N-body simulation with a semi-analytic model of galaxy formation to study in detail the visibility of the BAO in forthcoming

galaxy surveys. An interesting issue that we present in this chapter is how changes in the power spectrum due to galaxy bias, redshift-space distortions and nonlinear evolution degrade the visibility of the BAO. We then apply realistic methods to extract information about the dark energy equation of state, and in this way we forecast the performance of future cosmological experiments.

Since many of the currently planned galaxy surveys will be carried out in photometric space, i.e. the distance of galaxies will be inferred from their photometric redshift, in Chapter 3, we extend the analysis presented in Chapter 2 to include the effects of such an approach.

Chapter 4 investigates the dependence of the clustering of dark matter haloes on internal properties. Here we expand previous studies towards higher order statistics. We first present a novel approach to estimate the higher order correlation functions of haloes which we then apply to N-body simulations to compute, in narrow mass bins, the bias parameters up to order four. Finally, we explore whether the high order clustering of haloes correlates with an internal property of the sample, namely the concentration of the halo.

We then zoom to smaller scales to discuss the characteristics of the halo substructures. In Chapter 5 we use the Millennium Simulation (Springel et al., 2005) to investigate some general properties of these structures. In particular, we examine the mass function, radial distribution and spatial orientation of substructures with respect to their host halo. In the second half of this chapter we study the possibility of substructure mergers and destruction.

This thesis ends with two short studies. Chapter 6 contains a proof of concept about how to extend the dynamical range over which low resolution dark matter simulations can be used. In this chapter we present and test the algorithm that creates a halo population based on the dark matter density. Finally, we apply the procedure to create an ensemble of galaxy catalogues and in this way investigate the uncertainties in the two-point correlation function as well as the covariance matrix. Chapter 7 illustrates and discusses a method to create a light-cone containing consisting of haloes.

Finally, In Chapter 8 we provide a brief summary of the main conclusions of this thesis and list some prospects for future work.

# Chapter 2

## *The detectability of baryonic acoustic oscillations in future galaxy surveys*

### 2.1 Introduction

The discovery that the rate of expansion of the Universe is apparently accelerating was one of the key advances in physical cosmology in the 1990s (Riess et al. 1998; Perlmutter et al. 1999). Understanding the nature of the dynamically dominant dark energy, which is believed to be responsible for this behaviour, is one of the biggest challenges now facing cosmologists.

Over the past decade our knowledge of the basic cosmological parameters, which describe the content of the Universe, its expansion history and ultimate fate has improved tremendously. This progress is the result of advances on two fronts: the advent of datasets which have provided fresh views of the Universe with unprecedented detail and the development of the theoretical machinery required to interpret these new measurements. Currently, the values of many cosmological parameters are known to an accuracy of around 10% (albeit with caveats regarding degeneracies between certain combinations of parameters and also regarding the precise number of parameters that are allowed to vary in the cosmological model; see, for example, Sánchez et al. 2006).

The cold dark matter (CDM) model has emerged as the most plausible description of our Universe. In the most successful version of this model, more than 70% of the density required to close the Universe is in the form of dark energy. Currently, there is no model which can reconcile the magnitude of the dark energy component with the value expected from particle physics arguments. A simple phenomenological description of the

dark energy is provided by the equation of state that relates its pressure,  $P$ , and density,  $\rho$ , which is encapsulated in the parameter  $w = P/\rho c^2$ . If the dark energy has the form of the cosmological constant,  $w = -1$ . The indications are that the dark energy now has a form close to that expected for a cosmological constant (Riess et al. 2004; Sánchez et al. 2006). However, in the absence of a theoretical model for the dark energy, it is possible that the equation of state could depend on space and/or time.

A whole range of experiments and surveys is being planned which number amongst their goals determining the equation of state of the dark energy as a function of redshift (for a discussion, see Albrecht et al. 2006 and Peacock and Schneider 2006). Several techniques are being considered, which are sensitive to the influence of the dark energy on various features of the cosmological world model. These include the Hubble diagram of Type IA supernovae, counts of clusters of galaxies, the weak gravitational lensing pattern of faint galaxies and the measurement of the baryonic acoustic oscillation scale in the matter distribution as a function of redshift. The measurements and data analysis required to obtain useful constraints on the equation of state parameter are so demanding, and so open to potential systematic errors, that it is necessary to pursue as many different avenues as possible.

In this chapter, we focus on the test using the baryonic acoustic oscillations (BAO). The BAO is the name given to a series of peaks and troughs on scales on the order of  $100 h^{-1}$  Mpc, imprinted on the power spectrum of matter fluctuations prior to the epoch of last scattering, when the matter and radiation components of the Universe were coupled (Peebles and Yu 1970). The BAO are the counterpart of the acoustic peaks seen in the power spectrum of the temperature of the cosmic microwave background radiation, though they have a different phase and a much smaller amplitude (Sunyaev and Zel'dovich 1970; Press and Vishniac 1980; Hu and Sugiyama 1996; Eisenstein and Hu 1998; Meiksin et al. 1999). The wavelength of the BAO is related to the size of the sound horizon at recombination. This does not depend on the amount or nature of the dark energy, but on the physical density of matter ( $\Omega_m h^2$ ) and baryons ( $\Omega_b h^2$ ). Given the values of these parameters, for example, from the cosmic microwave background or large scale structure data, the sound horizon scale is known and can be treated as a standard ruler. The *apparent* size of this feature in the power spectrum of galaxies or galaxy clusters does depend on the dark energy and its equation of state through the angular diameter distance-redshift relation (e.g. Blake and Glazebrook 2003; Hu and Haiman 2003)

BAO in the galaxy distribution were first glimpsed in the early stages of the “2-

degree-field galaxy redshift survey” (Percival et al. 2001) and finally detected in the power spectrum of the completed 2dFGRS (Cole et al. 2005). The equivalent feature, a spike, was also found in the correlation function measured from the luminous red galaxy (LRG) sample of the Sloan Digital Sky Survey (SDSS) (Eisenstein et al. 2005). Cole et al used the BAO to constrain the parameter combination  $(\Omega_M/\Omega_b, \Omega_M)$  (where  $\Omega_M$  and  $\Omega_b$  denote the matter and baryon density parameters respectively). Eisenstein et al used the location of the spike in the correlation function to constrain the absolute distance to the median redshift of the SDSS LRG sample and hence constrained the value of  $\Omega_M$ . Hütsi (2006b,a) carried out a power spectrum analysis of a similar LRG sample, and combined this measurement with other datasets to constrain the values of cosmological parameters. More recently, the BAO have been extracted from the power spectrum measured from a much larger sample of SDSS LRGs to constrain  $\Omega_M$  and  $\Omega_b/\Omega_m$  (Tegmark et al. 2006; Blake et al. 2007; Padmanabhan et al. 2007; Percival et al. 2007). To date, measurements of the BAO have only yielded constraints on the dark energy equation of state when combined with other datasets, such as the spectrum of temperature fluctuations in the microwave background or when restrictive priors have been adopted on certain parameters, such as the Hubble constant.

The bulk of the work in the literature on the usefulness of the BAO has relied upon linear perturbation theory to assess the detectability of the features and to forecast the errors on the recovered value of  $w$  (Blake and Glazebrook 2003; Hu and Haiman 2003; Glazebrook and Blake 2005; Blake and Bridle 2005; Blake et al. 2006; Parkinson et al. 2007). There are, however, a range of dynamical and statistical effects which can alter the appearance of the power spectrum relative to the linear theory prediction, even on the scale of the BAO, which we review in this chapter (Seo and Eisenstein 2003; Angulo et al. 2005; Springel et al. 2005; Seo and Eisenstein 2005; Eisenstein et al. 2007b). Some simulation work has been done to study these effects, mostly using computational cubes of side  $500 h^{-1}$  Mpc (Seo and Eisenstein 2003, 2005; Springel et al. 2005; Eisenstein et al. 2007b). These are only a small factor (2-3) bigger than the scale of the fluctuations of interest. Calculations with small boxes are subject to large sampling fluctuations and may even miss some features of the nonlinear growth of large scale fluctuations through the absence of long wavelength density fluctuations (Crocce and Scoccimarro 2006b). Very recently, larger simulation volumes have been used, of around a cubic gigaparsec and larger (Schulz and White 2006; Huff et al. 2007; Angulo et al. 2005; Koehler et al. 2007). However, such studies have tended to have relatively poor mass resolution, making it

difficult to model galaxies without resorting to simplified biasing prescriptions (e.g. Cole et al. 1998).

Given the significant commitment of resources required by the proposed galaxy surveys and the level of precision demanded by the BAO approach, it is imperative to ensure that accurate theoretical predictions are available both to help in the design of the survey strategy and to extract the maximum amount of information from the observations. This is a tough challenge computationally, because it requires ultra-large volume N-body simulations with sufficient mass resolution to identify the haloes likely to host the galaxies to be seen in the surveys, and a realistic model to populate these haloes with galaxies.

In this chapter, we use a combination of suitable N-body simulations and a semi-analytical model of galaxy formation to assess the visibility of the BAO. In Section 2.2, we describe the suite of N-body simulations used and outline the semi-analytical model. Section 2.3 gives a blow-by-blow account of how the power spectrum changes relative to the simple prediction of linear perturbation theory, as additional layers of realism are added to the modelling, starting with dark matter and ending with galaxies. We set out our approach for constraining the dark energy equation of state in Section 2.4, and present our results in Section 2.5. We give our conclusions in Section 2.6.

## 2.2 Method

In this section, we introduce the theoretical tools used to produce synthetic galaxy catalogues. First, we describe the N-body simulations (§2.2.1) which consist of a high resolution run (§2.2.1.1) and an ensemble of lower resolution runs (§2.2.1.2). Next, we discuss the measurement of power spectra from discrete distributions of objects and use the ensemble of low resolution simulations to estimate the errors on the power spectrum measurement (§2.2.1.3). In the second part of this section, we explain how a galaxy formation model is used to populate the high resolution N-body simulation with galaxies (§2.2.2).

### 2.2.1 N-Body Simulations

The N-body method is a long-established computational technique which is used to follow the growth of cosmological structures through gravitational instability (see, for example, the reviews by Bertschinger 1998 and Springel et al. 2006). Our goal in this chapter is to simulate the formation of structure within a sufficiently large volume to follow the growth of fluctuations accurately on the scale of the BAO, and with similar statistics for

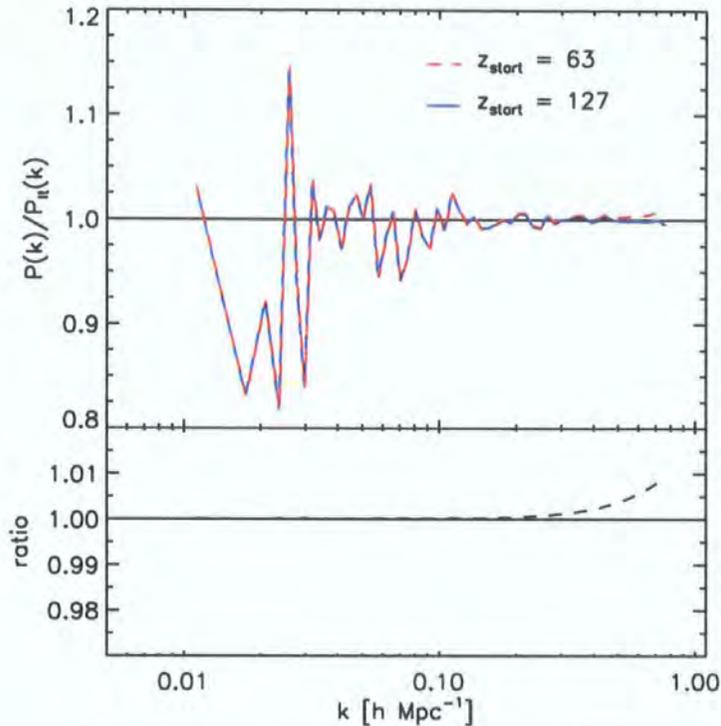


Figure 2.1: A test of the choice of starting redshift used in the N-body simulations. The upper panel compares the power spectrum measured at  $z = 15$  in the BASICC when the simulation is started at  $z = 63$  (dashed red curve) and at  $z = 127$  (solid blue curve). The power spectra plotted in the upper panel have been divided by the linear perturbation theory prediction for the dark matter power spectrum at  $z = 15$ . The lower panel shows the ratio between the power spectrum measured from the simulation started at redshift 63 to that measured from the run which started at redshift 127.

power spectrum measurements to those expected in forthcoming surveys. At the same time, we require a mass resolution which is adequate to identify the dark matter haloes likely to host the galaxies which will be seen in these surveys. To achieve these aims, we use a memory-efficient version of the GADGET-2 code of Springel (2005), which was kindly provided to us by Volker Springel and the Virgo Consortium.

We use two types of calculation: a high resolution simulation, labelled the ‘‘Baryon Acoustic Simulation at the ICC’’ or BASICC, which is able to track galactic haloes, and an ensemble of lower resolution simulations, labelled L-BASICC, which we use to study the statistics of power spectrum measurements on large scales. Here, we describe some

	$N_p$	$m_{\text{dm}}$	$\epsilon$
		$[h^{-1} M_{\odot}]$	$[h^{-1} \text{kpc}]$
BASICC	$3.03 \times 10^9$	$5.49 \times 10^{10}$	50
L-BASICC	$8.99 \times 10^7$	$1.85 \times 10^{12}$	200

Table 2.1: The values of some of the basic parameters used in the simulations. The columns are as follows: (1) The name of the simulation. (2) The number of particles. (3) The mass of a dark matter particle. (4) The softening parameter used in the gravitational force. In both cases, the length of the computational box is  $1340 h^{-1} \text{Mpc}$ , and the same cosmological parameters are used, as given in Section 2.2.1.

of the common features of the simulations, before moving on to outline specific details in §2.2.1.1 and §2.2.1.2.

We adopt a  $\Lambda$ CDM cosmology with the same parameters used in the Millennium Simulation (Springel et al. 2005), which are broadly consistent with the latest constraints from the cosmic microwave background data and large scale structure measurements (Sánchez et al. 2006; Spergel et al. 2007). The values of the parameters are: the matter density parameter,  $\Omega_M = 0.25$ , the energy density parameter for the cosmological constant,  $\Omega_{\Lambda} = 0.75$ , the normalization of density fluctuations,  $\sigma_8 = 0.9$  and Hubble constant,  $h = H_0/(100 \text{kms}^{-1} \text{Mpc}^{-1}) = 0.73$ .

Due to memory restrictions, the Fourier mesh used to set up the initial particle displacements has a dimension of  $1580^3$  grid points which is not commensurate with the cube root of the particle number mesh. We therefore avoided using a regular particle grid to set up the initial conditions, as this would have led to a spurious feature in the power spectrum of the initial conditions at the beat frequency between the particle grid and the Fourier mesh. Instead, we used a glass-like distribution (White 1994; Baugh et al. 1995). The input power spectrum of density fluctuations in linear perturbation theory is calculated using the CAMB package of Lewis et al. (2000). The amplitude of the Fourier modes is drawn from a Rayleigh distribution with mean equal to the linear theory power spectrum and the phase is drawn at random from the interval 0 to  $2\pi$ . The initial density field is generated by perturbing particles from the glass-like distribution, using the approximation of Zel'dovich (1970).

The simulations were started at a redshift of  $z = 63$ . The Zel'dovich (1970) approxi-

mation used to set up the initial pattern of density fluctuations produces transients which can be seen in clustering signal measured for the dark matter at expansion factors close to the starting redshift (Efstathiou et al. 1985; Baugh et al. 1995; Crocce et al. 2006). Later on, we will use the power spectrum from a high redshift output from the simulation,  $z = 15$ , as a proxy for linear perturbation theory, so it is important to check that this power spectrum in particular, and also the power spectra measured at all subsequent outputs are insensitive to the choice of starting redshift. We test this by comparing the power spectrum of the dark matter at  $z = 15$  in our standard run with the spectrum measured in a test run which started at  $z = 127$ , but which did not run all the way through to  $z = 0$ . The top panel of Fig. 2.1 shows that the power spectra measured for the dark matter in these two cases, divided by the power spectrum predicted by linear perturbation theory at  $z = 15$ . The fluctuations in the measured power at low wavenumbers around the linear theory prediction reflect the sample variance noise which is not negligible even in a simulation of the volume of the BASICC. The lower panel in Fig. 2.1 shows the  $z = 15$  power spectrum measured from the run started at  $z = 63$  divided by that measured from the run started at  $z = 127$ . At large wavenumbers, the effect of transients is visible, although quite small,  $\sim 1\%$ . The focus of this chapter, however, is the form of the power spectrum over wavenumbers smaller than  $k = 0.4 h\text{Mpc}^{-1}$ , for which the spectra measured at  $z = 15$  for the two different choices of starting redshift agree to better than 0.3%. Our results are therefore unaffected by any transients resulting from the use of the Zel'dovich approximation.

### 2.2.1.1 The high resolution simulation: the BASICC

The BASICC simulation covers a comoving cubical region of side  $1340 h^{-1} \text{Mpc}$ , in which the dark matter is represented by more than 3 billion ( $1448^3$ ) particles. The equivalent Plummer softening length in the gravitational force is  $\epsilon = 50 h^{-1} \text{kpc}$ , giving a dynamic range in length of almost 27,000. The volume of the computational box,  $2.41 h^{-3} \text{Gpc}^3$ , is almost twenty times the volume of the Millennium Simulation (Springel et al. 2005), and more than three times the volume of the catalogue of luminous red galaxies from the SDSS used to detect the acoustic peak by Eisenstein et al. (2005). The BASICC volume is within a factor of two of that proposed for a survey with WFMOS at  $z \sim 1$  (Glazebrook and Blake 2005). The simulation occupied the full 0.5 Terabytes of RAM of the second upgrade of the Cosmology Machine at Durham. The run took 11 CPU days on 506 processors, the equivalent of 130,000 CPU-hours.

The particle mass in the BASICC simulation is  $m_p = 5.49 \times 10^{10} h^{-1} M_\odot$ . This is approximately 64 times larger than the particle mass used in the Millennium Simulation. The mass resolution limits the usefulness of dark matter halo merger trees from the BASICC, so we have chosen to output at a modest selection of redshifts:  $z=0, 0.3, 0.5, 1, 2, 3, 4, 6, 8, 10, 15$  and 63. Each of these outputs occupies  $\sim 100$  Gb of disk space. In each snapshot we have identified groups of dark matter particles using a friends-of-friends algorithm (Davis et al. 1985) with a linking length of 0.2 times the mean inter-particle separation. We have stored groups with 10 or more particles, i.e. haloes more massive than  $5.49 \times 10^{11} h^{-1} M_\odot$ . There are 17 258 579 haloes in the  $z = 0$  output of the simulation with ten or more particles. The most massive halo has a mass of  $6.74 \times 10^{15} h^{-1} M_\odot$  and 860 haloes have a mass in excess of the Coma cluster ( $\approx 10^{15} h^{-1} M_\odot$ ).

The BASICC simulation sits between the Millennium and Hubble Volume (Evrard et al., 2002) simulations. Its unique combination of mass resolution and volume makes it ideal for studying the large scale distribution of galaxies and clusters alike.

### 2.2.1.2 The ensemble of low resolution simulations: L-BASICC

We also generated an ensemble of 50 “low-resolution” simulations to study the sample variance in the BASICC and to test an analytic model for the errors expected on measurements of the power spectrum, which we discuss in the next subsection. These low resolution runs (L-BASICC) have exactly the same cosmological parameters as the BASICC and the same box size (see Table 2.1), but they have fewer particles ( $448^3$ ). For each realization, a different random seed is used to set up the initial density field. The starting redshift of these simulations is  $z = 63$ . The particle mass is comparable to that employed in the Hubble Volume simulation (Evrard et al. 2002). Each L-BASICC simulation took 0.8 days to run on 16 processors of the third upgrade of the Cosmology Machine. The total volume of the ensemble is  $120 h^{-3} \text{Gpc}^3$ , more than four times that of the Hubble Volume, making this a unique resource for studying the frequency of rare objects in a  $\Lambda$ CDM universe. For L-BASICC, the position and velocity are stored for every particle at 4 output times ( $z = 0.0, 0.5, 0.9, 3.8$ ); we also produce a halo catalogue at each redshift retaining objects with ten or more particles (corresponding to a mass of  $1.8 \times 10^{13} h^{-1} M_\odot$ ). As we shall see in later sections, the ensemble allows us to assess whether or not a particular result is robust or simply due to sampling fluctuations. Due to their limited mass resolution, it is not feasible to populate these simulations with galaxies using the method outlined below (§2.2.2).

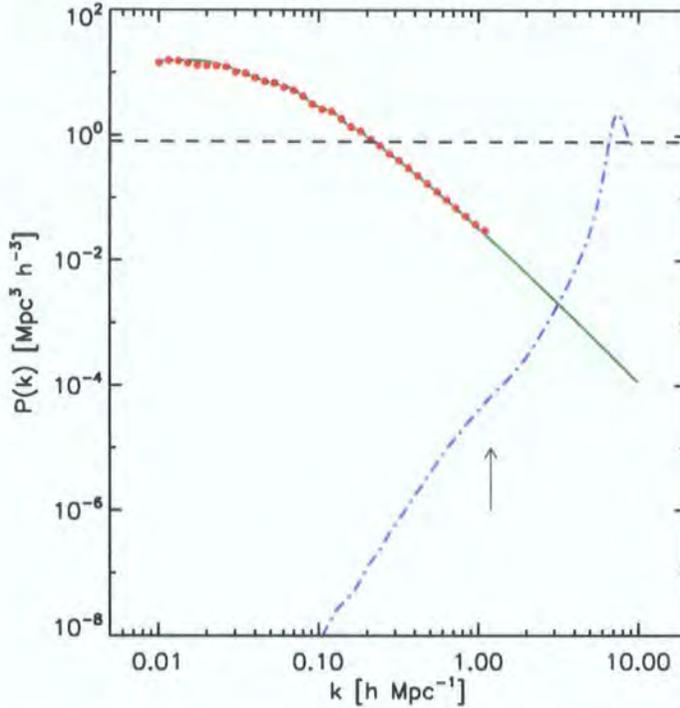


Figure 2.2: The power spectrum of the dark matter in real space measured at the starting redshift of the BASICC,  $z = 63$  (red points). The corresponding prediction of linear perturbation theory is shown by the green (solid) line. The blue (dot-dashed) curve shows the power spectrum of the *unperturbed* glass-like distribution of particle positions. The dashed line shows the Poisson noise expected for the number density of dark matter particles used in the BASICC. The noise of the initial particle distribution is much less than Poisson. The arrow marks the position of the Nyquist frequency of the FFT grid.

### 2.2.1.3 Power spectrum estimation and errors

The two point statistics of clustering, the correlation function, and its Fourier transform, the power spectrum,  $P(k)$ , are the most commonly employed measurements of clustering. In this chapter we focus on the power spectrum; in Sanchez et al. (2008), we address the visibility of the acoustic oscillations in the correlation function. The standard way to quantify the amplitude of a density fluctuation is by means of the density contrast,  $\delta(x, t) = (\rho(x, t) - \bar{\rho})/\bar{\rho}$ . If we consider the Fourier transform of the density contrast,  $\rho_k$ , then the power spectrum is defined as the modulus squared of the mode amplitude,  $P(k) = \langle |\delta_k|^2 \rangle$ .

There are two steps in the computation of the power spectrum from a distribution of discrete objects, such as dark matter particles, dark haloes or galaxies. Firstly, a density field is constructed by assigning the objects to mesh points on a cubic grid. In the simplest mass assignment scheme, the nearest grid point, the contribution of each object to the density field is confined to the cell in which it is located. In higher-order assignment schemes, the mass of the particle is shared with adjacent cells. Here, we use the cloud-in-cell assignment scheme (see Hockney and Eastwood 1981). Secondly, we perform a Fast Fourier Transform of the density field. The power spectrum is obtained by spherically averaging the resulting Fourier mode amplitudes in annuli of radius  $\delta k = 2\pi/L = 0.0047 h \text{ Mpc}^{-1}$ .

The mesh we use to store the density field has  $N_{\text{FFT}}^3 = 512^3$  grid points. Estimating the density on a grid alters the form of the power spectrum at wavenumbers approaching the Nyquist frequency of the grid ( $k_{\text{Nyquist}} = 2\pi/L N_{\text{FFT}}/2 = 1.2 h \text{ Mpc}^{-1}$  in our case). The degree of modification and the precise wavenumber above which the power spectrum is distorted depend upon the choice of assignment scheme (Hatton 1999; Jing 2005). In practice, for the size of FFT mesh we use, this has little impact on the recovered power spectrum for wavenumbers of interest; the measured amplitude differs by less than 1% from the true value at a wavenumber  $k \sim 0.8 h \text{ Mpc}^{-1}$ ; in most cases we focus on the form of the power spectrum on large scales,  $k < 0.4 h \text{ Mpc}^{-1}$ . Nevertheless, we correct for the effects of the cloud-in-cell assignment scheme by dividing each mode by the Fourier transform of a cubical top hat:

$$\delta(k_x, k_y, k_z) \Rightarrow \frac{\delta(k_x, k_y, k_z)}{\text{sinc}^2\left(\frac{k_x L}{2N_{\text{FFT}}}\right) \text{sinc}^2\left(\frac{k_y L}{2N_{\text{FFT}}}\right) \text{sinc}^2\left(\frac{k_z L}{2N_{\text{FFT}}}\right)}, \quad (2.1)$$

where

$$\text{sinc}(x) = \frac{\sin(x)}{x}. \quad (2.2)$$

Note this is different from the approach taken by Jing (2005), who applied a correction to the spherically averaged power spectrum.

A further possible distortion to the form of the measured power spectrum is discreteness noise and the associated Poisson or shot noise. Poisson-sampling a continuous density field with point objects of space density,  $\bar{n}$ , introduces a spurious contribution that should be subtracted from the measured power spectrum:  $P_{\text{corr}}(k) = P_{\text{meas}}(k) - 1/\bar{n}$ . In the case of dark matter halo centres or galaxies, the need for such a correction is jus-

tified. However, in the case of dark matter particles in our simulations, one should *not* subtract Poisson shot noise from the power spectrum because the particles were initially laid down by perturbing a glass-like configuration which is sub-Poissonian in nature. This is clear from Fig. 2.2, which shows the power spectrum measured for the dark matter in the initial conditions of the BASICC. The red curve shows the spectrum measured in the simulation and the smooth green curve shows the input spectrum predicted by linear perturbation theory. The two agree remarkably well over a wide range of wavenumbers. The power spectrum of the *unperturbed* glass-like particle distribution is shown by the blue curve. For the wavenumbers of interest, the power spectrum of the glass is many orders of magnitude below the discreteness noise expected for a Poisson distribution of objects with the same space density as the dark matter particles, as shown by the dashed line. In this chapter, we do not apply any shot noise correction to power spectra measured for the dark matter, but we do make such a correction for spectra estimated for samples of haloes and galaxies.

To close this subsection, we turn our attention to the error on the measurement of the power spectrum. A commonly used expression for the fractional error in the measured power spectrum was derived by Feldman et al. (1994) (see also Efstathiou (1988), for a similar argument applied to the two point correlation function):

$$\frac{\sigma}{P} = \sqrt{\frac{2}{n_{\text{modes}}}} \left( 1 + \frac{1}{P\bar{n}} \right), \quad (2.3)$$

where  $n_{\text{modes}}$  is the number of Fourier modes present in a spherical shell of width  $\delta k$ , which depends upon the survey volume  $V$ : for  $k \gg 2\pi/V^{1/3}$ , this is given by  $n_{\text{modes}} = V4\pi k^2 \delta k / (2\pi)^3$ . The first term on the right hand side of Eq. 2.3 quantifies the sample variance in the measurement, which decreases as the square root of the number of modes or, equivalently, as the square root of the volume probed. The second term arises from the discreteness of the objects under consideration. The combination  $P\bar{n}$  quantifies the amplitude of the power spectrum in units of the Poisson shot noise, effectively giving the contrast of the power spectrum signal relative to the shot noise level. In the case where  $P\bar{n} \gg 1$ ,  $\sigma/P \propto 1/k$ . On the other hand, when the amplitude of the power spectrum is comparable to the shot noise, and if  $P(k) \propto k^{-1}$ , then the fractional error in the power is approximately independent of wavenumber. We have tested this prescription in both regimes against the diagonal element of the covariance between power spectrum measurements extracted from the ensemble of low resolution simulations, as shown in Fig. 2.3. Over the wavenumber range of interest, the agreement is reasonably good for

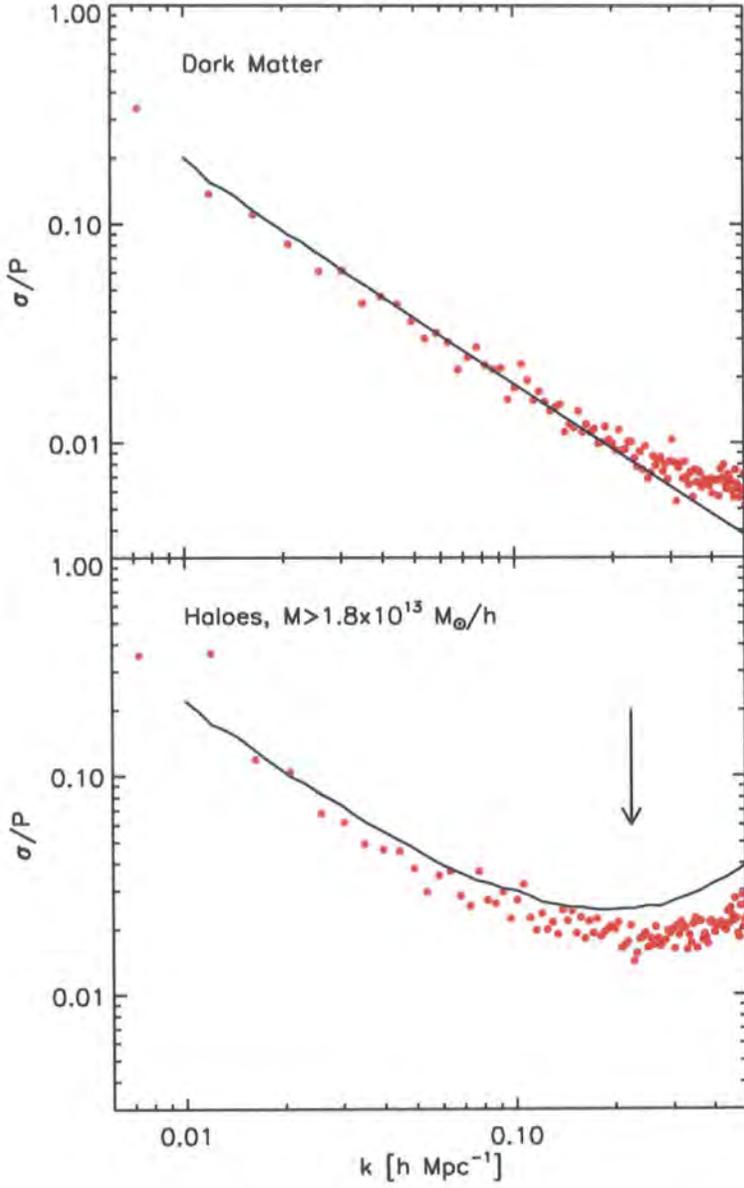


Figure 2.3: The fractional error in the power spectrum of the dark matter (top panel) and in the power spectrum of haloes more massive than  $1.8 \times 10^{13} h^{-1} M_{\odot}$  (bottom panel), estimated using the low resolution simulations from the dispersion of  $P(k)$  around the ensemble mean. The smooth black curves show the error predicted by the analytical expression given in Eq. 2.3. The red points show the scatter from the ensemble of low resolution simulations. The arrow in the bottom panel shows the wavenumber for which  $\bar{n}P(k = 0.2h\text{Mpc}^{-1}) = 1$ .

samples in which the shot noise is negligible compared to the clustering signal. For samples with low contrast power measurements, such as is the case for dark matter haloes used in the bottom panel of Fig. 2.3, the analytic expression works well until  $k \sim 0.1h\text{Mpc}^{-1}$  and then overpredicts the errors by up to 50%. We note that nonlinearities and the impact of the window function of a realistic survey could introduce off-diagonal terms in the power spectrum covariance matrix. In Section 2.5.3, we compare the constraints on the recovered oscillation scale using the scatter from the ensemble and using the simple mode-counting argument outlined above. We find good agreement which suggests that mode-coupling does not make a significant contributions to the errors on the scales relevant to the BAO.

### 2.2.2 Modelling the formation and evolution of galaxies

The N-body simulations described in the previous section follow the growth of fluctuations in the mass which is dominated by collisionless matter. To connect the predictions of the cold dark matter theory to forthcoming galaxy surveys, we need to predict which structures host galaxies and how galaxy properties depend on halo mass.

Some authors have chosen to incorporate galaxies into an N-body simulation empirically by using a parametric model called a halo occupation distribution function (HOD) to describe the probability distribution of galaxies expected in haloes of a given mass (Benson et al. 2000). The form of the HOD is constrained to reproduce a particular clustering measurement, such as the galaxy correlation function (e.g. Peacock and Smith 2000; Seljak 2000; Scoccimarro et al. 2001; Cooray and Sheth 2002). This approach has been applied to the study of the detectability of acoustic oscillations by several authors (Seo and Eisenstein 2005; Schulz and White 2006; Huff et al. 2007). Two assumptions are made when using the HOD to populate an N-body simulation with galaxies. Firstly, the parametrisation used for the HOD is assumed to provide an accurate description of the manner in which galaxies populate haloes across a wide range of halo mass. Detailed comparisons between the clustering predictions made using HODs and those obtained directly from simulations of galaxy formation show that in practice, the HODs do a reasonable job (Berlind et al. 2003; Zheng et al. 2005). Recently, one of the fundamental assumptions which underpins the HOD approach has been called into question. Using the Millennium simulations, Gao et al. (2005) demonstrated that the clustering of dark matter haloes depends on a second parameter, such as the formation time of the halo, in addition to halo mass (see also Harker et al. 2006 and Wechsler et al. 2006, Wetzel et al. 2007). In practice, for typical galaxy samples, this effect is largely washed out due to the

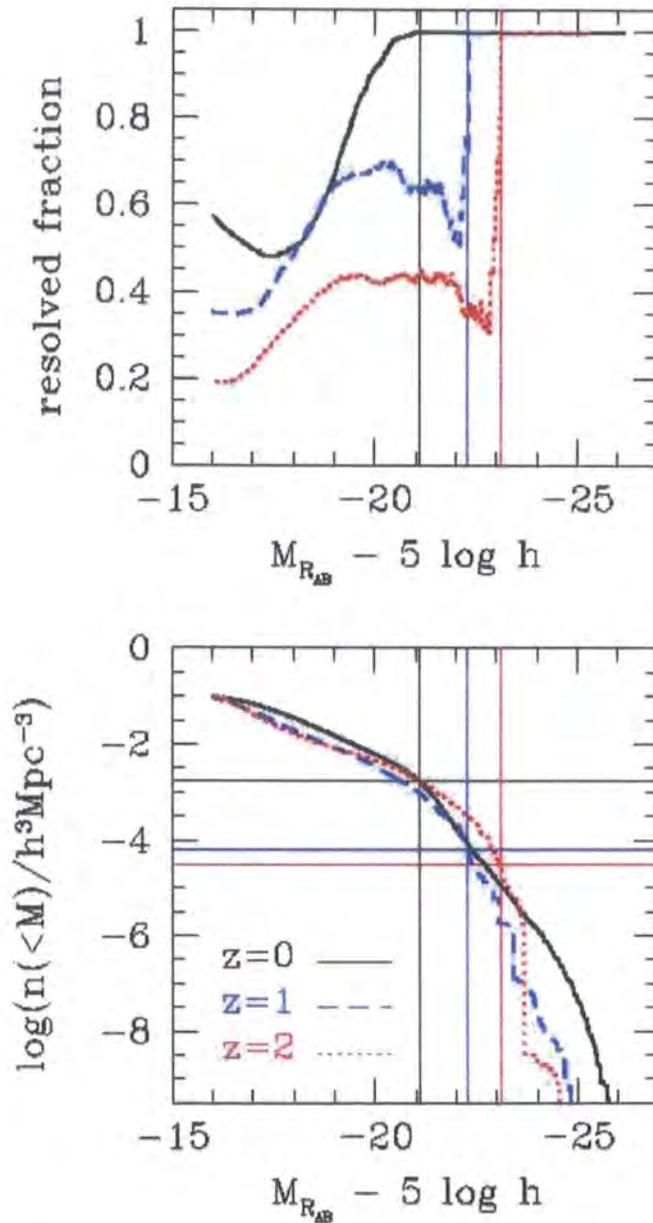


Figure 2.4: Upper panel: the fraction of ‘resolved galaxies’ in the high resolution N-body simulation as a function of magnitude, at different output redshifts (as given by the key in the lower panel). The magnitude is in the observer-frame  $R$ -band; to obtain an apparent  $R$ -band magnitude, the distance modulus corresponding to the redshift should be added to the plotted magnitude. The vertical lines mark the magnitude at which the galaxy sample is 100% complete at each redshift. Lower panel: the cumulative luminosity function of galaxies brighter than a given  $R$ -band magnitude, for different redshifts as given in the key. The vertical lines show the 100% completeness limits at each redshift and the horizontal lines indicate the associated space density of galaxies.

mix of halo properties sampled (Croton et al. 2007). The second implicit assumption in the HOD method when applied to an N-body simulation is that all of the haloes in which galaxies are expected to be found can be resolved in the simulation; if the mass resolution of the simulation turns out to be inadequate, then the HOD realized will be distorted to compensate, compared with the true, underlying HOD in the Universe.

In this chapter, we take a more physical approach and make an *ab initio* prediction of which dark matter haloes should contain galaxies by modelling the physics of the baryonic component of the Universe. We do this using a semi-analytic model of galaxy formation (for a review of this technique see Baugh 2006). The semi-analytic model describes the key physical processes which are thought to determine the formation and evolution of galaxies. We use the GALFORM code introduced by Cole et al. (2000) and developed in a series of papers (Benson et al. 2002, 2003; Baugh et al. 2005; Bower et al. 2006). The specific model we use is the one proposed by Baugh et al. (2005), which reproduces the abundance of Lyman-break galaxies at  $z = 3$  and  $z = 4$ , the number counts of sub-mm detected galaxies (with a median redshift  $z \sim 2$ ), and a rough match to the abundance of luminous red galaxies (Almeida et al., 2008), whilst at the same time giving a reasonable match to the observed properties of local galaxies (e.g. Nagashima et al. 2005a,b; Almeida et al. 2007).

A key advantage of using a semi-analytic model is that we can investigate how the manner in which galaxies are selected affects the accuracy with which the acoustic oscillations can be measured. The model predicts the star formation history of each galaxy and uses this to compute a spectrum, broadband magnitudes and emission line strengths (for examples of the latter, see Le Delliou et al. 2005, 2006). We can therefore select samples of model galaxies by applying precisely the same criteria which will be applied in the proposed surveys.

Our methodology mirrors the hybrid schemes introduced by Kauffmann et al. (1997) and Benson et al. (2000). We use a Monte Carlo technique to generate merger trees for dark matter haloes since our simulation outputs do not have the resolution in time or mass necessary to allow the construction of merger trees. (See Baugh 2006 for a discussion of the relative merits of these two approaches.)

We first construct a grid of halo masses at the redshift of interest, which extends to lower mass haloes than can be resolved in the simulation. We then generate a number of Monte-Carlo realizations of mass assembly histories for each mass on the grid, using the algorithm introduced by Cole et al. (2000). The number of realizations is chosen

to allow robust predictions to be made for observables such as the galaxy luminosity function. The halo merger history is input into the semi-analytic code and the properties of the galaxy population are output at the redshift for which the galaxy catalogue is to be constructed. In the calculations in this chapter, we output the broadband magnitudes in the  $R$ ,  $I$  and  $K$  bands and the equivalent widths of  $H_\alpha$  and OII[3727] for each galaxy. Finally, haloes from the grid are matched with haloes of similar mass identified in the N-body simulation. The central galaxy in each halo is assigned to the centre of mass of the matched halo in the simulation. The satellite galaxies are assigned randomly to dark matter particles in the halo. Galaxies placed in the simulation box in this way are called ‘resolved galaxies’. The Monte Carlo merger trees will not, of course, correspond in detail with those of the matched halos in the N-body simulation. However, to the extent that the halo assembly bias discussed by Gao et al. (2005) can be neglected, the properties of the trees are statistically similar for haloes in the same mass range.

Because of the finite mass resolution of the N-body simulation, galaxy samples generated by populating resolved haloes will be incomplete fainter than some magnitude limit. In principle, since we are using Monte-Carlo merger trees, we can follow galaxies down to arbitrarily faint magnitudes *within* a resolved dark matter halo. However, as we consider progressively fainter objects, some fraction of these galaxies should also appear in haloes which the simulation cannot resolve, causing the sample to become incomplete. Thus, in some instances we need to consider galaxies which we would expect to find in haloes below the mass resolution of the simulation. These galaxies are called “unresolved galaxies” and are placed in the box in the following way. A volume-limited sample of galaxies is generated using the semi-analytic model, with a volume equal to that of the simulation cube. Only galaxies which reside in haloes from the grid which are less massive than the resolution limit of the N-body simulation are considered. (Recall that the grid of halo masses used in the semi-analytic calculation extends to lower mass than those resolved in the simulation). These galaxies are assigned to randomly selected dark matter particles which have *not* been identified as members of halos identified by the friends-of-friends algorithm. This approach was adopted for one of the mock catalogues used in Cole et al. (2005). As we will see below, the unresolved galaxies are a minority within any of the samples we consider. They have little effect on the measured power spectrum, producing only a modest change in the amplitude of the clustering signal.

We can use the semi-analytic calculation carried out on the grid of halo masses to find the completeness limit of the galaxy catalogue in the N-body simulation. To do this, we

use the galaxy formation calculation carried out using the grid of halo masses to compute the cumulative luminosity function of galaxies, starting with the brightest galaxy, for two cases: 1) without any restriction on the mass of the halo which hosts the galaxy and 2) considering only those galaxies which reside in haloes above the resolution limit of the simulation. We then divide the second estimate of the cumulative luminosity function by the estimate made without any restriction on halo mass.

The completeness ratios calculated in this way are shown for  $z = 0, 1$  and  $2$  in Fig. 2.4. The vertical lines show the magnitude limit down to which the ‘resolved galaxy’ catalogues are 100% complete. The lower panel shows the cumulative luminosity function in the model at the same redshifts, with horizontal lines marking the space density of galaxies at the sample completeness limit. (The magnitudes plotted are observer-frame absolute magnitudes in the  $R$ -band. The apparent magnitude is obtained by adding the appropriate distance modulus for each redshift. All magnitudes are on the AB scale.) The  $z = 2$  sample is complete down to  $M_R - 5 \log h = -23$ , or, equivalently to a space density of  $3.2 \times 10^{-5} h^3 \text{Mpc}^{-3}$ . Faintwards of this magnitude, the completeness drops sharply to around 30 – 40%. The situation is much more encouraging at  $z = 1$ . Here, the galaxy catalogue is complete to  $M_R - 5 \log h = -22.3$  (corresponding to a space density of just under  $10^{-4} h^3 \text{Mpc}^{-3}$ ) and faintwards of this there is a much more modest drop in the fraction of galaxies resolved in the simulation. The simulation resolves around two thirds of the space density of galaxies expected in the proposed WFMOS survey. At  $z = 0$ , the galaxy samples are complete to a much higher space density, in excess of  $10^{-3} h^3 \text{Mpc}^{-3}$ .

### 2.3 The power spectrum of galaxy clustering

In this section we examine the various phenomena which are responsible for changing the form of the power spectrum of galaxy clustering from that expected in linear perturbation theory. We systematically add in new effects and elements of sample selection, considering first the power spectrum of the dark matter, looking at nonlinear evolution (§2.3.1) and the impact of peculiar velocities (§2.3.2), before moving onto dark matter haloes (§2.3.3) and finally to synthetic galaxy samples (§2.3.4).

For completeness, we first explain some of the terminology we use in this section. There are three types of phenomena responsible for distorting the linear theory power spectrum: i) non-linear growth of fluctuations, ii) redshift-space distortions and iii) bias. Non-linear growth refers to the coupled evolution of density fluctuations on different scales.

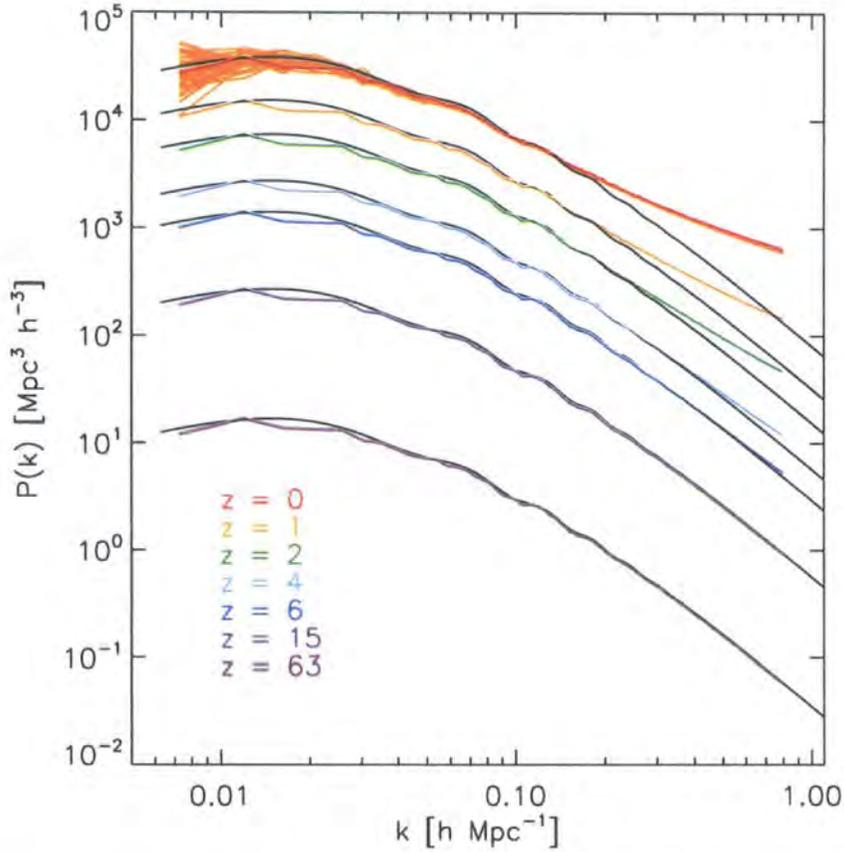


Figure 2.5: The growth of the power spectrum of density fluctuations in the dark matter, as measured in real space. The smooth curves show the predictions of linear perturbation theory at the redshifts indicated by the key. The power spectra measured in the low resolution ensemble at  $z = 0$  are plotted to show the sampling variance for a simulation box of side  $1340 h^{-1} \text{ Mpc}$ . The smallest wavenumber plotted corresponds to the fundamental mode in the simulation,  $2\pi/L = 0.0469 h^{-1} \text{ Mpc}$ . The maximum wavenumber shown is 0.67 times the Nyquist frequency of the FFT grid, chosen to avoid any aliasing effects.

Redshift-space distortions describe the impact of gravitationally induced peculiar motions on the clustering pattern. We will refer to clustering measurements as being made in “real space” or “redshift space”; in the latter case peculiar motions are taken into account, as we describe in §2.3.2. The term “bias” has a range of meanings in the literature. Bias is used to describe the boost in the clustering of a particular tracer (e.g. galaxies or clusters) relative to a reference point, which could be the clustering of the dark matter in either linear perturbation theory or taking into account nonlinear evolution. One of the earliest uses of the concept of bias was in the application of the high peaks model to explain the enhanced clustering of Abell clusters (Kaiser 1984). In this model, clusters are associated with rare peaks in the initial, Gaussian density field. The bias is defined as the square root of the ratio of the two-point correlation function of peaks of a certain minimum height to the clustering of the mass expected in linear perturbation theory. When considering galaxies, it is perhaps more natural to think in terms of a modulation of clustering relative to that displayed by the underlying mass at the same epoch, since galaxies populate dark matter haloes. In this case, the galaxy clustering will be measured relative to that of the evolved matter distribution. On large scales, these two reference points, the clustering of the matter expected in linear perturbation theory or the evolved clustering, should be essentially the same. We shall see later that this is approximately the case for the scales over which we compare clustering signals to measure bias factors.

### 2.3.1 The nonlinear growth of matter fluctuations

The early stages of the growth of a density fluctuation are particularly simple to describe analytically. The fluid equations can be written in terms of the perturbation to the density and Fourier transformed. In the simplest case, when the density contrast  $\delta \ll 1$ , the Fourier modes evolve independently of one another. This is called linear growth. In this regime, the power spectrum changes in amplitude with time, but not in shape. The shift in amplitude is described by the growth factor  $D$ , which is a function of the densities of matter and dark energy (as quantified by the present day density parameters,  $\Omega_M$  and  $\Omega_\Lambda$ , for matter and dark energy respectively) and redshift (see Heath 1977; Peebles 1980):

$$P(k, z) = D^2(z, \Omega_M, \Omega_\Lambda) P(k, z = 0), \quad (2.4)$$

where  $D(z = 0) = 1$ .

We plot the power spectrum of the dark matter in real space measured from the

BASICC at different output redshifts in Fig. 2.5. The approximately linear growth of the power spectrum is readily apparent on large scales (low  $k$ ). In an Einstein - de Sitter universe ( $\Omega_M = 1$ ), the growth factor is equal to the expansion factor. If dark energy plays a role in setting the rate at which the Universe expands, the growth of fluctuations is suppressed relative to the Einstein - de Sitter case at late times. The BASICC started at  $z_s = 63$ , so if  $\Omega_M = 1$ , we would expect to see the power spectrum grow in amplitude by a factor of  $(1 + z_s)^2 = 4096$  by  $z = 0$ . Using the approximate formula provided by Carroll et al. (1992), we expect a suppression in the growth of the power by a factor of 0.5537 for the cosmological parameters used in the simulation. This gives an overall growth in power from the initial conditions to the present of a factor of 2268. This agrees to within 0.6% with the factor expected from a direct numerical integration of the equation giving the growth factor (Eqs. 28 and 9 from Carroll et al. 1992), which gives 2281.01. In the simulation, we find that the power in the fundamental mode grows by a factor of 2285.21 from the initial conditions at  $z = 63$  to  $z = 0$ , which agrees with the growth predicted by linear perturbation theory to 0.02%.

Fig. 2.5 shows that the growth of the power spectrum is clearly not linear at high wavenumbers. The shape of the spectrum at high  $k$  at late times is different from that at high redshift, because the growth of modes of different  $k$  becomes coupled. This behaviour can be followed to some extent using second- and higher-order perturbation theory (Peebles 1980; Baugh and Efstathiou 1994; Jain and Bertschinger 1994; Crocce and Scoccimarro 2006a). However, as the density contrast approaches unity, second-order perturbation theory breaks down (Baugh and Efstathiou 1994). The coupled evolution of the Fourier modes starts on surprisingly large scales, which demonstrates the necessity of a large volume simulation to accurately follow the development of the power spectrum (Smith et al. 2007).

This can be seen more clearly if we divide the measured spectrum by the growth expected according to linear perturbation theory, as is done approximately in Fig. 2.6. In this plot, we have divided the power spectra measured from the simulation by the spectrum measured at  $z = 15$ , scaled by the square of the appropriate growth factor. This reduces the noise in the ratio arising from the finite number of modes realized at small wavenumbers in the simulation volume (Baugh and Efstathiou 1994; Springel et al. 2005). Any deviation away from unity signifies a departure from linear perturbation theory due to coupling between modes. The ratio shows a characteristic dip at low  $k$ , i.e. less power than expected in linear theory, before showing a strong enhancement at

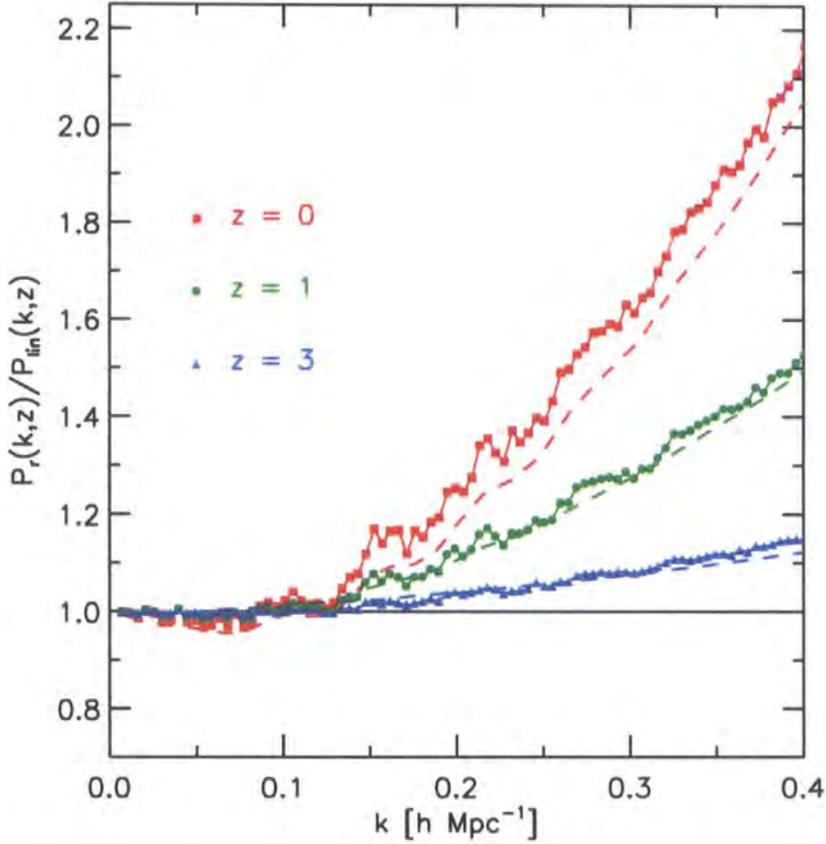


Figure 2.6: The nonlinear growth of the power spectrum. Here we divide the power spectrum in real space measured at the redshift indicated by the key by the power spectrum at  $z = 15$ , after taking into account the change in the growth factor. Any deviation of the resulting ratio from unity indicates a departure from linear perturbation theory. The dashed lines show the same ratio as predicted using the ansatz of Smith et al. (2003).

higher wavenumbers (Baugh and Efstathiou 1994). It is remarkable that the transition between a deficit and excess of power happens at the same wavenumber,  $k \sim 0.1 h \text{Mpc}^{-1}$ , at different epochs. The suppression in power at low  $k$ , on the order of a 3%, is not as strong as that seen in an Einstein - de Sitter universe (see figure 4 of Baugh and Efstathiou 1994). Nevertheless, this drives the spectacular boost in power seen at higher wavenumbers. The dip in power is largest around  $k \sim 0.05 h \text{Mpc}^{-1}$ , which corresponds to a length scale of  $2\pi/k \sim 125 h^{-1} \text{Mpc}$ , close to the wavelength of the acoustic oscillations. Several authors have proposed ansatzes which transform the linear perturbation theory power spectrum into the non-linear power spectrum (e.g. Hamilton et al. 1991; Peacock and Dodds 1994, 1996; Smith et al. 2003). We plot the predictions of the model proposed by Smith et al. (2003) in Fig. 2.6 using dashed lines. The ratio is computed by dividing the power spectrum at the epoch of interest by the suitably scaled prediction of the model for  $z = 15$ . The agreement is excellent at high redshift. At  $z = 0$ , at higher wavenumbers, the Smith et al. (2003) formula recovers the simulation results to within 5% over the range plotted.

### 2.3.2 The impact of redshift-space distortions on the power spectrum

In a spectroscopic galaxy survey, the radial distance to an object is inferred from its measured redshift. The shift in the spectral features of the galaxy is produced by two contributions to its the apparent velocity: the expansion of the Universe, which is responsible for the Hubble flow at the true distance to the galaxy, and local inhomogeneities in the gravitational field around the object, which generate an additional, “peculiar” velocity. Since we cannot correct a priori for the effects of the local gravitational field when inferring the radial distance from the Hubble law and the measured redshift, an error is made in the distance determination. The impact of such errors on the form of the measured power spectrum of clustering is called the redshift-space distortion.

Peculiar motions display two extremes which produce different types of distortion to the power spectrum: i) On large scales, coherent bulk flows out of voids and into overdense regions lead to an enhancement in the density inferred in redshift space, and hence to a boost in the recovered power. Kaiser (1987) derived a formula for the enhancement of the spherically averaged power, under the assumption of linear perturbation theory for an observer situated at infinity (the plane parallel approximation):

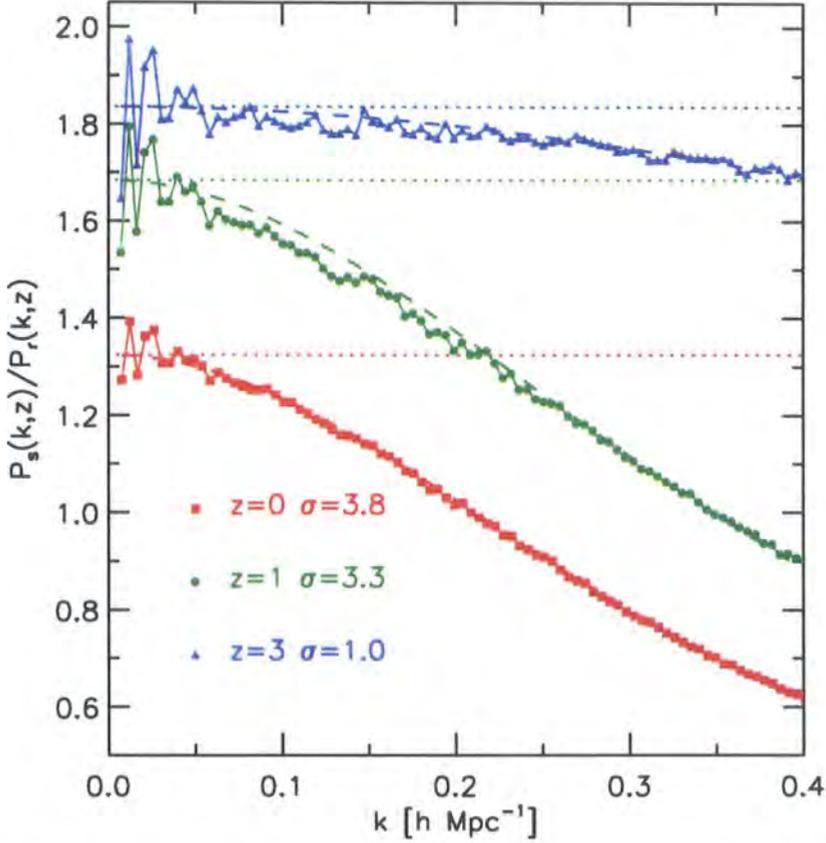


Figure 2.7: The ratio of the power spectrum measured for the dark matter in redshift space, i.e. including the impact of peculiar motions in the distance determination, to the power spectrum measured in real space. The deviation from unity shows the redshift-space distortion to the nonlinear power spectrum. The results are shown for selected output redshifts, as indicated by the key. The horizontal dotted lines indicate the boost in the redshift-space power expected due to coherent flows, as predicted by Eq. 2.5. The dashed lines show a simple fit to the distortions (see Eq. 2.6).

$$f = \frac{P_s(k)}{P_r(k)} = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right), \quad (2.5)$$

where  $P_s(k)$  is the power spectrum in redshift space,  $P_r(k)$  is the spectrum in real space and  $\beta = (d \log \delta / d \log a) / b \simeq \Omega_M^{0.6}(z) / b$ , where  $b$  is the bias factor ( $b = 1$  for the dark matter; for a discussion of the dependence of the growth factor on  $\Omega_M$ , see Linder 2005; Linder and Cahn 2007). ii) On small scales, the random motions of objects inside virialized dark matter haloes cause structures to appear elongated when viewed in redshift space, leading to a damping of the power. Peacock and Dodds (1994) discussed a model for the redshift-space power spectrum, which takes into account both limits of peculiar motions (see also Scoccimarro 2004).

Fig. 2.7 shows the ratio of the power spectrum measured for the dark matter in redshift space to that measured in real space, at redshifts  $z = 3, 1$  and  $0$ . The dotted lines indicate the boost expected in the redshift-space power, computed using the expression in Eq. 2.5 (Kaiser 1987). This factor changes with redshift because the matter density parameter is changing. Fig. 2.7 shows that this behaviour is only approached asymptotically, on scales in excess of  $100 h^{-1}$  Mpc. At higher wavenumbers, the power measured in redshift space is suppressed by random motions. The dashed lines in this plot show a simple fit to this ratio

$$f = \frac{P_s(k)}{P_r(k)} = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right)(1 + k^2\sigma^2)^{-1}, \quad (2.6)$$

where  $\sigma$  is a free parameter, which is loosely connected to the pairwise velocity dispersion. The degree of damping grows between  $z = 3$  and  $z = 1$ , but changes relatively little by  $z = 0$ . We shall see in later sections that the form of the redshift-space distortion to the power spectrum depends on the type of object under consideration.

### 2.3.3 The power spectrum of dark matter halos in real and redshift space

In modern theories of galaxy formation, dark matter haloes play host to galaxies. It is therefore instructive to compare the power spectra measured for different samples of haloes to that of the dark matter as a step towards understanding the power spectrum of galaxies.

A common conception is that the clustering of haloes is a scaled version of the clustering of the underlying mass, with the shift in clustering amplitude quantified in terms

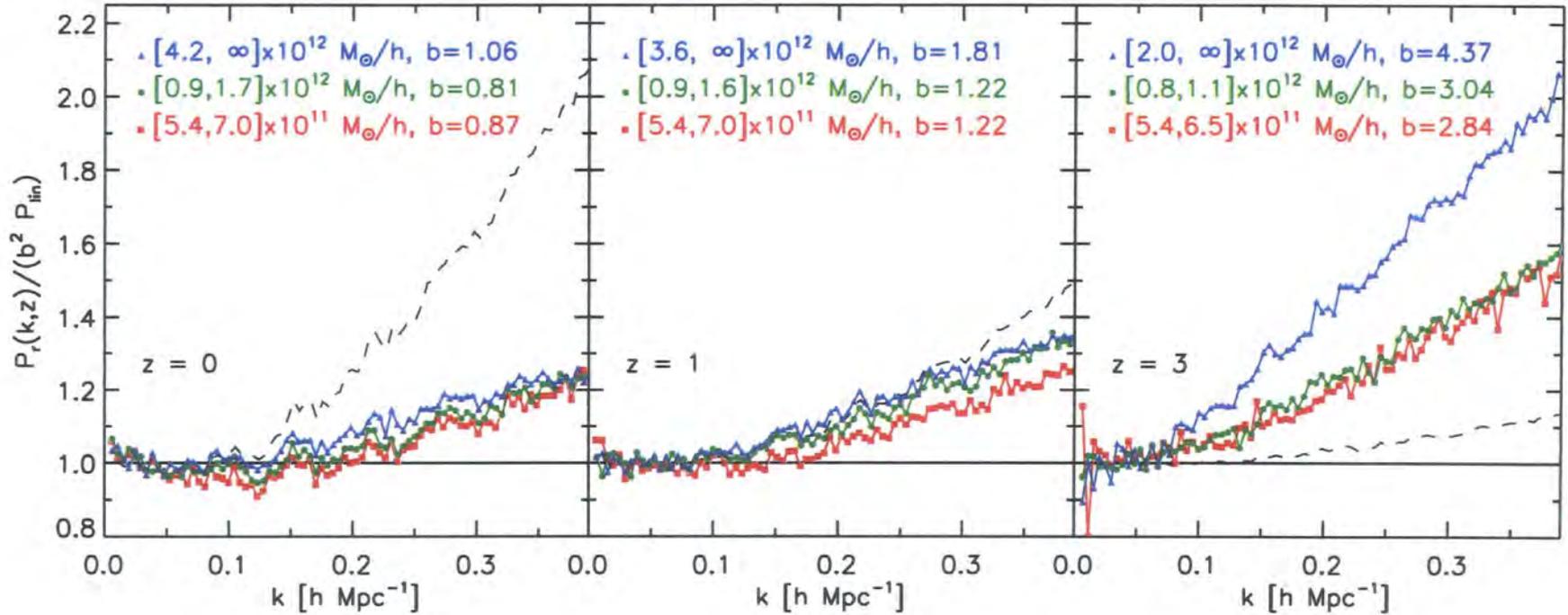


Figure 2.8: The power spectrum of dark matter haloes measured in real space compared to a scaled version of the prediction of linear perturbation theory, which takes into account the growth factor and an effective bias computed on large scales  $k < 0.1 h \text{Mpc}^{-1}$ . Each panel corresponds to a different output redshift. Different mass samples are considered, as indicated by the key, which correspond to low, average and high masses, defined in terms of the average halo mass present at each output time. The black dashed line shows the real-space power spectrum of the mass divided by the appropriate linear perturbation theory prediction.

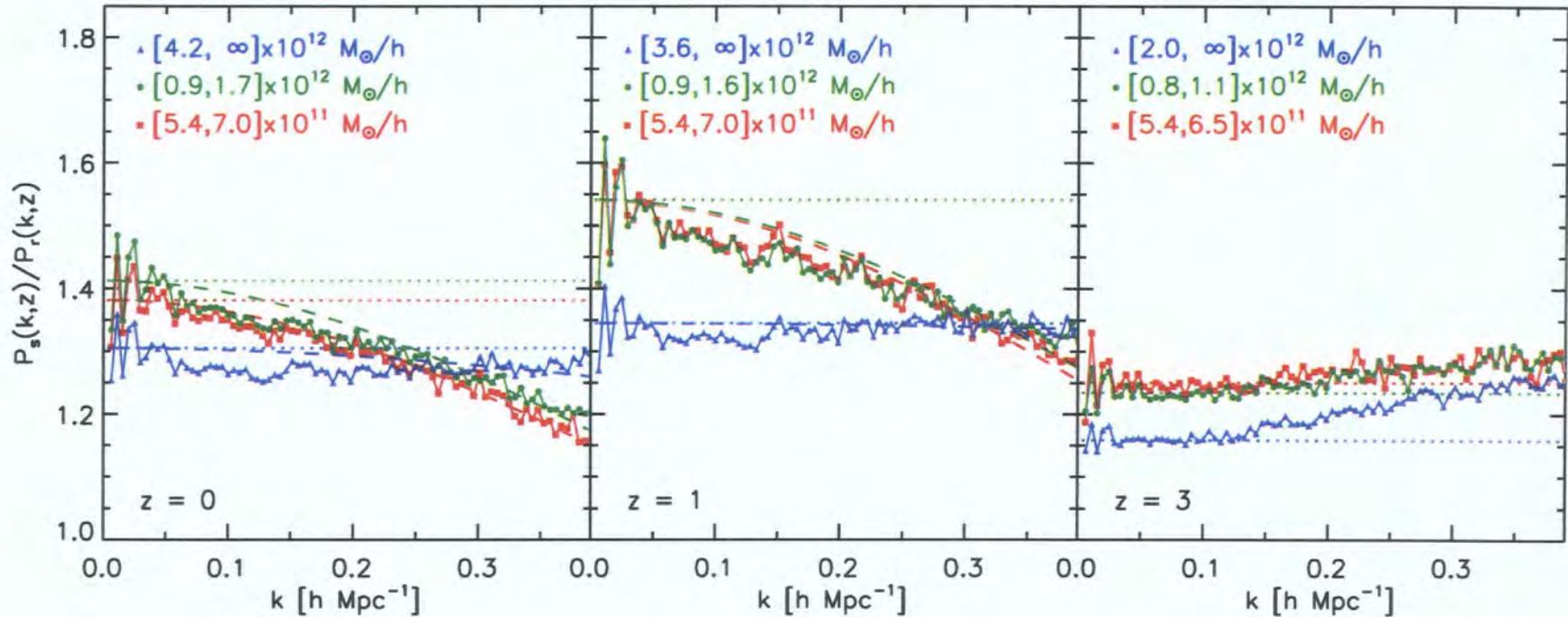


Figure 2.9: The power spectrum of dark matter haloes measured in redshift space divided by the power spectrum measured in real space for the same sample. Each panel corresponds to a different output redshift. Different mass samples are considered, as indicated by the key, which correspond to low, average and high masses, defined in terms of the average halo mass present at each output time. The horizontal dotted lines show the expected ratio for the boost in the amplitude of the redshift-space power spectrum due to coherent flows, computed using an effective bias factor estimated on large scales. The dashed lines show the best fit model of Eq. 2.6, which turns out to be a poor description of the redshift-space distortions. No suitable fits were obtained at  $z = 3$ .

of a bias factor,  $b$ , where  $b^2 = P_{\text{halos}}/P_{\text{dm}}$  (Cole and Kaiser 1989; Mo and White 1996). As we commented earlier, since we use the dark matter power spectrum on large scales to define a bias, this is approximately the same as using the linear perturbation theory spectrum. Many authors have tested analytical prescriptions for computing the bias parameter using extensions of the theory of Press and Schechter (1974) (e.g. Mo et al. 1997; Sheth et al. 2001; Jing 1998; Governato et al. 1999; Colberg and others 2000; Seljak and Warren 2004). In the extended Press-Schechter theory, the bias is only a function of halo mass and redshift. However, recent analyses of high resolution, large volume simulations have revealed some dependence of halo clustering on a second parameter besides mass, such as the halo’s formation redshift or concentration parameter (Gao et al. 2005; Harker et al. 2006; Wechsler et al. 2006).

In Fig. 2.8, we show that this simple picture, in which the clustering of haloes is a shifted version of that of the dark matter, is actually a poor approximation to what we find in the simulation. We show the ratio of the power spectrum of a sample of dark matter haloes measured in real space to a scaled version of the linear perturbation theory power spectrum. The amplitude of the linear theory spectrum used in the ratio takes into account the growth factor appropriate to the output redshift and an effective bias, which is set by matching the linear theory prediction for the mass spectrum to the measured halo spectrum on large scales, i.e. for wavenumbers in the range  $0.0046 < (k/h\text{Mpc}^{-1}) < 0.1$ . Each panel in Fig. 2.8 corresponds to a different output redshift from the simulation. For each redshift, we have defined three samples of dark matter haloes, which contain the same number of objects. The mass intervals are set relative to the average halo mass present in the respective outputs, with “low”, “mean” and “high” mass samples considered. Each of these contains 20% of the total number of haloes present at each epoch, with the mass ranges used at each redshift indicated on the keys. The effective bias factors of the halo samples are also written in the key. For comparison, the dashed line in each panel shows the corresponding ratio for the dark matter.

Fig. 2.8 shows that at  $z = 3$ , all of the haloes considered have effective biases much greater than unity, indicating they are more strongly clustered than the mass. This situation is reversed at  $z = 0$ . At this epoch, the halo mass resolution of the BASICC is smaller than the corresponding value of  $M_*^1$  ( $= 5.78 \times 10^{12} h^{-1} M_{\odot}$  at  $z = 0$ ). The  $z = 0$

---

<sup>1</sup> $M_*$  is a characteristic mass scale defined as the mass within a sphere for which the *rms* variance in linear perturbation theory is  $\sigma(M) = \delta_{\text{crit}}(z)$ , where  $\delta_{\text{crit}}$  is the extrapolated critical linear overdensity given by the spherical collapse model at redshift  $z$ .

samples have a bias of unity or smaller. In addition to the difference in the effective bias parameters, the shape of the spectrum of the haloes in these extremes is also different (see also Smith et al. 2007). The plot shows the shape of the power spectrum, after accounting for the effective bias on large scales. Any difference between the curves plotted for the haloes and that for the dark matter (dashed line) shows a difference in the clustering signal over and above that quantified by a constant effective bias. Similar behaviour was found for samples of cluster mass haloes in the Hubble Volume simulation by Angulo et al. (2005).

We now consider the clustering of haloes as viewed in redshift space, taking the centre of mass velocity of the halo as its peculiar velocity. In Fig. 2.9, we plot the ratio of the redshift-space power spectrum for the halo samples used in Fig. 2.8 to the power spectrum measured in real space. As we did before for the case of the dark matter (Fig. 2.7), we indicate the boost in power expected on large scales (small  $k$ ) due to coherent bulk flows of haloes. The boost is calculated from Eq. 2.5 using the effective bias of the halo sample. The plot shows that the redshift-space power spectrum at low wavenumbers is in reasonable agreement with this simple model. However, a range of behaviour is seen at higher wavenumbers. For haloes comparable to  $M_*$ , the boost in power in redshift space is less than predicted by Eq. 2.5. For the more extreme, massive haloes, there is actually more power in redshift space than is suggested by Kaiser’s formula. This “excess” power was previously noted by Padilla and Baugh (2002) and Angulo et al. (2005). The Kaiser formula assumes linear perturbation theory and breaks down in the case of objects with strongly nonlinear clustering. In the case of the less extreme haloes, the reduction in power is *not* due to virialized motions of haloes within larger structures. The halo finder we have used is designed to return an overdensity corresponding to virialized structures and not substructures. If the haloes were really part of a larger structure and were executing random motions, the group finder would simply have lumped them together as one larger structure. We are perhaps seeing instead haloes that have started to merge with one another, and whose motions have broken away from a coherent large scale flow. We know of no analytical description of the redshift-space clustering of dark matter haloes which explains this behaviour.

### 2.3.4 The power spectrum of galaxies

The galaxy power spectrum can be very different from the power spectrum of a sample of dark matter haloes. The way in which the galaxies are distributed among haloes changes

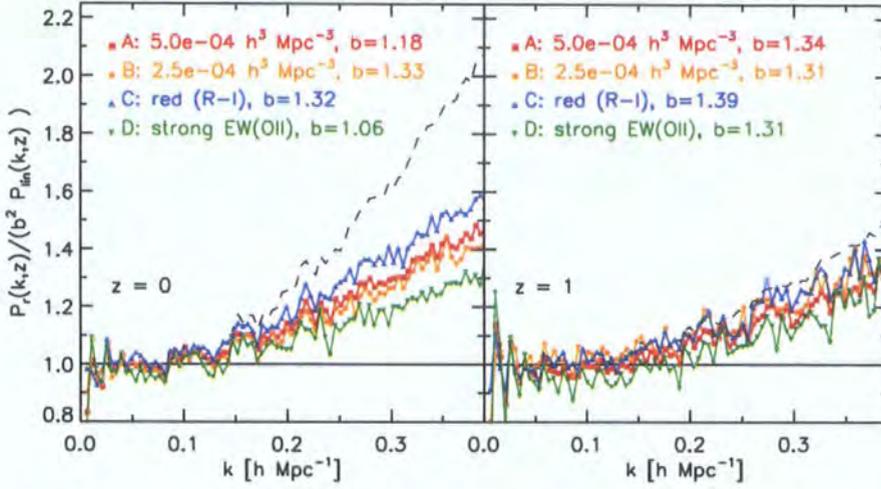


Figure 2.10: The power spectrum of different galaxy samples measured in real space, divided by the square of an effective bias parameter and the appropriately scaled linear perturbation theory power spectrum. The sample definition and the value of the effective bias used are given by the key. The power spectrum of the dark matter spectrum in real space, also divided by the linear perturbation theory spectrum, is shown by the black dashed line. The left hand panel shows the ratios at  $z = 0$  and the right hand panel at  $z = 1$ .

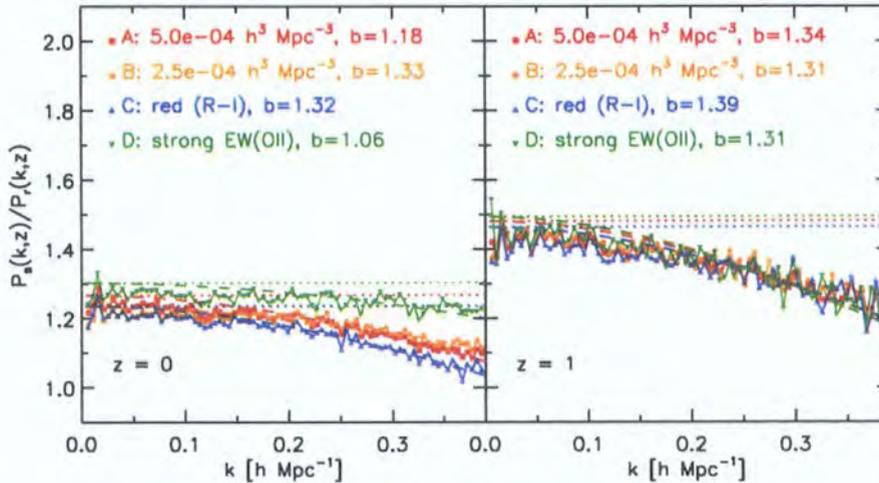


Figure 2.11: The ratio of the power spectrum of galaxies measured in redshift space to that in real space, at  $z = 0$  (left) and  $z = 1$  (right). The samples are defined by the key in each panel. The dotted horizontal lines show the predictions of Eq. 2.5 for the various samples.

the form of the power spectrum. In a mass-limited sample of haloes, the contribution of each halo to the power spectrum can be determined through its space density, which acts as a weighting factor when computing the contribution of the halo to the clustering signal. The number of galaxies per halo acts to modify this weight e.g. more massive haloes could contain more galaxies than less massive haloes. Furthermore, the presence of satellite galaxies within a halo means that one expects to see a damping in power on small scales in redshift space, due to the random motions of the satellites within the virialized dark halo. The precise modification of the power spectrum depends in detail on how galaxies populate dark matter haloes. As we discussed in §2.2.2, we have carried out an *ab initio* calculation of the number of galaxies per halo, using a semi-analytic model of galaxy formation. We are able to predict observable properties of galaxies, such as broadband magnitudes and the strength of emission lines. We consider a range of galaxy samples, defined either by a magnitude limit alone (set in the R-band) or by combining an R-band magnitude limit with a colour selection (in R-I) or a cut on the strength of the OII[3727] emission line:

- Sample A: magnitude-limited to reach a space density of  $5 \times 10^{-4} h^3 \text{ Mpc}^{-3}$ .
- Sample B: magnitude-limited to reach half the space density of sample A, i.e.  $2.5 \times 10^{-4} h^3 \text{ Mpc}^{-3}$ .
- Sample C. The reddest 50% of galaxies from sample A, using the  $R - I$  colour.
- Sample D. The 50% of galaxies from sample A with the strongest emission lines, using the equivalent width of OII[3727].
- Sample E. The bluest 50% of galaxies from sample A, using the  $R - I$  colour.
- Sample F. The 50% of galaxies from sample A with the weakest emission lines, using the equivalent width of OII[3727].

The power spectra measured in real space from the various galaxy samples are plotted in Fig. 2.10. The spectra have been divided by the linear perturbation theory power spectrum multiplied by the square of an effective bias factor, which was estimated by comparing the galaxy spectra to the power spectrum measured for the dark matter for wavenumbers  $k < 0.1h\text{Mpc}^{-1}$ . In all cases, for the space densities we have chosen, the effective bias factors estimated for the samples are modest. For comparison, the ratio of the power spectrum of the dark matter in real space to the linear theory prediction

is also plotted, using a dashed line. The deviation of the dashed line from unity shows where nonlinear effects are important for the dark matter. Any differences between the plotted ratios for galaxies and mass indicate a scale dependent bias. The comparison between the dashed and solid curves in Fig 2.10 shows that a constant bias is only a good approximation on large scales,  $k < 0.15h\text{Mpc}^{-1}$ .

The redshift-space distortion in the galaxy power spectrum is shown in Fig 2.11, where we plot the ratio of the redshift-space spectrum to the real-space spectrum for the galaxy samples shown in Fig. 2.10. The horizontal lines show the Kaiser boost (Eq. 2.5) expected for the effective bias of the galaxy sample. This ratio is only attained on the very largest scales and seems to be an overestimate of the size of the effect at  $z = 1$ . The damping of the power on intermediate and small scales is readily apparent and, unlike the case with dark matter haloes, is well described by the form given in Eq. 2.6.

## 2.4 Constraining the Dark Energy Equation of state

In this section we outline the procedures we follow to place constraints on the dark energy equation of state parameter,  $w$ , by measuring the length scale imprinted by baryonic acoustic oscillations on the power spectrum of the various tracers of the density field. The transformation of a measurement of a distance scale into a constraint on  $w$  requires various approximations to be made, and depends upon the survey in question and upon the time variation assumed for the dark energy. Nevertheless it is instructive to go through this exercise, bearing these caveats in mind, to get a feel for how well future experiments will be able to measure  $w$  for the case of a constant equation of state.

The form of the power spectrum of density fluctuations contains information about basic cosmological parameters, and measurements of the galaxy power spectrum on large scales have been exploited to extract the values of these parameters (e.g. Cole et al. 2005; Sánchez et al. 2006; Tegmark et al. 2006; Padmanabhan et al. 2007; Percival et al. 2007). The apparent scale of features in the power spectrum offers another route to constrain selected cosmological parameters through the dependence of the distances parallel and perpendicular to the line of sight on the matter density parameter,  $\Omega_M$ , the dark energy density parameter,  $\Omega_{DE}$ , the dark energy equation of state parameter,  $w$  and the Hubble constant. For such an approach to work, we either need to know the true physical scale of a particular feature in the power spectrum beforehand or to compare the relative size of a feature when measured parallel and perpendicular to the line of sight (Alcock and

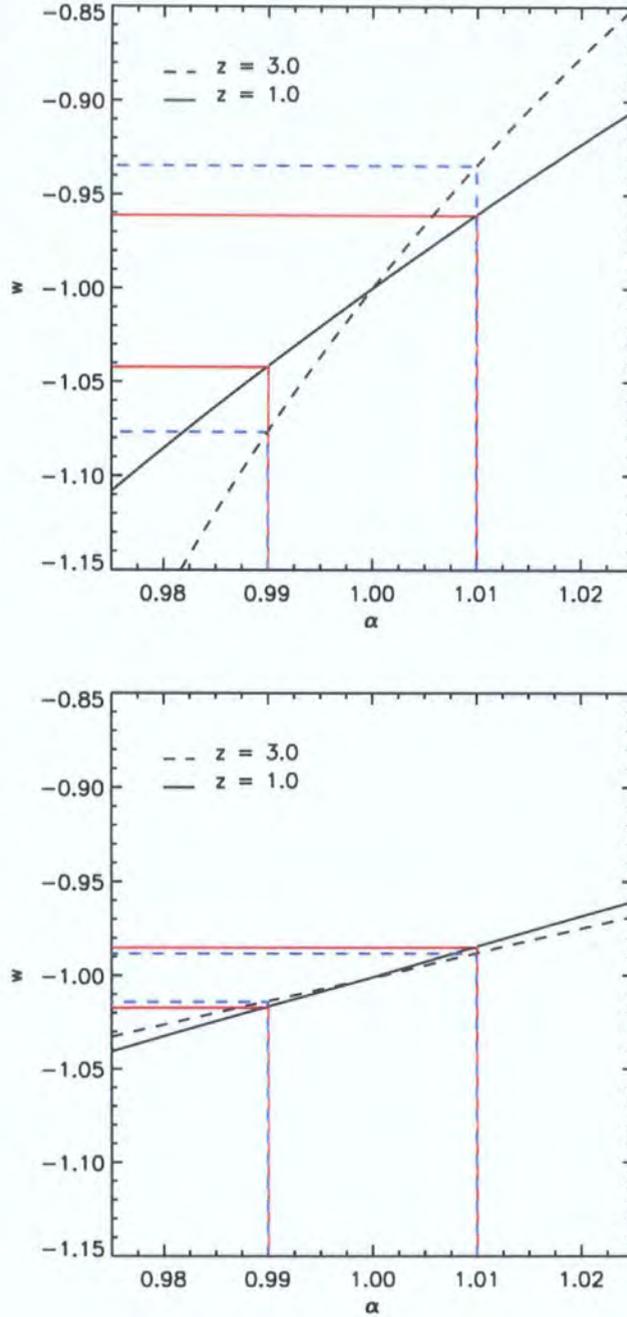


Figure 2.12: The relation between the dark energy equation of state parameter,  $w$ , and the scale factor,  $\alpha$ , defined by Eq. 2.8, for perturbations in the equation of state around  $w_{\text{true}} = -1$ . Two cases are shown. In the upper panel, the values of the other cosmological parameters are kept fixed. In the lower panel, the ratio of the sound horizon scale to the angular diameter distance to the last scattering surface is held fixed. The relation between  $\alpha$  and  $w$  is shown for  $z = 1$  (solid lines) and  $z = 3$  (dashed lines). The horizontal and vertical lines guide the eye to show how a 1% error in  $\alpha$  translates into an error in  $w$ .

Paczynski 1979). The baryonic oscillations present a promising candidate for such a feature. If we assume for the sake of argument that the cosmological parameters, apart from the equation of state of the dark energy, are well constrained, then the scale of the acoustic oscillations becomes a standard ruler. These features are expected on smaller scales than the turnover and have already been seen in current surveys at low redshift, although at too low a signal-to-noise ratio to use in isolation to extract a competitive constraint on the dark energy equation of state (Cole et al. 2005; Eisenstein et al. 2005).

We can see how the value of the equation of state parameter of the dark energy influences the form of the BAO with the following simple argument. To measure the power spectrum of galaxy clustering, we need to convert the angular positions and redshifts of the galaxies into comoving spatial separations. This requires a choice to be made for values of the cosmological parameters, including  $w$ . In our case, we set the parameters equal to the values used in the N-body simulations, with  $w = w_{\text{true}} = -1$  for the particular case we have run. The effect of a change in the value of  $w$ ,  $w_{\text{assumed}} = w_{\text{true}} + \delta w$  is to change the separations between pairs of galaxies, which leads to a change in the appearance of the power spectrum. For small perturbations away from the true equation of state, we assume that the alteration in the measured power spectrum can be represented by a rescaling of the wavenumber from  $k_{\text{true}}$  to  $k_{\text{app}}$ . The ratio of these wavenumbers gives a “stretch” parameter,  $\alpha$ , which describes the change in the recovered oscillation scale:

$$\alpha = \frac{k_{\text{app}}}{k_{\text{true}}}. \quad (2.7)$$

If  $w_{\text{assumed}} = w_{\text{true}}$ , then there is no shift in the BAO in the estimated power spectrum and  $\alpha = 1$ . In the case of a wide-angle, deep galaxy survey with spectroscopic redshifts, the stretch parameter can be approximated by:

$$\alpha \approx \left( \frac{D_A(z, w_{\text{assumed}})}{D_A(z, w_{\text{true}})} \right)^{-2/3} \left( \frac{H(z, w_{\text{true}})}{H(z, w_{\text{assumed}})} \right)^{1/3}, \quad (2.8)$$

where

$$H(z, w) = H_0 \left[ \Omega_m (1+z)^3 + \Omega_{\text{DE}} (1+z)^{3(1+w)} \right]^{1/2} \quad (2.9)$$

$$D_A(z, w) = \frac{c}{1+z} \int_0^z \frac{dz}{H(z)}. \quad (2.10)$$

The values of the exponents in Eq. 2.8,  $2/3$  for the distance transverse to the line of sight and  $1/3$  for the distance parallel to the line of sight are motivated by the number of

Cartesian components in these directions (e.g. Eisenstein et al. 2005). The precise value of these exponents will depend upon the geometry and construction of the galaxy survey. For example, in a survey which relies upon photometric redshifts, the exponent parallel to the line of sight would be greatly reduced and it would be beneficial to compute the power spectrum transverse to the line of sight. Note that in Eqs. 2.9 and 2.10 we assume that  $w$  is independent of redshift. There are many models in which  $w$  is a function of redshift. In this case, the exponent of  $\Omega_{\text{DE}}$  in the expression for the Hubble parameter (Eq. 2.9) would be replaced by an integral over  $w(z)$ .

It is instructive to see how the constraints on  $\alpha$  translate into limits on the value of  $w$ . We can do this approximately using Eq. 2.8, for the case of a redshift independent equation of state, considering perturbations around  $w_{\text{true}} = -1$ . We consider two illustrative cases: a “pessimistic” case in which we consider the constraints from BAO in isolation from any other data which constrains the cosmological parameters and an “optimistic” case, in which we perturb  $w$  and only consider cosmological models that give similar predictions for the CMB. The translation in the pessimistic case is shown in the upper panel of Fig. 2.12 for two different redshifts. Here we have assumed fixed values for  $\Omega_{\text{M}}$  and  $\Omega_{\Lambda}$  and we have not marginalized over these parameters. This is the case discussed most commonly in the literature. Under these conditions, at  $z = 1$ , a 1% error in  $\alpha$  corresponds approximately to a 4% error in the value of  $w$ . At  $z = 3$ , the boost is about 50% larger, with  $\delta w \approx 6\delta\alpha$ .

In the “optimistic” case, we only consider models which give the same angular location for the first peak in the CMB spectrum. Hence, when the value of  $w$  is perturbed, we restrict our attention to those models which give the same ratio of the sound horizon scale to the angular diameter distance to the last scattering surface as our default cosmology. Given the parametric forms quoted for these distances by Eisenstein and Hu (1998)), this is equivalent to keeping  $\Omega_{\text{b}}/\Omega_{\text{M}}$  and  $h$  fixed, and varying  $\Omega_{\text{M}}$ . We have called this case “optimistic” because it does not include any error on the fixed parameters. In this scenario, shown in the lower panel of Fig. 2.12, the error on  $w$  is now only around 50% larger than the corresponding error on  $\alpha$ .

We now explore two of the approaches which have been advocated in the literature to measure the value of  $w$ . Both methods involve making fits to the ratio of a measured power spectrum divided by a smooth reference spectrum. In the first approach, a parametric form is assumed for the ratio (Blake and Glazebrook 2003). The second approach is more general as it does not assume a specific form for the ratio, but instead uses the linear perturbation theory power spectrum without any further approximations (Percival et al.

2007; see also Eisenstein et al. 2005). We shall henceforth refer to these methods as the parametric and general schemes respectively. In their original forms, there are also differences in the way in which a “featureless” reference spectrum is constructed, as we will briefly discuss when describing these approaches below.

Blake and Glazebrook (2003) (see also Glazebrook and Blake 2005) studied the feasibility of extracting measurements of the acoustic oscillations from forthcoming galaxy surveys using linear perturbation theory. Their starting point is to divide the power spectrum, including the imprint of baryons, divided by a smooth reference spectrum which is chosen to be free from any signature of acoustic oscillations. This method therefore does not use any of the information contained in the overall shape of the power spectrum, which Blake & Glazebrook argue could be susceptible to large scale gradients arising from the effects we discussed in Section 2.3, such as galaxy bias or redshift-space distortions. Instead, they focused on the location and amplitude of the acoustic oscillations. The smooth reference spectrum is obtained using the zero-baryon transfer function written down by Eisenstein and Hu (1998). The parametric form suggested by Blake & Glazebrook as a fit to the resulting ratio is a Taylor expansion of the ratio of a power spectrum for cold dark matter plus a small baryonic component, divided by a pure cold dark matter power spectrum. The sound horizon, which is a free parameter in their method, is treated as the oscillation wavelength in this parametric form. This is an approximation, as the wavelength of the acoustic oscillations actually changes with wavenumber, albeit slowly, and is therefore not a constant (see eqn. 22 of Eisenstein and Hu 1998). Some authors have criticized this approach due to the sensitivity of the ratio to the choice of the reference power spectrum. Angulo et al. (2005) describe how realistic power spectra, which include nonlinear growth, bias effects and redshift-space distortions, require a “linearization” process before they become adequately described by the parametric form put forward by Blake & Glazebrook. Due to the sensitivity of the ratio to the choice of reference spectrum at low wavenumbers, Koehler et al. (2007) proposed ignoring power spectrum measurements below  $k \sim 0.05h\text{Mpc}^{-1}$  to avoid this problem (although we note that they also discuss a different approach to measuring the equation of state parameter).

Percival et al. (2007) proposed a new technique which has a number of appealing features compared with that of Blake & Glazebrook. Firstly, the shortcut of fitting an approximate parametric form to the ratio of the measured power spectrum to a reference is dropped in favour of using a full linear perturbation theory power spectrum (with a modification; see later) to model the ratio. This is completely general, and permits one

to use the most accurate description available of the linear perturbation theory power spectrum, such as the tabulated output of **CAMB**. Secondly, the reference power spectrum is defined separately in the case of the data and the linear theory model, by using a coarse rebinning of the relevant power spectrum. The reference is constructed using a spline fit to a reduced number of wavenumber bins over the range in which the spectrum in question is defined. Thus, any deviations in the general form of the measured spectrum away from linear theory are naturally accounted for in the reference spectrum. Thirdly, Percival et al. allow for a damping of the amplitude of the oscillations in the theoretical ratio beyond some wavenumber, which is treated as a free parameter in their fit. The quality of the fits is dramatically improved when damping of the higher harmonics is allowed. Percival et al. applied their method to extract the matter density parameter from the power spectrum of luminous red galaxies in the SDSS.

The majority of the results we present are obtained using the general method suggested by Percival et al. For completeness, and because Percival et al. did not actually apply their method to the extraction of the equation of state parameter, we set out the general approach step-by-step below:

1. A smooth reference spectrum (i.e. without any oscillatory features),  $P_{\text{ref}}$ , is constructed from the measured power spectrum using a cubic spline fit over the wavenumber range  $0.0046 < (k/h\text{Mpc}^{-1}) < 1.2$ , using the measured spectrum smoothed over 25 bins in wavenumber. The spline is constrained to pass through the data points in this coarse rebinning of the measured power spectrum.
2. We compute the ratio,  $R(k)$ , of the measured power spectrum,  $P(k)$ , to the reference spectrum,  $P_{\text{ref}}(k)$ , obtained in step 1:

$$R(k) = \frac{P(k)}{P_{\text{ref}}(k)}. \quad (2.11)$$

3. A linear perturbation theory power spectrum is generated with **CAMB** for the cosmological parameters used in the **BASICC** simulation. A smooth reference spectrum,  $P_{\text{ref}}^{\text{L}}$ , is defined for this spectrum in the same manner as described for the measured spectrum in Step 1, using the same wavenumber bins. A ratio,  $R_{\text{L}}$ , is derived for the linear perturbation theory spectrum by dividing by this reference spectrum.
4. The linear theory ratio,  $R_{\text{L}}$ , is compared with the measured ratio,  $R$ . Two modifications are considered to the linear theory ratio. The first is a stretch or scaling

of the wavenumber used in the linear theory ratio, as described above, to mimic the act of changing the dark energy equation of state parameter,  $w$ . The goal here is to see what variation in  $w$  can be tolerated before  $R_L$  is no longer a good fit to the measured ratio  $R$ . The second change is to allow for a damping of the oscillations beyond some characteristic wavenumber by multiplying the theoretical power spectrum by a Gaussian filter:

$$W(k) = \exp\left(-\frac{k^2}{2k_{\text{nl}}^2}\right), \quad (2.12)$$

where  $k_{\text{nl}}$  is a free parameter. Hence, the linear theory ratio is modified to:

$$R_L(k) = \left(\frac{P^L}{P_{\text{ref}}^L}(\alpha k) - 1\right) \times W(k, k_{\text{nl}}) + 1 \quad (2.13)$$

5. A likelihood is computed for each combination of the parameters  $k_{\text{nl}}$  and  $\alpha$ , assuming Gaussian errors:

$$-2 \ln L = \chi^2 = \sum_i \left(\frac{R^i - R_L^i}{\sigma^i/P^i}\right)^2 \quad (2.14)$$

where the summation is over wavenumber and  $\sigma^i$  is the error on the power spectrum estimated in the  $i^{\text{th}}$  bin (as given by Eq. 2.3). We generate a grid of models using  $200^2$  different combinations of  $\alpha$  and  $k_{\text{nl}}$  in the ranges  $[0.9, 1.1]$  and  $[0, 0.4]$  respectively.

6. Finally, the best fit values for  $\alpha$  and  $k_{\text{nl}}$  correspond to those for the model with the maximum likelihood. We obtain confidence limits on the parameter estimation by considering the models within  $\Delta\chi^2$  equal to 2.3 and 6.0; in the case of a Gaussian likelihood, these would correspond to the 68% ( $1\text{-}\sigma$  error) and 95% ( $2\text{-}\sigma$  error) confidence levels on the best fit. We note that in some cases presented later (see Fig. 2.15), the distribution of the likelihood is not Gaussian.

In some cases, we also present constraints on  $w$  derived using a slightly modified version of the approach of Blake & Glazebrook. The main difference is that we follow step 1 to construct a ratio from the measured power spectrum, rather than using a zero-baryon transfer function.

One issue to be resolved is the range of wavenumbers which should be used in the fitting process. To address this, we used the power spectrum of the dark matter measured at  $z = 6$ . We systematically varied the minimum and maximum wavenumbers used in our fit and compared the values of the scaling parameter,  $\alpha$ , recovered. Our results are

fairly insensitive to the choice of the maximum wavenumber, particularly when damping of the oscillations is included in the fitting algorithm. However, the recovered  $\alpha$  shows a systematic shift once the minimum wavenumber exceeds  $k \sim 0.1 h \text{Mpc}^{-1}$ . For minimum wavenumbers smaller than this, there is little difference in the recovered value of  $\alpha$  or in the size of the errors on  $\alpha$ , as these modes have relatively large errors in our simulation. This is encouraging news for realistic survey geometries, for which the power spectrum measured at low wavenumbers will be distorted due to the window function of the survey. In the rest of the chapter, we use the power spectrum in the wavenumber interval  $k/(h \text{Mpc}^{-1}) = [0, 0.4]$  to constrain the value of  $\alpha$ .

## 2.5 Results

In this Section, we present the expected constraints on the dark matter equation of state using the power spectra measured from our simulations. We first show how our algorithm for extracting the equation of state parameter works in practice, for dark matter particles, haloes and galaxies, comparing the results obtained in real space and redshift space (§2.5.1). We then assess the need for an accurate model of the linear theory power spectrum and the relative merits of the general and parametric fitting procedures (§2.5.2). In §2.5.3, we present our main results, which are summarized in Fig. 2.19 and Tables 2.2 and 2.3, which list the best-fitting value of  $\alpha$  and the estimated error for different samples of galaxies at  $z = 1$ , along with the corresponding fractional error in  $w$ . Finally, in §2.5.4, we use the results presented in §2.5.3 to make forecasts for the accuracy with which several forthcoming surveys will be able to measure the value of  $w$ .

### 2.5.1 The algorithm to extract the scale of the acoustic oscillations in action

We present a series of plots for samples at  $z = 0$ , which illustrate the various stages in the fitting process. Fig. 2.13 shows the power spectra measured for different tracers, both in real space and redshift space. The sample of dark matter haloes includes all objects with a mass in excess of  $5.4 \times 10^{12} h^{-1} M_{\odot}$ . The galaxy sample is magnitude-limited with a space density of  $\bar{n} = 5 \times 10^{-4} h^{-3} \text{Mpc}^3$ . For reference, the linear perturbation theory power spectrum for the mass at  $z = 0$  is shown by the blue line in each panel: this is the power spectrum of the dark matter measured in real space at  $z = 15$ , scaled by the ratio of growth factors in order to have the amplitude expected at  $z = 0$ . It is

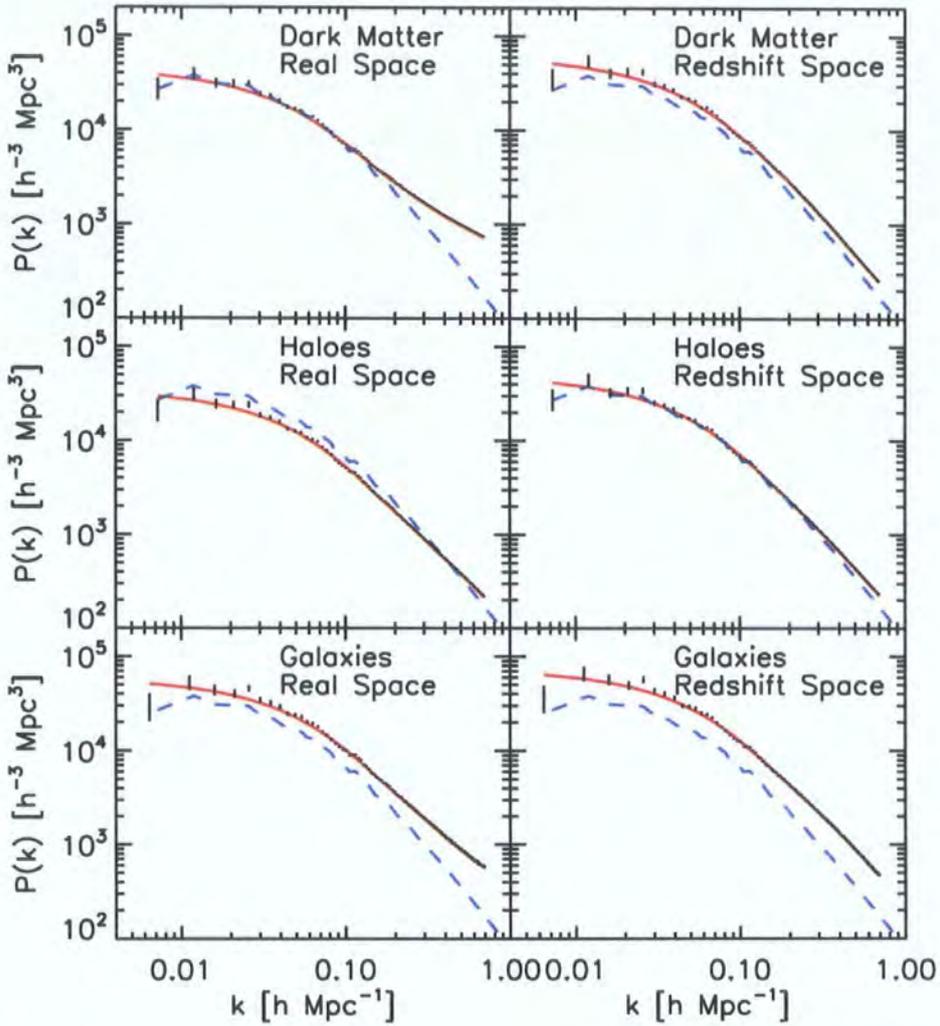


Figure 2.13: The power spectra of dark matter particles, dark matter haloes and galaxies at  $z = 0$  (error bars). The real-space power spectra are plotted in the left hand column and the redshift-space power spectra appear in the right hand column. The red curves show the reference spectra derived from the measured spectra using a cubic spline fit, as described in Section 2.4. The blue curve is the same in each panel, showing the linear perturbation theory prediction for the  $z = 0$  matter power spectrum (plotted using a high redshift output obtained from the BASICC simulation, which has been scaled in amplitude according to the difference in growth factors between the two epochs expected in linear perturbation theory) The errors on the power spectrum are estimated using Eq. 2.3.

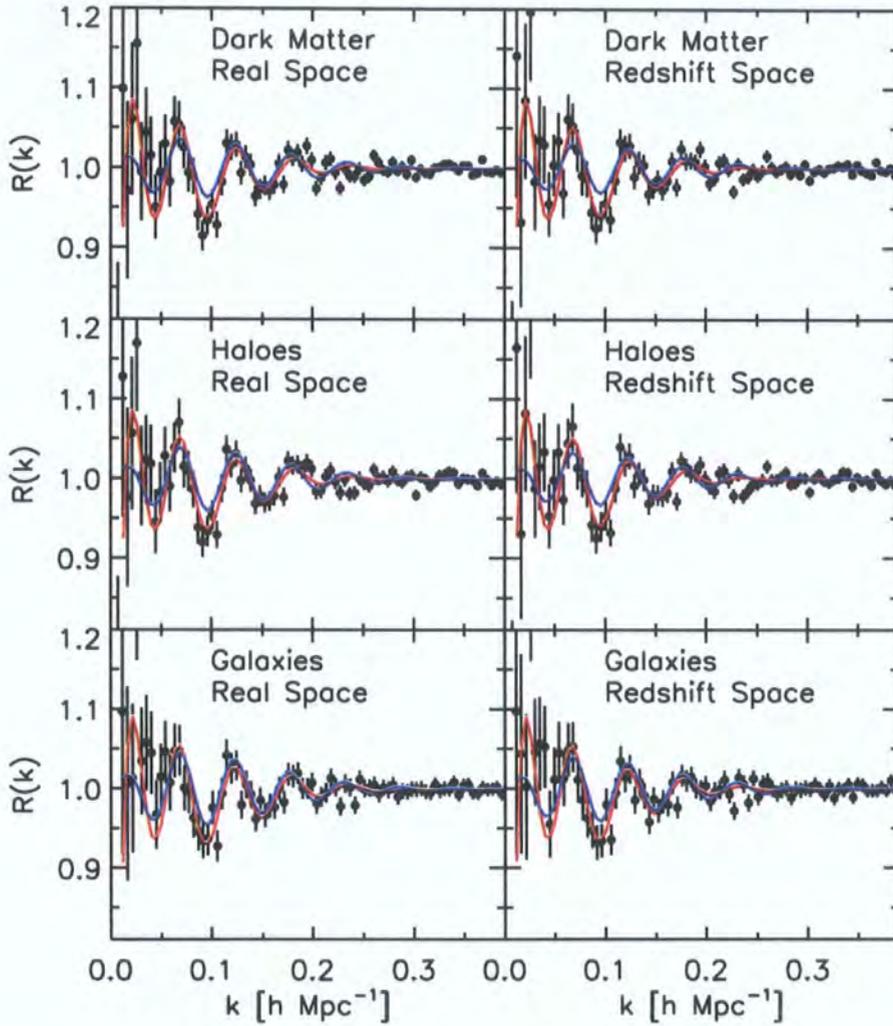


Figure 2.14: The ratio of the measured power spectrum divided by a smooth reference spectrum. The symbols correspond to the measurements plotted in Fig. 2.13 divided by the red curve in each panel of that figure. The red lines here show the best-fitting model in each case using the general method and the blue curves show the best fit for the parametric method. The errors on the power spectrum are estimated using Eq. 2.3.

important to bear in mind that the  $y$ -axis in this plot covers more than a factor of one thousand in amplitude. Fig. 2.13 shows that there is considerable variation in the power spectra measured for different types of objects, and between the results in real space and redshift space, which re-enforces the points made in Section 2.3 regarding deviations from the predictions of linear perturbation theory on large scales. The red curve in each panel shows the corresponding reference power spectrum, which is constructed from the measured power spectrum as explained in Section 2.4.

In Fig. 2.14, the symbols show the ratio obtained by dividing the measured power spectrum by the appropriate reference spectrum for the same samples plotted in Fig. 2.13. The ratios look remarkably similar for the different tracers up to  $k \approx 0.15 h\text{Mpc}^{-1}$ . Beyond this wavenumber, the appearance of the oscillations varies from panel to panel, but the ratio stays close to unity. This similarity illustrates how well the approach for producing the reference spectrum works. The red curves in each panel show the best-fitting model produced in the general scheme whilst the blue curves show the fit obtained in the parametric approach. The best fits have somewhat different forms at wavenumbers below  $k \sim 0.05 h\text{Mpc}^{-1}$ . The constraints on the values of the parameters  $k_{\text{nl}}$  and  $\alpha$  are presented in Fig. 2.15, where we show the 1, 2 and 3- $\sigma$  ranges in the case of two parameters, computed assuming Gaussian errors. There is a weak systematic trend for the best-fitting result for  $\alpha$  to shift to lower values when galaxies are considered instead of the dark matter. The errors on the recovered parameters are larger in the case of galaxies than for the dark matter or for haloes, reflecting the lower signal-to-noise of the predicted galaxy power spectrum.

### 2.5.2 Two tests of the algorithm

Before presenting the main results of applying our algorithm to extract the acoustic oscillation scales for various samples drawn from the BASICC run, we use the L-BASICC ensemble to address two questions: 1) How accurately do we need to model the linear perturbation theory matter power spectrum to avoid introducing a systematic bias into the results for the oscillation scale? 2) How does the performance of the new method for constraining the oscillation scale introduced in this chapter compare with earlier approaches? To help answer these questions, we use the power spectrum of the dark matter measured from the L-BASICC runs in real space at  $z = 0$  and  $z = 3.8$ , the highest output redshift besides the initial conditions. The results of applying our standard algorithm for extracting the oscillation scale are shown by the red histogram labelled **CAMB** in Fig. 2.16,

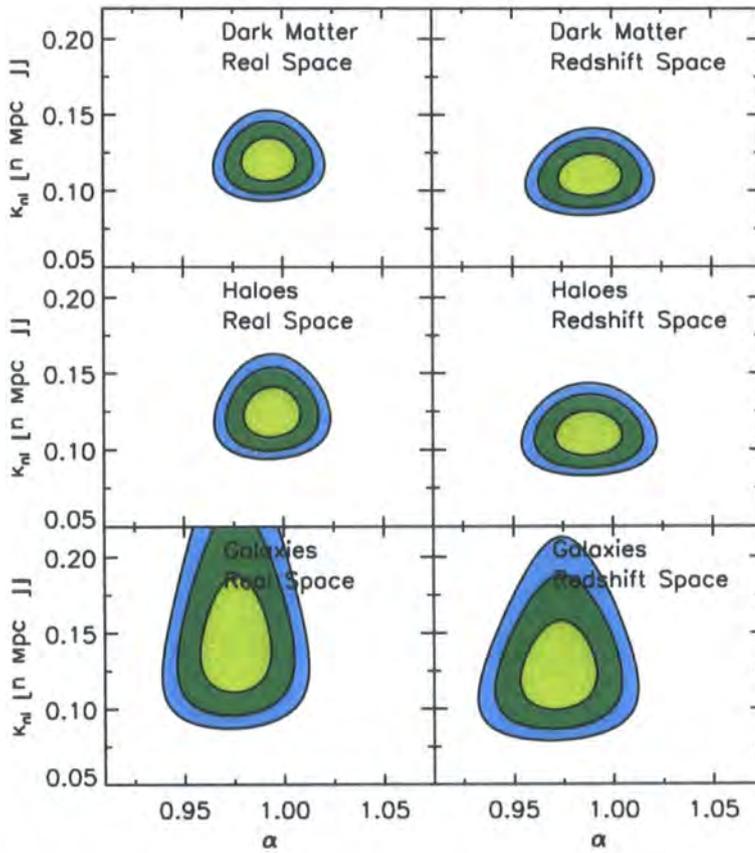


Figure 2.15: The constraints on the parameters  $k_{nl}$  and  $\alpha$  for the power spectra plotted in Fig. 2.13. The contours show the 1, 2 and 3- $\sigma$  confidence limits for two parameters.

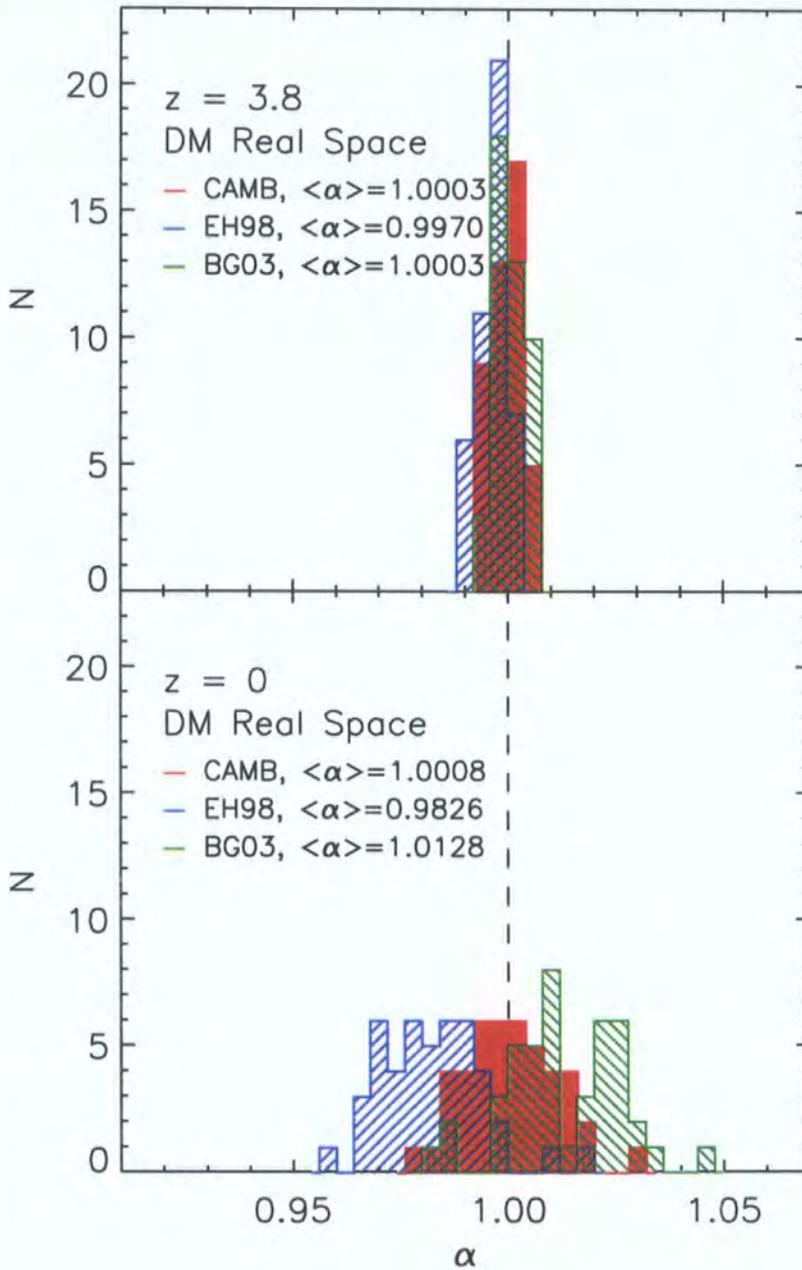


Figure 2.16: The best fit value for the scaling parameter  $\alpha$ , recovered from the ensemble of low resolution simulations, using the dark matter power spectrum in real space. The results are shown for two different redshifts:  $z = 3.8$  (top) and  $z = 0$  (bottom). The histograms marked **CAMB** and **BG03** show the results for the general and parametric fitting procedures, respectively. The blue histogram shows the results if the general method is followed with the **CAMB** power spectrum replaced by the formula for the linear theory power spectrum presented by Eisenstein & Hu (1998).

which gives the distribution of the best-fitting value of  $\alpha$ . The ensemble returns an unbiased mean value for the stretch parameter,  $\alpha = 1$ . At  $z = 3.8$ , the standard deviation on the best fit is 0.3%; by  $z = 0$ , this rises to 1%.

To address the first issue above, regarding how well we need to model the linear theory power spectrum to get an unbiased result for the oscillation scale, we replace the **CAMB** generated power spectrum in our algorithm by the approximation introduced by Eisenstein and Hu (1998). These authors proposed a physically motivated expression for the linear theory power spectrum, with parameters set to achieve a reasonable match to the results obtained from detailed calculations using Boltzmann codes over a much wider range of wavenumbers than are typically considered for baryonic acoustic oscillations. Eisenstein & Hu's motivation was to provide physical insight into the form of the power spectrum in a cold dark matter universe and to produce a code which could rapidly calculate large numbers of power spectra for grids cosmological parameters. Of course, the correct approach in our fitting procedure is to use the same code to compute the linear theory spectrum as was used to generate the initial conditions in the N-body simulation. In the case of real data, we do not have the luxury of knowing which Boltzmann code to use, so we should use the one which claims to be the most accurate representation of the model we are testing. Nevertheless, it is instructive to perform this test to see what error is introduced by using a less accurate calculation of the transfer function. The choice of Eisenstein & Hu's code is particularly relevant for this purpose as Blake & Glazebrook used this formalism to inspire their parametric expression to fit the acoustic oscillations. The use of Eisenstein & Hu's formalism to model the linear theory power spectra generated with **CAMB** introduces a small but measurable systematic shift in the mean value of  $\alpha$ . At  $z = 0$ , the mean  $\alpha$  indicated by the blue histogram in Fig. 2.16 is  $0.98 \pm 0.01$ .

We answer the second question by adopting the fitting algorithm of Blake & Glazebrook (2003), which assumes a parametric form for the ratio of the power spectrum with baryons to a smooth, cold dark matter only power spectrum. Changing the fitting method in this way also introduces a similar magnitude of shift in the best-fitting value of  $\alpha$ . The green histogram shows the results when we use the parametric approach introduced by Blake and Glazebrook (2003). The mean value of  $\alpha$  in this case is  $1.01 \pm 0.01$ . These shifts are small but one must bear in mind that the corresponding bias in the dark energy equation of state parameter is several times larger than the shift in  $\alpha$ .

id	Sel I	Sel II	Real space					Redshift space						
	$\bar{n}$ $h^3\text{Mpc}^{-3}$		$b$	$\bar{n}P$	$k_{nl}$ $h/\text{Mpc}$	$\alpha$	$\Delta\alpha$ %	$\Delta\alpha$ %	$b$	$\bar{n}P$	$k_{nl}$ $h/\text{Mpc}$	$\alpha$	$\Delta\alpha$ %	$\Delta\alpha$ %
							(SE07)						(SE07)	
DM			0.99	3567	0.120	0.993	0.91	1.02	1.15	3635	0.110	0.989	1.05	1.17
A	$5.0e-4$		1.18	1.78	0.144	0.975	1.16	1.10	1.32	2.15	0.125	0.972	1.26	1.23
B	$2.5e-4$		1.33	1.11	0.155	0.971	1.34	1.18	1.47	1.34	0.139	0.966	1.35	1.23
C	$2.5e-4$	red	1.32	1.15	0.152	0.978	1.35	1.21	1.46	1.36	0.127	0.975	1.49	1.37
D	$2.5e-4$	strong	1.06	0.67	0.155	0.956	1.75	1.41	1.20	0.86	0.138	0.956	1.67	1.42
E	$2.5e-4$	blue	1.03	0.66	0.141	0.964	1.92	1.56	1.17	0.83	0.130	0.962	1.79	1.53
F	$2.5e-4$	weak	1.30	1.16	0.132	0.980	1.55	1.40	1.44	1.34	0.115	0.972	1.66	1.54
haloes	$5.9e-5$		1.56	0.81	0.197	0.980	1.32	1.07	1.71	1.04	0.148	0.975	1.43	1.25

Table 2.2: The results of applying the general fitting procedure described in §2.4 to power spectra measured for different galaxy catalogues at  $z = 0$ . The first row gives the results for the dark matter and the final row lists results for a sample of dark matter haloes (all haloes with mass in excess of  $2.7 \times 10^{13} h^{-1} M_{\odot}$ ). The first column gives the label of the sample, as defined in Section 2.2.2. The second column gives the space density of galaxies. The first two samples, A and B, are constructed by applying a magnitude limit. Samples C-F are derived from sample A by applying a second selection criterion, as listed in the third column. Samples C and E correspond to the red and blue halves of sample A respectively. Samples D and F comprise the 50% of galaxies from sample A with the strongest and weakest (in terms of equivalent width) OII[3727] emission lines, respectively. Column 4(10) gives the effective bias of the sample, computed from the square root of the ratio of the measured galaxy power spectrum in real (redshift) space to the real space power spectrum of the dark matter over the wavenumber interval  $0.01 < (k/h\text{Mpc}^{-1}) < 0.05$ . Column 5(1) gives the ratio of the clustering signal to the shot noise for the power spectrum measurement, averaged over the wavenumber range  $0.19 < (k/h\text{Mpc}^{-1}) < 0.21$ . Columns 6 and 7 (12 and 13) give the best-fitting values of the scaling parameter  $\alpha$  and the  $1 - \sigma$  error on the fit, in real (redshift) space. Column 9 (15) gives the error expected on the scale parameter from Seo & Eisenstein (2007). The rms Lagrangian displacement was set equal to 1 over the best fit non linear scale ( $1/k_{nl}$ ) for each case.

id	Sel I	Sel II	Real space					Redshift space						
	$\bar{n}$ $h^3\text{Mpc}^{-3}$		$b$	$\bar{n}P$	$k_{\text{nl}}$ $h/\text{Mpc}$	$\alpha$	$\Delta\alpha$ %	$\Delta\alpha$ %	$b$	$\bar{n}P$	$k_{\text{nl}}$ $h/\text{Mpc}$	$\alpha$	$\Delta\alpha$ %	$\Delta\alpha$ %
			(SE07)					(SE07)						
DM			0.99	1269	0.163	0.997	0.61	0.68	1.29	1710	0.133	0.991	0.77	0.88
A	$5.0e-4$		1.34	0.87	0.188	0.980	1.30	1.10	1.60	1.19	0.164	0.976	1.21	1.07
B	$2.5e-4$		1.31	0.43	0.212	0.975	2.02	1.47	1.57	0.59	0.174	0.970	1.72	1.38
C	$2.5e-4$	red	1.39	0.48	0.235	0.977	1.81	1.32	1.65	0.65	0.208	0.975	1.52	1.17
D	$2.5e-4$	strong	1.31	0.40	0.624	0.971	1.90	1.14	1.57	0.55	0.186	0.970	1.79	1.31
E	$2.5e-4$	blue	1.30	0.40	0.219	0.973	2.31	1.47	1.56	0.54	0.159	0.962	1.98	1.48
F	$2.5e-4$	weak	1.37	0.47	0.218	0.987	1.91	1.38	1.63	0.64	0.190	0.978	1.61	1.25
haloes	$5.9e-5$		3.07	0.59	0.226	1.000	1.65	1.24	3.34	0.77	0.146	0.994	1.82	1.53

Table 2.3: Same as 2.2 but at  $z = 1$ .

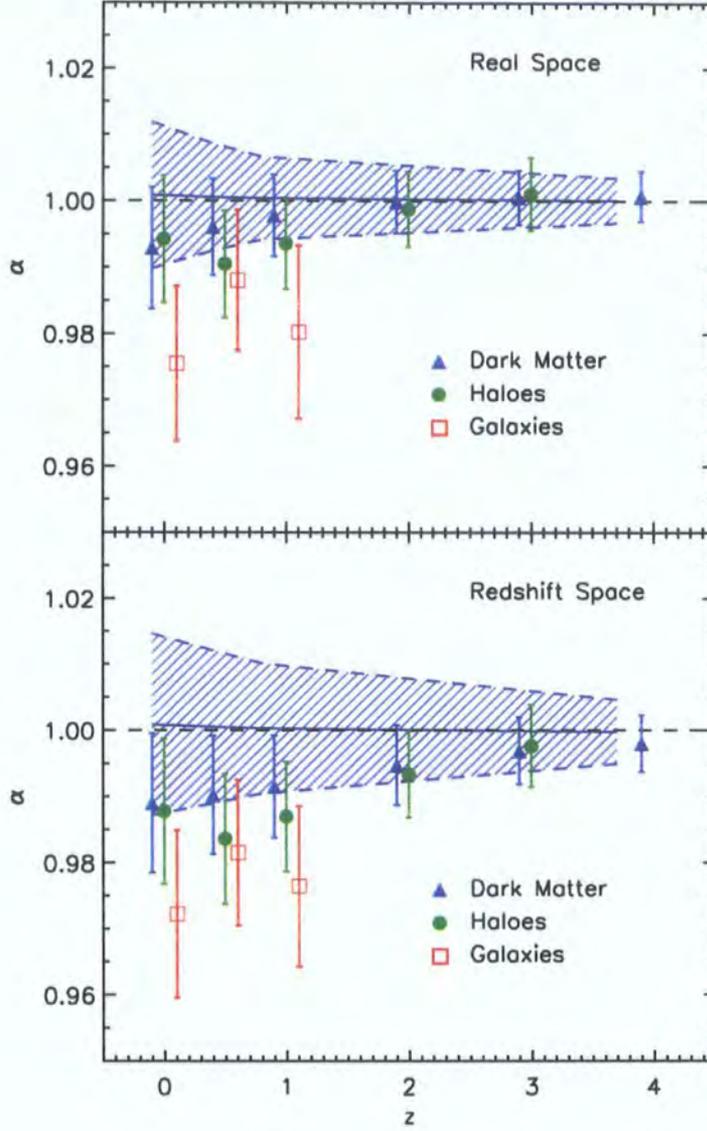


Figure 2.17: The best-fitting value of the scale factor  $\alpha$  as a function of redshift, for different tracers of the density distribution, in real space (top) and redshift space (bottom). The symbols show results from the high resolution BASICC simulation: dark matter (blue triangles), dark matter haloes with mass in excess of  $5.4 \times 10^{12} h^{-1} M_{\odot}$  (green circles) and galaxies (red squares). The error bars show the  $1\text{-}\sigma$  range on  $\alpha$ , calculated from  $\Delta\chi^2$ . The hatched region shows the central 68% range of the results obtained using the dark matter in the ensemble of low resolution simulations. Recall that  $\alpha = 1$  corresponds to an unbiased measurement of the equation of state parameter,  $w$ , and that  $\delta w \approx 4\delta\alpha$  at  $z = 1$ .

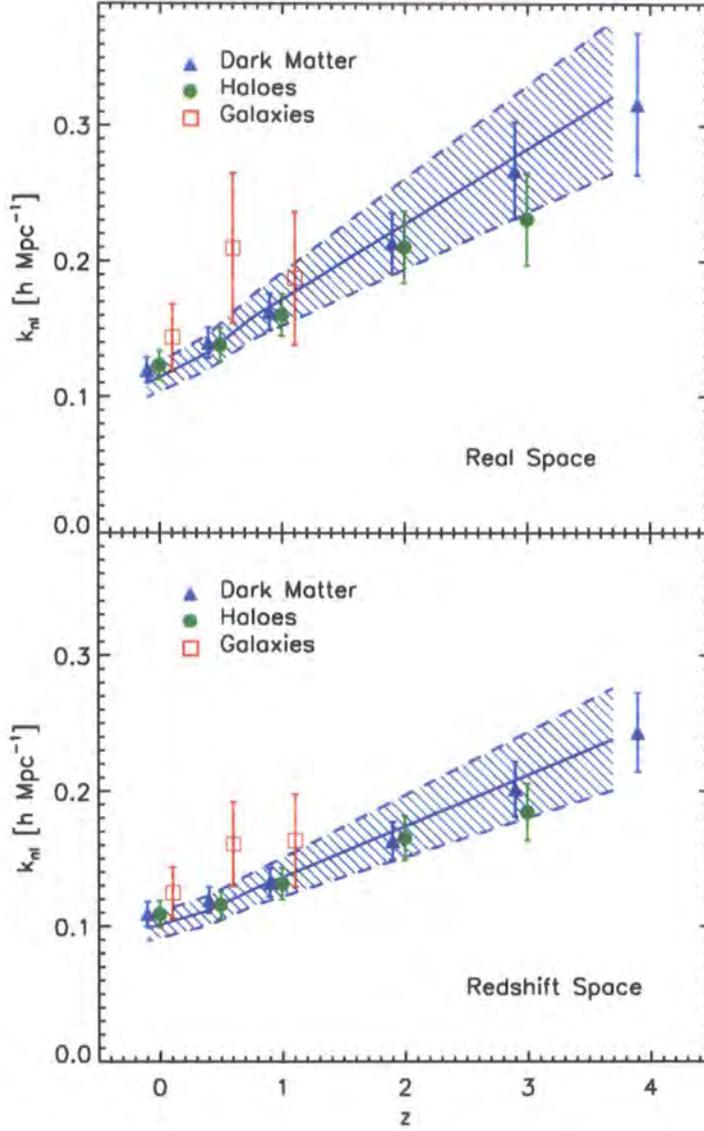


Figure 2.18: The best-fitting value of the damping scale  $k_{nl}$  as a function of redshift, for different tracers of the density distribution, in real space (top) and redshift space (bottom). The symbols show results from the high resolution BASICC simulation: dark matter (blue triangles), dark matter haloes with mass in excess of  $5.4 \times 10^{12} h^{-1} M_{\odot}$  (green circles) and galaxies (red squares). The error bars show the  $1\text{-}\sigma$  range on  $\alpha$ . The hatched region shows the central 68% range of the results obtained using the dark matter in the ensemble of low resolution simulations.

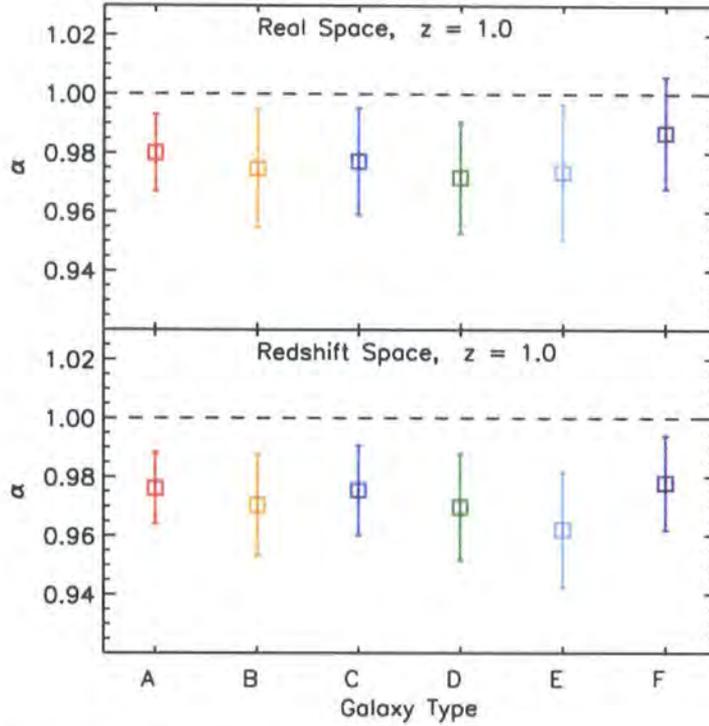


Figure 2.19: The recovered value of the stretch parameter  $\alpha$  for the galaxy samples listed in Table 2.1. Recall that  $\alpha = 1$  corresponds to the equation of state parameter  $w = -1$ . At  $z = 1$ , a shift in  $\alpha$  away from unity implies a shift in the recovered value of  $w$  given by  $\delta w \approx 4\delta\alpha$ .

### 2.5.3 The main results

We now turn our attention back to the general results shown in Fig. 2.17 and Fig. 2.18, and discuss the conclusions for different tracers of the density field in turn. In these plots, the symbols refer to the constraints obtained from the high resolution BASICC simulation and the shading shows results from the ensemble of low resolution simulations, L-BASICC.

The blue triangles in Fig. 2.17 show the values obtained for  $\alpha$  from the power spectrum of the dark matter. There is a trend for the best-fitting value to deviate away from unity with decreasing redshift, although the result at  $z = 0$  is still within  $1\text{-}\sigma$  of  $\alpha = 1$ . The mean of the ensemble of low resolution runs does not, however, show any deviation away from  $\alpha = 1$  as a function of redshift, although the scatter on the recovered value of  $\alpha$  increases towards the present day. If we examine the analogous results for individual simulations taken from the low resolution ensemble, we find a wide range of behaviour for the best-fitting value of  $\alpha$  for the dark matter. Some low resolution runs give results which look like the high resolution one, whereas others show deviations away from  $\alpha = 1$ ,

with values of  $\alpha > 1$ , as  $z = 0$  is approached. The trend seen for the dark matter in the high resolution run serves to illustrate the importance of sampling fluctuations, even in such large volumes. In redshift space, the scatter in the recovered value of  $\alpha$  is larger than in real space (see also Seo and Eisenstein 2005; Eisenstein et al. 2007b).

To obtain the errors quoted in Tables 2.2 and 2.3 on the parameters  $\alpha$  and  $k_{\text{nl}}$ , we assume Gaussian mode counting errors on the power spectra measured in the BASICC simulation, as given by Eq. 2.3. In Fig. 2.3, we showed that this simple estimate of the errors on the power spectrum agreed fairly well with the scatter found in the measurements from the L-BASICC ensemble, particularly for the case of the dark matter. We have extended this comparison to look at how the errors on  $\alpha$  and  $k_{\text{nl}}$  quoted in Tables 2.2 and 2.3 match the scatter in these parameters obtained from the L-BASICC runs. We find the scatter estimated from the ensemble is somewhat larger than the error inferred using the mode counting argument. At  $z = 0$ , the mode counting errors are 20% smaller for  $\alpha$  for the dark matter in real space. In redshift space, the discrepancy increases to nearly 30%. The mismatch between the two estimates is smaller at  $z = 1$ . The level of disagreement is not remarkable. It could be the case that the scatter from the ensemble has not converged, even with 50 realizations of the density field. A more likely explanation, particularly in view of the redshift dependence of the discrepancy, is mode coupling in the power spectrum measurements arising from nonlinearities and redshift space distortions, which could increase the variance in the power spectrum compared with the Gaussian estimate.

Fig. 2.18 shows that there is a strong trend for the best-fitting value of the smoothing scale,  $k_{\text{nl}}$ , to decrease with decreasing redshift. This results from the oscillations being erased and modified down to smaller wavenumbers as the nonlinearities in the density field grow. The variation of the smoothing scale  $k_{\text{nl}}$  on redshift is well described by a linear relation:  $k_{\text{nl}} = a + bz$ . In real space,  $a = 0.108 \pm 0.0082$  and  $b = 0.054 \pm 0.0110$ . In redshift space,  $a = 0.096 \pm 0.0074$  and  $b = 0.036 \pm 0.0094$ .

The constraints on  $\alpha$  and  $k_{\text{nl}}$  for dark matter haloes (with masses in excess of  $5 \times 10^{12} h^{-1} M_{\odot}$ ) are plotted with green circles in Figs. 2.17 and 2.18. The parameter constraints obtained for this sample of haloes are very similar to those found for the dark matter, except for the value of  $k_{\text{nl}}$  at high redshift. Considering haloes in place of dark matter represents a step closer to the observations, so it is reassuring that the conclusions do not change significantly.

Finally, in Figs. 2.17 and 2.18, we show using red squares the results for magnitude-limited samples of galaxies. The magnitude limit is varied with redshift such that in each

case the galaxy sample has a space density  $\bar{n} = 5 \times 10^{-4} h^{-3} \text{Mpc}^3$ . There is a weak systematic shift in the best-fitting values of  $\alpha$  compared with the results obtained for the dark matter. At the same time the signal to noise of the power spectrum measurement is lower for the galaxy samples than for the dark matter, so the errors on the best-fitting parameters are correspondingly larger for the galaxies. The galaxy samples are consistent with  $\alpha = 1$  at slightly over  $1\text{-}\sigma$ . The size of this systematic shift is comparable to the random measurement errors, so we cannot reach a firm conclusion. It will be very interesting to repeat our calculation with a larger simulation volume to reduce the size of the random errors and to assess if such shifts could genuinely provide an ultimate limitation to the accuracy of this method.

As a result of using a semi-analytic galaxy formation model which makes predictions for the observable properties of galaxies, we can vary the selection criteria used to construct samples and compare the constraints on the equation of state. The results of this exercise at  $z = 1$  are presented in Tables 2.2 and 2.3 and in Fig. 2.19, where we consider a range of samples defined either by a simple magnitude limit, or by a magnitude limit applied in combination with a colour cut or a restriction on the strength of an emission line. The key result from comparing the constraints for different samples is that whilst there are no strong systematic differences between the results, the accuracy of the constraints varies significantly. For example, using a catalogue of red galaxies, we predict that one could measure the dark energy equation of state with an accuracy 40% better than with the same number density of galaxies chosen by the strength of their emission lines.

We compare the error on the acoustic scale extracted from our simulations with the results of the prescription set out by Seo and Eisenstein (2007). The Seo and Eisenstein (2007) algorithm contains a parameter which is equivalent to  $1/k_{\text{nl}}$ . If we use our best fitting values of  $k_{\text{nl}}$ , we find that the Seo & Eisenstein prescription gives a similar estimate of the error on the acoustic scale to that we obtain by fitting directly to the simulation results. However, if we use the value of  $k_{\text{nl}}$  suggested by Seo and Eisenstein (2007), which they extract from a dark matter simulation, we find that their prescription gives an optimistic estimate of the error on  $\alpha$ . The reason we recover a larger value of  $k_{\text{nl}}$  from our galaxy samples than we do for the dark matter is due to the increased discreteness shot noise in these samples, which results in noisier power spectra at high  $k$ . This causes an elongation in the confidence levels in the  $k_{\text{nl}}$  versus  $\alpha$  plane.

It is interesting to compare the results for the dark matter and for the galaxy samples with those for a set of massive haloes. Tables 2.2 and 2.3 also give the constraints on  $\alpha$

and  $k_{\text{nl}}$  for a sample of massive haloes (see also Angulo et al. 2005). There are 142 000 haloes in the BASICC output at  $z = 1$  with a mass in excess of  $2.7 \times 10^{13} h^{-1} M_{\odot}$ . Although the effective bias of this sample of massive haloes is greater than that of any of the galaxy samples listed in Tables 2.2 and 2.3, the reduction in space density means that  $\bar{n}P \approx 1$  and the estimated error on  $w$  is comparable to that found for the galaxy samples.

### 2.5.4 Forecasts for future surveys

We can use the results presented in Tables 2.2 and 2.3 to make a rough estimate of the accuracy with which future surveys are likely to be able to constrain the scale of the acoustic oscillations. This can be done using a simple calculation motivated by the expression for the fractional error in the power spectrum given by Eq. 2.3. We assume that the error in the distance scale,  $\Delta\alpha$ , scales with the volume of the survey,  $V_{\text{survey}}$ , and the product of the space density of galaxies and the power spectrum,  $\bar{n}P(k = 0.2h\text{Mpc}^{-1})$ , as:

$$\Delta\alpha \propto \frac{1}{\sqrt{V_{\text{survey}}}} \left( 1 + \frac{1}{\bar{n}P} \right). \quad (2.15)$$

The constant of proportionality can be set for a particular galaxy sample using the results given in Table 2.2 and 2.3 .

The WiggleZ survey is currently underway and will measure redshifts for 400,000 blue galaxies over 1000 square degrees in the redshift interval  $z = 0.5 - 1.0$  (Glazebrook et al. 2007). For the cosmological parameters adopted in this chapter, this gives a comoving volume of  $1.13 h^{-3} \text{Gpc}^3$ . Using the blue colour selected sample or the large equivalent width sample from Tables 2.2 and 2.3, and assuming  $\bar{n}P \sim 1$  for WiggleZ galaxies, somewhat higher than we find in our simulation, we estimate that this survey will measure the distance scale to an accuracy of  $\Delta\alpha \sim 2\%$ , which is similar to that claimed by Glazebrook et al. using linear perturbation theory.

The WFMOS survey has been proposed to motivate the construction of a new spectrograph for the Subaru telescope (Glazebrook et al. (2005)). This will target galaxies with a space density of  $\bar{n} = 5 \times 10^{-4} h^3 \text{Mpc}^{-3}$  in the redshift interval  $z = 0.5 - 1.3$  over 2000 square degrees, covering a volume of  $4.4 h^{-3} \text{Gpc}^3$ . (There is also a WFMOS survey which will target  $z = 3$  galaxies but over a much smaller solid angle.) Using sample A from Tables 2.2 and 2.3, and adopting  $\bar{n}P = 1$ , we obtain an estimated error of  $\Delta\alpha = 0.83\%$ , again in good agreement with Glazebrook et al.

Photometric surveys can generally cover a larger solid angle than spectroscopic surveys down to a fainter magnitude limit. The fainter magnitude limit results in a higher median redshift and a broader redshift distribution for the survey galaxies, which means that a larger volume is covered. However, the limited accuracy of photometric redshift estimates means that in practice Fourier modes are lost and the effective volume of the survey is greatly reduced. Blake & Bridle (2005) estimate that the factor by which the survey volume is reduced is  $\approx 12(\delta z/(1+z)/0.03)$ , where  $\delta z/(1+z)$  is the error in the photometric redshifts.

The Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) survey will map  $3\pi$  steradians of the sky (<http://pan-starrs.ifa.hawaii.edu/public/home.html>). Cai et al. (2008) show that the median redshift of the  $3\pi$  survey will be  $z \approx 0.5$ , with a tail extending to  $z \approx 1.2$ . The volume of the survey, assuming that 20,000 square degrees cover low-extinction parts of the sky and give high quality clustering measurements, is around  $41 h^{-3} \text{Gpc}^3$ . Taking sample A from Tables 2.2 and 2.3, and setting  $\bar{n}P \gg 1$ , as appropriate for the relatively high space density of galaxies in a photometric sample, and allowing for the reduction in the effective volume caused by a photometric redshift error of  $\delta z/(1+z) = 0.03$ , gives a forecast error on the oscillation scale of  $\Delta\alpha \sim 0.5\%$ . In the more likely event that the photometric redshift errors are twice as large,  $\delta z/(1+z) \sim 0.06$ , this figure increases to  $\Delta\alpha \sim 0.7\%$ .

Remembering the crude conversion  $\Delta w \approx 4\Delta\alpha$  from Section 2.4, this means that the next generation of galaxy surveys is unlikely to deliver 1% errors on a constant equation of state from BAO measurements used in isolation from other cosmological data. A survey with almost an order of magnitude more effective volume than Pan-STARRS will be needed to achieve this target. This will require an all-sky, spectroscopic galaxy redshift survey, such as the SPACE mission being proposed to ESA's Cosmic Vision call. SPACE will measure redshifts for galaxies in the interval  $0.5 < z < 2$ , covering around  $150h^{-3}\text{Gpc}^3$ . Extrapolating from Sample A, we forecast that an error in the oscillation scale of  $\Delta\alpha \sim 0.15\%$  could be achieved with SPACE. In the case of the pessimistic translation to an error on  $w$  considered in Section 2.4, this corresponds to  $\Delta w \sim 0.6\%$ ; in the optimistic scenario, we expect a constraint of  $\Delta w \sim 0.23\%$ .

## 2.6 Conclusions

In the next five to ten years, several proposed galaxy surveys will allow high precision measurements of the clustering of galaxies on the scale of the acoustic oscillations at intermediate and high redshifts. Both photometric and spectroscopic surveys are planned, which will cover volumes up to tens of cubic gigaparsecs and will contain hundreds of thousands to hundreds of millions of galaxies. There is a clear need to ensure that theoretical predictions develop apace with sufficient accuracy and realism to allow such datasets to be fully exploited and to uncover any possible systematic errors in this cosmological test to uncover the nature of the dark energy.

Early theoretical work in this area used linear perturbation theory (Blake and Glazebrook 2003; Hu and Haiman 2003; Glazebrook and Blake 2005). Recently, more physical calculations have been carried out using N-body simulations with cubes of side  $500 - 1100 h^{-1} \text{Mpc}$  (Seo and Eisenstein 2003, 2005; Schulz and White 2006; Huff et al. 2007; Seo and Eisenstein 2007). In this chapter, we have improved upon previous modelling work in three ways. Firstly, we have used a simulation volume comparable to the largest of the currently proposed spectroscopic surveys. This allows us to accurately follow the growth of density fluctuations on an ultra large scales in excess  $100 h^{-1} \text{Mpc}$ , the scales of interest for the acoustic oscillations, which can only be followed approximately in smaller computational volumes. In particular, a large volume is necessary to obtain accurate predictions for bulk flows, which are sensitive to the power spectrum at low wavenumbers. The only published work with a larger simulation volume used the Hubble Volume simulation (Angulo et al. 2005; Koehler et al. 2007). The Hubble Volume has a larger particle mass than the **BASICC**, which restricted these studies to consider either cluster mass dark matter haloes (Angulo et al. 2005) or a simple biasing scheme to add galaxies (Koehler et al. 2007). Secondly, through the use of a large number of particles, we are able to resolve the majority of the haloes which are likely to host the galaxies which will be observed in the forthcoming surveys. Thirdly, we use a semi-analytic galaxy formation model to populate the simulation with galaxies. Unlike other studies which use phenomenological biasing schemes or the halo occupation model to add galaxies, this allows us to predict the shape and amplitude of the galaxy power spectrum and the signal-to-noise of the clustering expected for different galaxy selections.

We use our N-body simulation in combination with a galaxy formation model to make the connection between the linear perturbation theory prediction for the matter power

spectrum and the power spectrum of galaxies. We do this in a series of steps, starting with power spectrum of the dark matter, looking at the impact of the nonlinear growth of fluctuations and peculiar motions or redshift-space distortions, before examining the power spectrum of dark matter haloes and, finally, galaxies. A number of conclusions are reached from this study: i) The nonlinear evolution of the dark matter power spectrum is apparent even on scales larger than the sound horizon scale. Although the deviation from linear theory is only a few percent, the coupled evolution of different Fourier modes means that these scales need to be followed accurately to get the correct behaviour at higher wavenumbers. ii) The form of the distortion of the power spectrum due to peculiar motions is extremely sensitive to the type of object under consideration, being quite different for the cases of dark matter, dark haloes and galaxies. Moreover, different galaxy selections give different redshift-space distortions. iii) Galaxy bias is scale dependent and sensitive to the selection applied for wavenumbers  $k > 0.15h\text{Mpc}^{-1}$ . Eisenstein et al. 2007a discuss a technique which attempts to reconstruct the linear density field from an observed distribution of objects. The reconstruction can reduce the damping of the higher harmonic oscillations in the power spectrum, thereby increasing the statistical significance of the acoustic scale measurement and diminishing any systematic effects caused by departures from linearity. It will be interesting to apply this method to the galaxy samples presented in this chapter, to see if this approach still works at the required level in the case of biased tracers of the linear density field.

We also present a new method to extract the dark energy equation of state parameter, based upon an approach put forward by Percival et al. (2007). The method involves dividing the measured power spectrum by a smooth reference spectrum and comparing the resulting ratio to the predictions of linear perturbation theory. The algorithm has three key advances over earlier work, which can be credited to Eisenstein et al. (2005) and Percival et al. (2007): i) The reference spectrum is derived from the measured spectrum, which avoids the need to apply major corrections to a linear theory reference. ii) The measured ratio is compared to a prediction generated using **CAMB**, which is more accurate than assuming a parametric form for the ratio based on a Taylor expansion. iii) The linear theory ratio is modified by allowing the higher-order oscillations to be damped, which improves the fit to the measured ratio. Changing the value of the equation of state parameter is approximately equivalent to rescaling the wavenumber in the predicted power spectrum ratio; at  $z = 1$ , a 1% shift in wavenumber is equivalent to a 4% shift in the recovered value of  $w$ .

We explore the constraints on the dark energy equation of state using different tracers of the density field. By applying our algorithm for extracting the oscillation scale to the L-BASICC ensemble, we have provided the most stringent test to date of usefulness of baryonic acoustic oscillations for measuring the equation of state of the dark energy. For the case of the dark matter, there is no significant bias in the recovered oscillation scale, compared with the value expected from linear perturbation theory. Within a given simulation, we find that 1% deviations from the underlying length scale are possible although these are only at the  $1\text{-}\sigma$  level. Such excursions are the result of sampling variance arising from the finite volume of the computational box, which are important even in a simulation of the volume of the BASICC. The error on the scale factor recovered from galaxy samples is larger than that found for the dark matter, reflecting the lower signal-to-noise of the galaxy power spectrum measurements. Different galaxy selections lead to variations in the clustering strength and hence in the error expected in the scale factor.

Currently, the best constraints on the equation of state parameter come either from combining datasets, such as the power spectrum of galaxy clustering and measurements of the microwave background radiation (e.g. Sánchez et al. 2006) or from the Hubble diagram of Type Ia, with priors on the flatness of the Universe and the matter density (Riess et al. 2004). For example, Wood-Vasey et al. (2007) combine high redshift SNe Ia from the ESSENCE Supernova Survey with the measurement of the BAO made by Eisenstein et al. (2005), and, assuming a flat universe, constrain a constant equation of state to have  $w = -1.05_{-0.12}^{+0.13}(\text{stat.}) \pm 0.11(\text{sys.})$ , consistent with a cosmological constant. Possible contributions to the systematic error include the degree of dust extinction in the SNe host galaxy, evolution in the properties of SNe with redshift and local calibration effects such as a “Hubble bubble”. We have used our simulation results to forecast the accuracy with which future galaxy surveys will use the BAO in isolation to constrain the scale of the acoustic oscillations, and under certain assumptions,  $w$ . We anticipate that Pan-STARRS, with accurate photometric redshifts, will have an accuracy comparable to that expected for the next generation of spectroscopic survey (WF MOS) and could potentially reduce the statistical errors on the value of  $w$  by a factor of 2 compared with the current constraints. However, the target of 1% random errors on  $w$  using BAO measurements is beyond the grasp of any of the surveys likely to be completed or even to start within the next decade.

The predictions we have presented here are idealized in a number of respects. The

accuracy with which we expect the dark energy equation of state parameter will be measured assumes that the values of the other cosmological parameters are known with infinite accuracy. We have also neglected the impact of the survey window function on the power spectrum measurement; this will be particularly important in the case of surveys which rely on photometric redshifts. In future work, we plan a number of improvements: i) Use of an even larger simulation volume, to exceed that proposed in forthcoming surveys. One caveat on our quoted error on  $w$  is that some of the planned surveys will be larger than the volume of the BASICC, and will consequently have smaller sampling fluctuations. ii) The inclusion of the evolution of clustering along the line of sight. Although we have focused on  $z = 1$ , proposed surveys will span a broad redshift interval centred on this value. iii) The inclusion of a survey window function, mimicking the angular and radial selections, and including the impact of errors on photometric redshifts. Such calculations represent huge challenges in computational cosmology, due to the volume coverage and mass resolution required in the N-body simulations used, and the post-processing needed to include galaxies. However, such calculations are essential if the BAO approach is to be used to its full potential.



## Chapter 3

# *Implications of Photometric Redshift Errors for BAO Detection*

In this section we investigate the impact of using photometric redshifts on the accuracy with which the baryonic acoustic oscillation (BAO) scale can be measured from the power spectrum of galaxy clustering. Our aim here is to provide a simple quantification of the factor by which the effective volume of a survey is reduced when photometric redshifts are used in place of spectroscopic redshifts.

The photometric redshift technique allows large solid angles of sky to be covered to depths exceeding those accessible spectroscopically at a low observational cost. However, the inaccurate determination of a galaxy's redshift results in an uncertainty in its position and this leads to a distortion in the pattern of galaxy clustering. We shall refer to a measurement of the power spectrum which uses photometric redshifts to assign radial positions as being in “photo- $z$ ” space.

The errors introduced by photometric redshifts can be modelled as random perturbations to the radial positions of galaxies. If we assume that they are Gaussian distributed with mean equal to the true redshift and width  $\sigma_z \approx \Delta z$ , then the Fourier transform of the measured density field,  $\delta_{\text{pz}}(\vec{k})$ , can be written as

$$\delta_{\text{pz}}(\vec{k}) = \delta_z(\vec{k}) \exp(-0.5k_z^2 \sigma_z^2), \quad (3.1)$$

where  $k_z = \vec{k} \cdot \hat{z}$  and  $\delta_z(\vec{k})$  is the density field measured in redshift space. From this expression, the spherically averaged power spectrum can be approximately<sup>1</sup> written as:

---

<sup>1</sup>It is an approximate expression since the redshift-space distortions and photometric redshift errors do not commute under a spherical average (see Peacock and Dodds, 1994).

$$P_{pz}(k) = P_z(k) \frac{\sqrt{\pi}}{2} \frac{\text{Erf}(k\sigma_z)}{k\sigma_z}, \quad (3.2)$$

where Erf is the error function defined as  $\text{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$ . In addition, the power spectrum in photo- $z$  space can be seen as that in redshift space with additional damping on small scales due to the large value of  $\sigma_z$ . On very large scales the main contribution to the power spectrum comes from modes with wavelengths larger than the typical size of the photometric redshift errors. Therefore, the clustering on these scales is essentially unaffected. On the contrary, on scales comparable to and smaller than the photo- $z$  errors, structures are smeared out along the line-of-sight. The modes describing these scales along the line-of-sight contain little information about the true distribution of galaxies and contribute only noise to the power spectrum.

We investigate these effects directly on the measurement of the matter power spectrum using large N-body simulations. We use the L-BASICC ensemble of Angulo et al. (2008), which consists of 50 low-resolution, large volume simulations. Each has a volume of  $2.4(\text{Gpc}/h)^3$  and resolves halos more massive than  $1 \times 10^{13} M_\odot/h$ . The assumed cosmological parameters are  $\Omega_m = 0.25$ ,  $\Omega_\Lambda = 0.75$ ,  $h = 0.73$ ,  $n = 1$  and  $\sigma_8 = 0.9$ . Their huge volume makes the L-BASICC simulations ideal to study the detectability of BAO in future surveys. Photometric redshift errors are mimicked as a random perturbation added to the particles' position along one direction (line-of-sight). The perturbations are drawn from a Gaussian distribution with various widths representing different degrees of uncertainty in the photometric redshift.

In the upper panels of Fig. 3.1 we show the mean, spherically averaged power spectrum of the dark matter measured from the L-BASICC simulations at  $z = 0.5$ , along with its variance, in photo- $z$  space (solid blue lines). The size of the photo- $z$  errors are  $\sigma_z = 0.01$  and  $\sigma_z = 0.04$  (equivalent to 15.8 and 63.4  $h^{-1}\text{Mpc}$  at  $z = 0.5$ ) in the left- and right-hand panels respectively. We have also plotted the power spectrum measured in redshift space (solid red lines) and the analytical expression of Eq. 3.2 (dashed red line). By comparing the spectra in redshift and photo- $z$  spaces, the additional damping described above is evident. Also, we see that Eq. 3.2 describes quantitatively this extra damping on scales where the power spectrum is not shot-noise dominated.

In the lower panels of Fig. 3.1 we take a closer look at the BAO by isolating them from the large-scale shape of the power spectrum. We do this by dividing the power spectrum by a smoothed version of the measurement. It is clear that since the number of “noisy modes” increases with the size of the photometric redshift errors, the error on

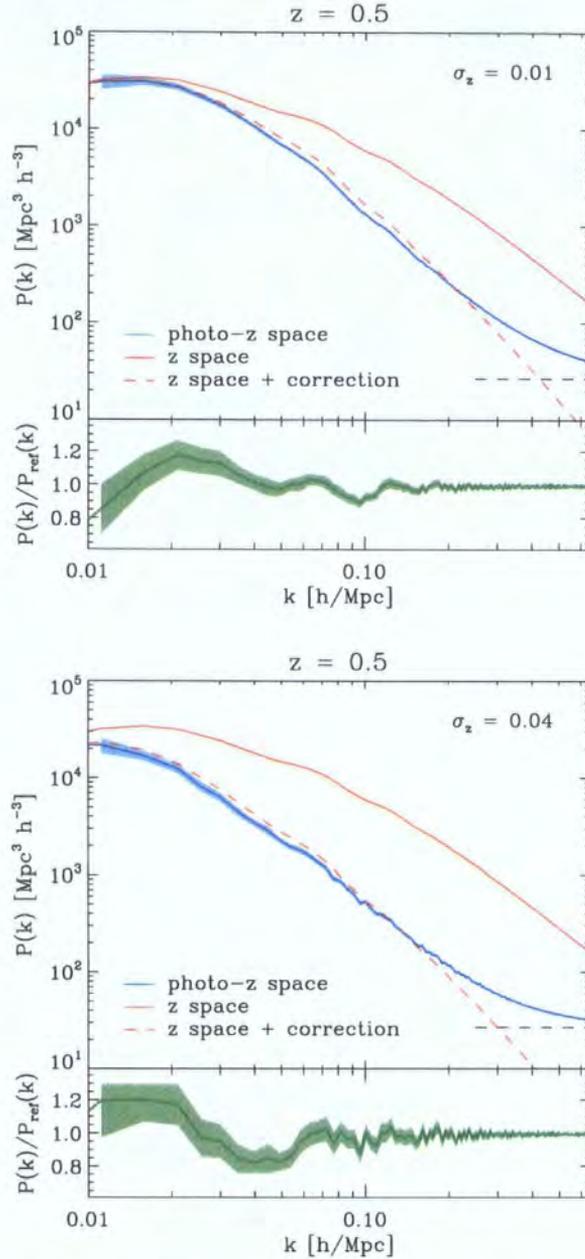


Figure 3.1: The mean and standard deviation of the dark matter power spectrum averaged over an ensemble of 50 N-body simulations at  $z = 0.5$ . The top-panels display the power spectrum in three different cases: (i) redshift space (solid red line), (ii) photo- $z$  space (blue line) in which the position of each dark matter particle has been perturbed to mimic the effect of photometric redshift errors, and (iii) the photo- $z$  space power spectrum derived from Eq. 3.2 and the measured redshift-space power spectrum (red dashed lines). The horizontal dashed line illustrates the shot-noise level. In the bottom panels we plot the photo- $z$  power spectrum divided by a smooth reference spectrum. This reveals the impact of photometric redshift errors directly on the baryonic acoustic oscillations (BAO).

the power spectrum and therefore on the BAO also increases. The visibility of the higher harmonic BAO is also reduced as the photometric redshift error increases. In order to quantify the loss of information, we have followed a standard technique to measure BAO as described in Angulo et al. (2008) (see also Percival et al. 2007 and Sanchez et al. 2008). The method basically consists of dividing the measured power spectrum by a smoothed version of the measurement. In this way, any long wavelength gradient or distortion in the shape of the power spectrum is removed which diminishes the impact of possible systematic errors due to redshift-space distortions, nonlinear evolution and, in the case described in this chapter, photometric-redshift distortions. Then, we construct a model ratio using linear perturbation theory,  $P_t/P_{\text{smooth}}$ , which we fit to the measured ratios. In the fitting procedure there are two free parameters: (i) a damping factor to account for the destruction of BAO peaks located at high  $k$  by non-linear effects and redshift-space distortions and (ii) a stretch factor,  $\alpha$ , which quantifies how accurately we can measure the BAO wavelength. The latter gives a simple estimate of how well we can constrain the dark energy equation of state from BAO measurements alone.

Fig. 3.2 shows the results of applying our fitting procedure to the L-BASICC ensemble at different redshifts. On the  $x$ -axis we plot the size of the photometric redshift error divided by  $(1+z)$ , whilst on the  $y$ -axis we plot the predicted error on  $\alpha$  divided by the error we infer for an ideal spectroscopic survey (i.e. from the power spectrum in redshift space). Since the error on  $\alpha$  scales with the error on the power spectrum and the latter is proportional to the square root of the volume of the survey, the square of the  $y$ -axis should be roughly equal to the factor by which the volume of a photometric redshift needs to be larger than the volume of a spectroscopic survey to achieve the same accuracy.

Several authors have investigated the implications of photometric redshift errors on the clustering measurements in general and on the BAO in particular (Seo and Eisenstein, 2003; Amendola et al., 2005; Dolney et al., 2006; Blake and Bridle, 2005). Our analysis improves upon these studies in several ways: (i) we have included photometric redshift errors directly into an realistic distribution of objects; (ii) by using N-body simulations, our calculation takes into account the effects introduced by nonlinear evolution, nonlinear redshift-space distortions and shot noise; (iii) the use of 50 different simulations enables a robust and realistic estimation of the errors on the power spectrum measurements.

These improvements lead to predictions that are somewhat different from previous ones. For example, for  $\Delta z = 0.03$ , Blake and Bridle (2005) predict a factor of  $\sim 10$  for the reduction of the effective volume of a photometric survey. Here, as shown in

Fig. 3.2, we find a reduction which is a factor 2 times smaller than this (i.e. a volume reduction factor of  $\sim 5$ ). The main difference between our analyses is that Blake and Bridle (2005) use only modes larger than  $k_{\max} = 2/\sigma_z$ , arguing that wavelengths shorter than the size of the photometric redshift errors contribute only noise. In reality, there is a smooth transition around  $k_{\max}$ , with signal coming from all wavenumbers (with different weighting, of course). In addition, the neglect of nonlinear evolution (which erases the BAO at high wavenumbers) also contribute to Blake and Bridle (2005) overestimating the reduction in effective volume. These two effects together could explain the disagreement between our results and those found by Blake and Bridle (2005).

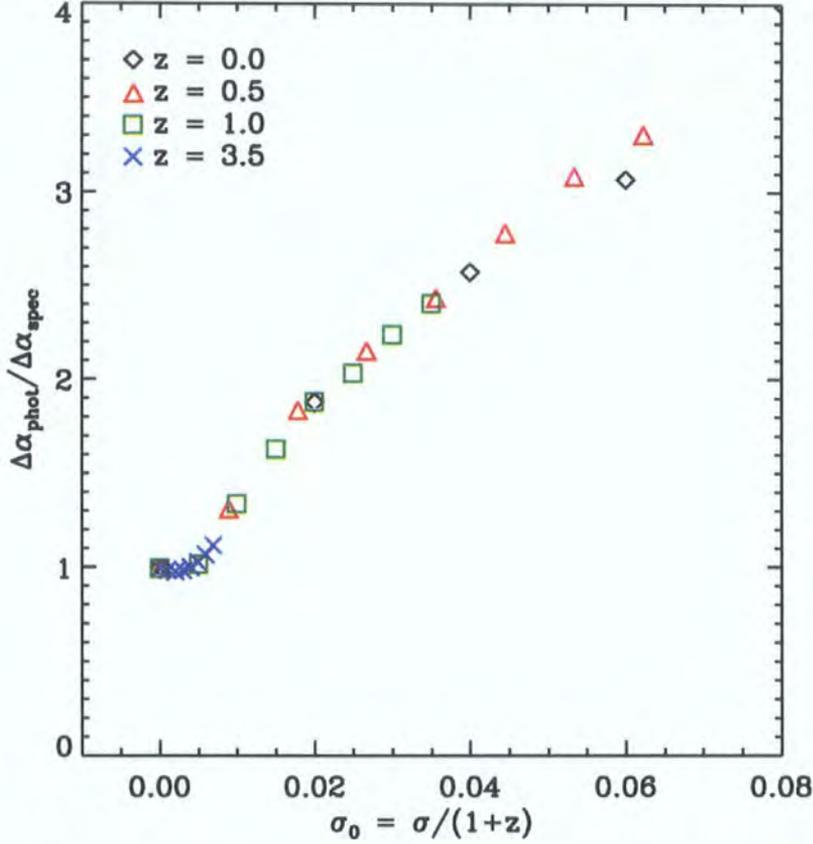


Figure 3.2: The ratio of the error on the measurement of the BAO scale in photo- $z$  space to that in redshift space (i.e. from a perfect spectroscopic redshift) as a function of the magnitude of the photometric redshift error. Assuming that the error on the measurement scales with the square root of the volume, then the  $y$ -axis gives the square root of the ratio of volumes of photometric to spectroscopic surveys which achieve the same accuracy in the measurement of the BAO scale. Note that this quantity is independent of the redshift at which the measurement is made, i.e. it is independent of the degree of nonlinearity present in the dark matter distribution.

# Chapter 4

## *The assembly bias of dark matter haloes to higher orders*

### 4.1 Introduction

The spatial distribution of dark matter haloes is not as simple as was once suspected. In the standard theoretical model for the abundance and distribution of haloes, the clustering strength of haloes is predicted to be a function of mass alone, with more massive haloes displaying stronger clustering (e.g. Kaiser 1984; Cole and Kaiser 1989; Mo and White 1996). However, recent numerical simulations of hierarchical cosmologies, by covering larger volumes with ever improving mass resolution, have been able to reveal subtle dependences of halo clustering on other properties such as formation redshift, the internal structure of the halo and its spin (Gao, Springel, and White 2005; Wechsler et al. 2006; Harker et al. 2006; Bett et al. 2007; Wetzel et al. 2007; Jing, Suto, and Mo 2007; Espino-Briones et al. 2007).

The dependence of halo clustering on a second parameter in addition to mass is generally referred to as assembly bias. However, the nature of the trend in clustering strength recovered depends upon the choice of property used to classify haloes of a given mass. Early simulation work failed to uncover a convincing assembly bias signal, as a result of insufficient volume and mass resolution, which meant that halo clustering could be measured for only a narrow range of mass and with limited statistics (Lemson and Kauffmann 1999; Percival et al. 2003; Sheth and Tormen 2004). The first clear indication of a dependence of halo clustering on a second property was uncovered by Gao, Springel, and White (2005). These authors reported that low mass haloes which form early are more clustered than haloes of the same mass which form later on. No effect was seen for massive haloes. Wechsler et al. (2006) were able to confirm this result but also found

that halo clustering depends on the density profile of the halo, as characterized by the concentration parameter (Navarro, Frenk, and White, 1997). The sense of the dependence of clustering strength on concentration changes with mass. Wechsler et al. found that massive haloes showed a dependence of clustering strength on concentration, with low concentration haloes being the more strongly clustered (as confirmed by Gao and White 2007, Jing, Suto, and Mo 2007 and Wetzel et al. 2007). This trend of clustering strength with concentration is reversed for low mass haloes. Although formation time and concentration are correlated (e.g. Neto et al. 2007), their impact on the clustering of haloes does not follow trivially from this correlation, suggesting that some other parameter may be more fundamental (as argued by Croton, Gao, and White 2007).

Previous studies of assembly bias have focused exclusively on the linear bias parameter, which relates the two-point correlations of haloes and dark matter. Measurements from local surveys have shown that galaxies have significant higher order correlation functions and that the spatial distribution of galaxies and haloes is not fully described by two-point statistics (e.g. Baugh et al. 2004; Croton et al. 2004; Nichol et al. 2006; Frith et al. 2006). With large surveys planned at higher redshifts, there is a clear need for accurate models of the higher order clustering of dark matter haloes, and to establish whether or not the higher order bias parameters depend on other properties in addition to mass.

In this chapter we measure the higher order bias parameters of dark matter haloes using a simulation which covers a volume more than an order of magnitude larger than the run analyzed by Gao and collaborators. We use a novel approach to estimate the higher order correlation functions of dark matter haloes. Our method builds upon the cross-correlation technique advocated for two-point correlations by Jing, Suto, and Mo (2007), Gao and White (2007) and Smith et al. (2007). By considering fluctuations in the density of haloes and dark matter within the same smoothing window, we can suppress discreteness noise in our measurements. This improved clustering estimator, which uses the counts-in-cells method, when coupled with the large volume of our simulation, allows us to recover the bias parameters from linear to fourth order, and to study the dependence of these parameters on the halo concentration.

In Section 4.2, we give the theoretical background to the counts-in-cells technique we use to estimate higher order clustering and explain how the clustering of haloes relates to the underlying dark matter at different orders. We also introduce the numerical simulations in that section. We present our results in Section 4.3 and a summary and discussion in Section 4.4.

## 4.2 Theoretical background and method

In this Section we give the theoretical background to the measurements presented in Section 4.3. We estimate the clustering of haloes and dark matter using a counts-in-cells approach. An overview of this method is given in §4.2.1, in which we explain how to obtain expressions for the higher order auto correlation functions of a density field from the moments of the distribution of counts-in-cells. We also introduce the concept of higher order *cross* correlation functions, which combine fluctuations in two density fields. The concept of hierarchical amplitudes, scaling relations between higher order correlation functions and the two-point correlation function, is introduced in §4.2.3. The key theoretical results relating the higher order cross correlation functions of haloes to the two-point function and hierarchical amplitudes of the dark matter are given in §4.2.2. The simulations we use to measure the clustering of dark matter haloes are described in §4.2.4.

### 4.2.1 The counts in cells approach to measuring clustering

Here we give a brief overview of the approach of using the distribution of counts in cells to estimate the higher order auto correlation functions of a set of objects. An excellent and comprehensive review of this material is given by Bernardeau et al. (2002). We first discuss the higher order correlation functions for the case of a continuous, unsmoothed density field, then introduce the concept of cross-correlations (§4.2.1.1), before explaining how these results are changed in the case of a smoothed distribution of discrete points (§4.2.1.2).

#### 4.2.1.1 Higher order correlations: unsmoothed and continuous density field

In general, the complete hierarchy of  $N$ -point correlation functions is required to fully characterize the spatial distribution of fluctuations in a density field. An exception to this occurs for the special case of a Gaussian density field, which can be described completely by its two-point correlation function.

The  $N$ -point correlation functions are usually written in terms of the dimensionless density fluctuation or density contrast at a point:

$$\delta(x) = \rho(x)/\langle\rho\rangle - 1, \quad (4.1)$$

where  $\langle\rho\rangle$  is the mean density; the average is taken over different spatial locations. By

definition,  $\langle \delta(x) \rangle = 0$  when the average is taken over a fair sample of the density field. The  $N^{\text{th}}$  order moment of the density field, sometimes referred to as a central moment, because  $\delta$  is a fractional fluctuation around the mean density, is given by:

$$\mu_N = \langle \delta(x_1), \dots, \delta(x_n) \rangle, \quad (4.2)$$

where, in general, the density fluctuations are correlated at different spatial locations.

The  $N^{\text{th}}$  order central moments defined in Eq. 4.2 can be decomposed into terms which include products of lower order moments. This is because there are different permutations of how the  $N$ -points can be “connected” or joined together. This idea is illustrated nicely by tree diagrams in the review by Bernardeau et al. (2002). The terms into which the central moments are broken down are called connected moments and these cannot be reduced further. In the tree diagram language, an  $N$ -point connected moment has no disjoint points; all  $N$ -points are linked to one another when the spatial averaging is performed. The distinction between connected and unconnected moments may become clearer if we write down the decomposition of the unconnected central moments up to fifth order:

$$\langle \delta^2 \rangle = \langle \delta^2 \rangle_c + \langle \delta \rangle_c^2 \quad (4.3)$$

$$\langle \delta^3 \rangle = \langle \delta^3 \rangle_c + 3\langle \delta^2 \rangle_c \langle \delta \rangle_c + \langle \delta \rangle_c^3 \quad (4.4)$$

$$\langle \delta^4 \rangle = \langle \delta^4 \rangle_c + 4\langle \delta^3 \rangle_c \langle \delta \rangle_c + 3\langle \delta^2 \rangle_c^2 + 6\langle \delta^2 \rangle_c \langle \delta \rangle_c^2 + \langle \delta \rangle_c^4 \quad (4.5)$$

$$\begin{aligned} \langle \delta^5 \rangle &= \langle \delta^5 \rangle_c + 5\langle \delta^4 \rangle_c \langle \delta \rangle_c + 10\langle \delta^3 \rangle_c \langle \delta^2 \rangle_c + 10\langle \delta^3 \rangle_c \langle \delta \rangle_c^2 + 15\langle \delta^2 \rangle_c^2 \langle \delta \rangle_c \\ &+ 10\langle \delta^2 \rangle_c \langle \delta \rangle_c^3 + \langle \delta \rangle_c^5, \end{aligned} \quad (4.6)$$

where the subscript  $c$  outside the angular brackets denotes a connected moment. Remembering that  $\langle \delta \rangle = 0$ , these equations simplify to:

$$\langle \delta^2 \rangle = \langle \delta^2 \rangle_c \quad (4.7)$$

$$\langle \delta^3 \rangle = \langle \delta^3 \rangle_c \quad (4.8)$$

$$\langle \delta^4 \rangle = \langle \delta^4 \rangle_c + 3\langle \delta^2 \rangle_c^2 \quad (4.9)$$

$$\langle \delta^5 \rangle = \langle \delta^5 \rangle_c + 10\langle \delta^3 \rangle_c \langle \delta^2 \rangle_c. \quad (4.10)$$

Hence, for the second and third order moments, there is no difference in practice between the connected and unconnected moments.

The  $N$ -point auto-correlation functions,  $\xi_N$ , are written in terms of the connected moments:

$$\xi_N(x_1, \dots, x_N) = \langle \delta(x_1), \dots, \delta(x_N) \rangle_c. \quad (4.11)$$

By analogy with the  $N$ -point auto correlation functions of fluctuations in a single density field, we can define the  $i + j$ -point cross correlation function of two, co-spatial density fields, with respective density contrasts given by  $\delta_1$  and  $\delta_2$ :

$$\xi_{i,j}(x_1, \dots, x_i \ ; \ y_1, \dots, y_j) = \langle \delta_1(x_1), \dots, \delta_1(x_i) \delta_2(y_1), \dots, \delta_2(y_j) \rangle_c. \quad (4.12)$$

In the application in this chapter, the first index will refer to the distribution of dark matter haloes and the second index to the dark matter. When the density contrasts are evaluated at the same spatial location, i.e.  $x_1 = \dots = x_i = y_1 = \dots = y_j = 0$ , the connected moments  $\xi_{i,j}$  are called cumulants of the joint probability distribution function of  $\delta_1$  and  $\delta_2$  (and are sometimes denoted as  $k_{i,j}$ ).

To generate expressions for the higher order correlation functions of the cross-correlated density fluctuations,  $\xi_{i,j}$ , we will use the method of generating functions (see §4.3.3.3 of Bernardeau et al. 2002). A moment generating function is defined for the central moments ( $\mu_{i,j}$ ) as a power series in  $\delta_1$  and  $\delta_2$ , which can be written as  $\chi \equiv \langle \exp(\delta_1 t_1 + \delta_2 t_2) \rangle$ , where  $t_1$  and  $t_2$  are random variables. This moment generating function can be related to the cumulant generating function ( $\psi$ ) for the cumulants by (see Bernardeau et al. 2002 for a proof):

$$\psi(t_1, t_2) \equiv \ln \chi(t_1, t_2). \quad (4.13)$$

Then, by taking partial derivatives of  $\psi$  and  $\chi$  evaluated at  $t_1 = t_2 = 0$ , one can “generate” the cumulants and moments:

$$\xi_{i,j}(0) = k_{i,j} = \frac{\partial^{i+j}}{\partial t_1^i \partial t_2^j} \psi |_{t_1=t_2=0} \quad (4.14)$$

$$\mu_{i,j} = \frac{\partial^{i+j}}{\partial t_1^i \partial t_2^j} \chi |_{t_1=t_2=0} = \langle \delta_1^i \delta_2^j \rangle. \quad (4.15)$$

Following this method we can obtain expressions for the cross-correlation cumulants up to order  $i + j = 5$ , grouping terms of the same order:

$$k_{1,1} = \mu_{1,1} \quad (4.16)$$

$$k_{2,0} = \mu_{2,0} \quad (4.17)$$

$$k_{3,0} = \mu_{3,0} \quad (4.18)$$

$$k_{2,1} = \mu_{2,1} \quad (4.19)$$

$$k_{4,0} = \mu_{4,0} - 3 \mu_{2,0}^2 \quad (4.20)$$

$$k_{3,1} = \mu_{3,1} - 3 \mu_{2,0} \mu_{1,1} \quad (4.21)$$

$$k_{2,2} = \mu_{2,2} - \mu_{2,0} \mu_{0,2} - 2 \mu_{1,1}^2 \quad (4.22)$$

$$k_{5,0} = \mu_{5,0} - 10 \mu_{3,0} \mu_{2,0} \quad (4.23)$$

$$k_{4,1} = \mu_{4,1} - 4 \mu_{3,0} \mu_{1,1} - 6 \mu_{2,0} \mu_{2,1} \quad (4.24)$$

$$k_{3,2} = \mu_{3,2} - \mu_{3,0} \mu_{0,2} - 6 \mu_{2,1} \mu_{1,1} - 3 \mu_{2,0} \mu_{1,2}. \quad (4.25)$$

Note that these results are symmetric with respect to exchanging the indexes and that we have used the fact that  $\mu_{1,0} = \mu_{0,1} = 0$ , since, by construction  $\langle \delta_1 \rangle = \langle \delta_2 \rangle = 0$ .

#### 4.2.1.2 Higher order correlations: smoothed and discrete density fields

Sadly, density fluctuations at a point are of little practical use as they cannot be measured reliably, as typically we have a finite number of tracers of the density field, i.e. galaxies in a survey or dark matter particles in an N-body simulation, and so have a limited resolution view of the density field. Furthermore, estimating the  $N$ -point correlations for a modern survey or simulation is time consuming and short cuts are often taken, such as restricting the number of configurations of points sampled. To overcome both of these problems, moments of the smoothed density field can be computed instead of the point moments.

The smoothed density contrast,  $\delta_R$ , is a convolution of the density contrast at a point with the smoothing window,  $W_R$ , which has volume  $V$ :

$$\delta(x)_R = \frac{1}{V} \int dx' \delta(x) W_R(x - x'). \quad (4.26)$$

Typically, the smoothing window is a spherical top-hat in which case  $W_R = 1$  for all points within distance  $R$  from the centre of the window and  $W_R = 0$  otherwise. After smoothing, the cumulants correspond to the  $i + j$ -point volume-averaged cross correlation functions:

$$\bar{\xi}_{i,j}(R) \equiv \int d^3x_1 \dots d^3x_i d^3y_1 \dots d^3y_j \frac{\xi_{i,j}(x_1, \dots, x_i, y_1, \dots, y_j)}{W_R(x_1) \dots W_R(x_i) W_R(y_1) \dots W_R(y_j)} \quad (4.27)$$

Eqs. 4.16-4.25 are still valid, with the cumulants replaced by volume-averaged cumulants.

Another issue introduced by the discreteness of the density field is the contribution of Poisson noise to the measurements of the cumulants. To take this into account, we can modify the moment generating function as follows (Peebles 1980):

$$\chi(t_1, t_2) = \langle \exp(f_1(t_1) + f_2(t_2)) \rangle, \quad (4.28)$$

$$f_1 = (\exp(t_1) - t_1 - 1) \bar{n}_1 + (\exp(t_1) - 1) \delta_1, \quad (4.29)$$

$$f_2 = (\exp(t_2) - t_2 - 1) \bar{n}_2 + (\exp(t_2) - 1) \delta_2. \quad (4.30)$$

Here,  $\bar{n}_1$  and  $\bar{n}_2$  are the mean number of objects in density field 1 and density field 2 respectively within spheres of radius  $R$ . Using this modified generating function, and defining

$$\mu'_{i,j} = \langle (n_1 - \bar{n}_1)^i (n_2 - \bar{n}_2)^j \rangle,$$

we obtain the following relations between the volume-averaged, connected  $i + j$ -point cross correlation functions,  $\bar{\xi}_{i,j}$ , and the central moments,  $\mu_{i,j}$ :

$$\bar{n}_1^2 \bar{\xi}_{2,0} = \mu'_{2,0} - \bar{n}_1 \quad (4.31)$$

$$\bar{n}_1 \bar{n}_2 \bar{\xi}_{1,1} = \mu'_{1,1} \quad (4.32)$$

$$\bar{n}_2^2 \bar{\xi}_{0,2} = \mu'_{0,2} - \bar{n}_2 \quad (4.33)$$

$$\bar{n}_1^3 \bar{\xi}_{3,0} = \mu'_{3,0} + 2\bar{n}_1 - 3\mu'_{2,0} \quad (4.34)$$

$$\bar{n}_1^2 \bar{n}_2 \bar{\xi}_{2,1} = \mu'_{2,1} - \mu'_{1,1} \quad (4.35)$$

$$\bar{n}_1 \bar{n}_2^2 \bar{\xi}_{1,2} = \mu'_{1,2} - \mu'_{1,1} \quad (4.36)$$

$$\bar{n}_2^3 \bar{\xi}_{0,3} = \mu'_{0,3} + 2\bar{n}_2 - 3\mu'_{0,2} \quad (4.37)$$

$$\bar{n}_1^4 \bar{\xi}_{4,0} = \mu'_{4,0} - 6\bar{n}_1 + 11\mu'_{2,0} - 6\mu'_{3,0} - 3\mu'_{2,0}{}^2 \quad (4.38)$$

$$\bar{n}_1^3 \bar{n}_2 \bar{\xi}_{3,1} = \mu'_{3,1} + 2\mu'_{1,1} - 3\mu'_{2,1} - 3\mu'_{1,1}\mu'_{2,0} \quad (4.39)$$

$$\bar{n}_1^2 \bar{n}_2^2 \bar{\xi}_{2,2} = \mu'_{2,2} - \mu'_{1,2} - \mu'_{2,1} + \mu'_{1,1} - \mu'_{2,0}\mu'_{0,2} - 2\mu'_{1,1}{}^2 \quad (4.40)$$

$$\bar{n}_1 \bar{n}_2^3 \bar{\xi}_{1,3} = \mu'_{1,3} + 2\mu'_{1,1} - 3\mu'_{1,2} - 3\mu'_{1,1}\mu'_{0,2} \quad (4.41)$$

$$\bar{n}_2^4 \bar{\xi}_{0,4} = \mu'_{0,4} - 6\bar{n}_2 + 11\mu'_{0,2} - 6\mu'_{0,3} - 3\mu'_{0,2}{}^2 \quad (4.42)$$

Note that these expressions revert to those in the literature for autocorrelation moments in the case of either  $i$  or  $j$  equal to zero (see for instance Baugh et al. 1995). Also note that in the limit  $\bar{n}_1 \rightarrow \infty$ ,  $\bar{n}_2 \rightarrow \infty$ , they correspond to the expressions given by Eqs. 4.16-4.25.

#### 4.2.2 Hierarchical amplitudes

At this point it is useful to define quantities called hierarchical amplitudes which are the ratio between the  $N$ -point, volume-averaged connected moments and the two-point volume-averaged connected moment raised to the  $N - 1$  power:

$$S_N \equiv \frac{\bar{\xi}_N}{\bar{\xi}_2^{N-1}}. \quad (4.43)$$

This form is motivated by the expected properties of a Gaussian field which evolves due to gravitational instability (Bernardeau et al. 2002). In the case of small amplitude fluctuations, i.e. on smoothing scales for which  $\bar{\xi}_2(R) \ll 1$ , the  $S_N$  depend only on the local slope of the linear perturbation theory power spectrum of density fluctuations and are independent of time (Juszkiewicz, Bouchet, and Colombi 1993; see Bernardeau 1994 for

expressions for the  $S_N$ ). Similar scalings, but with different values for the  $S_N$ , apply in the case of distributions of particles which have not arisen through gravitational instability, e.g. particles displaced according to Zel'dovich approximation (see Juszkiewicz, Bouchet, and Colombi 1993).

In the case of a Gaussian density field, all of the  $S_N$  are equal to zero. Initially, as perturbations grow through gravitational instability, the two-point connected moment increases. The distribution of fluctuations soon starts to deviate from a Gaussian, particularly as voids grow in size and cells become empty ( $\delta \rightarrow -1$ ). Voids evolve more slowly than overdense regions. There is in principle no limit on how overdense a cell can become. As a result, the distribution of overdensities becomes asymmetrical or skewed, with the peak of the distribution moving to negative density contrasts and a long tail developing to high density contrasts. To first order, this deviation from symmetry is quantified by the value of  $S_3$ , which is often referred to as the skewness of the density field. Higher order moments and hierarchical amplitudes probe progressively further out into the tails of the distribution of density contrasts.

#### 4.2.3 Higher order correlations: biased tracers

We are now in a position to consider the cross-correlation functions for the case of relevance in this chapter, when the set of objects making up one of the density fields is local function of the second density field; the first density field is a biased tracer of the second. In our application, one density field is defined by the spatial distribution of dark matter haloes and the other by the dark matter. In the case of a local bias and small perturbations, the density contrast in the biased tracers ( $\delta_1$ ) can be written as an expansion in terms of the underlying dark matter density contrast ( $\delta_2$ ), as proposed by Fry & Gaztanaga (1993):

$$\delta_1(R) = \sum_{k=0}^{\infty} \frac{b_k}{k!} \delta_2^k(R), \quad (4.44)$$

where the  $b_k$  are known as bias coefficients<sup>1</sup>;  $b_1$  is the linear bias commonly discussed in relation to two point correlations. Note that, by construction, we require that  $\langle \delta \rangle = 0$ , which implies  $b_0 = -\sum_{k=2}^{\infty} b_k \langle \delta_2^k \rangle / k!$ . The  $b_k$ , as we shall see later, depend on mass but this is suppressed in our notation.

Using this bias prescription, and following the treatment Fry and Gaztanaga (1993)

---

<sup>1</sup>See Dekel and Lahav (1999) for a discussion of the relation between stochastic and deterministic bias.

used for autocorrelations, we can write the volume-averaged cross-correlation functions of dark matter haloes in terms of the two-point volume averaged correlation function ( $\bar{\xi}_{0,2}$ ) and hierarchical amplitudes of the dark matter,  $S_N$ :

$$\bar{\xi}_{1,1} = b_1 \bar{\xi}_{0,2} + O(\bar{\xi}_{0,2}^2) \quad (4.45)$$

$$\bar{\xi}_{2,0} = b_1^2 \bar{\xi}_{0,2} + O(\bar{\xi}_{0,2}^2) \quad (4.46)$$

$$\bar{\xi}_{1,2} = b_1 \bar{\xi}_{0,2}^2 (c_2 + S_3) + O(\bar{\xi}_{0,2}^3) \quad (4.47)$$

$$\bar{\xi}_{2,1} = b_1^2 \bar{\xi}_{0,2}^2 (2c_2 + S_3) + O(\bar{\xi}_{0,2}^3) \quad (4.48)$$

$$\bar{\xi}_{3,0} = b_1^3 \bar{\xi}_{0,2}^2 (3c_2 + S_3) + O(\bar{\xi}_{0,2}^3) \quad (4.49)$$

$$\bar{\xi}_{1,3} = b_1 \bar{\xi}_{0,2}^3 (3S_3 c_2 + S_4 + c_3) + O(\bar{\xi}_{0,2}^4) \quad (4.50)$$

$$\bar{\xi}_{2,2} = b_1^2 \bar{\xi}_{0,2}^3 (S_4 + 6S_3 c_2 + 2c_2^2 + 2c_3) + O(\bar{\xi}_{0,2}^4) \quad (4.51)$$

$$\bar{\xi}_{3,1} = b_1^3 \bar{\xi}_{0,2}^3 (6c_2^2 + 9S_3 c_2 + S_4 + 3c_3) + O(\bar{\xi}_{0,2}^4) \quad (4.52)$$

$$\bar{\xi}_{4,0} = b_1^4 \bar{\xi}_{0,2}^3 (12c_2^2 + 12S_3 c_2 + S_4 + 4c_3) + O(\bar{\xi}_{0,2}^4) \quad (4.53)$$

$$\bar{\xi}_{1,4} = b_1 \bar{\xi}_{0,2}^4 (4c_2 S_4 + 6c_3 S_3 + c_4 + S_5 + 3c_2 S_3^2) + O(\bar{\xi}_{0,2}^5) \quad (4.54)$$

$$\begin{aligned} \bar{\xi}_{2,3} = & b_1^2 \bar{\xi}_{0,2}^4 (12S_3 c_3 + 6S_3^2 c_2 + 12S_3 c_2^2 + 6c_2 c_3 + 2c_4 + S_5 + \\ & 8c_2 S_4) + O(\bar{\xi}_{0,2}^5) \end{aligned} \quad (4.55)$$

$$\begin{aligned} \bar{\xi}_{3,2} = & b_1^3 \bar{\xi}_{0,2}^4 (12c_2 S_4 + 18c_3 S_3 + 18c_2 c_3 + 36c_2^2 S_3 + 9c_2 S_3^2 + \\ & S_5 + 6c_2^3 + 3c_4) + O(\bar{\xi}_{0,2}^5) \end{aligned} \quad (4.56)$$

$$\begin{aligned} \bar{\xi}_{4,1} = & b_1^4 \bar{\xi}_{0,2}^4 (4c_4 + 24c_2^3 + S_5 + 72c_2^2 S_3 + 16c_2 S_4 + 36c_2 c_3 + \\ & 24c_3 S_3 + 12c_2 S_3^2) + O(\bar{\xi}_{0,2}^5) \end{aligned} \quad (4.57)$$

$$\begin{aligned} \bar{\xi}_{5,0} = & b_1^5 \bar{\xi}_{0,2}^4 (20c_2 S_4 + 15c_2 S_3^2 + 60c_2^3 + 30c_3 S_3 + 5c_4 + \\ & 120c_2^2 S_3 + S_5 + 60c_2 c_3) + O(\bar{\xi}_{0,2}^5) \end{aligned} \quad (4.58)$$

where  $c_k = b_k/b_1$ . Note it has been shown that these transformations preserve the hierarchical nature of the clustering (Fry and Gaztanaga 1993).

#### 4.2.4 Numerical Simulations

To make accurate measurements of the higher order clustering of dark matter and dark matter haloes, we use the N-body simulations carried out by Angulo et al. (2008). Two simulation specifications were used: i) The **BASICC**, a high-resolution run which used  $1448^3$  particles of mass  $5.49 \times 10^{11} h^{-1} M_{\odot}$  to follow the growth of structure in the dark matter in a periodic box of side  $1340h^{-1}\text{Mpc}$ . ii) The **L-BASICC** ensemble, a suite of 50 lower resolution runs, which used  $448^3$  particles of mass  $1.85 \times 10^{12} h^{-1} M_{\odot}$  in the same box size as the **BASICC**. Each **L-BASICC** run was evolved from a different realization of the initial Gaussian density field. The simulation volume was chosen to allow the growth of fluctuations to be modelled accurately on a wide range of scales, including that of the baryonic acoustic oscillations (the **BASICC** acronym stands for Baryonic Acoustic oscillation Simulations at the Institute for Computational Cosmology). The extremely large volume of each box also makes it possible to extract accurate measurements of the clustering of massive haloes. The superior mass resolution of the **BASICC** run means that it can resolve the haloes which are predicted to host the galaxies expected to be seen in forthcoming galaxy surveys. The **L-BASICC** runs resolve haloes equivalent to group-sized systems. The independence of the **L-BASICC** ensemble runs makes them ideally suited to the assessment of the impact of cosmic variance on our clustering measurements.

In both cases, the same values of the basic cosmological parameters were adopted, which are broadly consistent with recent data from the cosmic microwave background and the power spectrum of galaxy clustering (Sánchez et al., 2006): the matter density parameter,  $\Omega_M = 0.25$ , the vacuum energy density parameter,  $\Omega_{\Lambda} = 0.75$ , the normalization of density fluctuations, expressed in terms of the linear theory amplitude of density fluctuations in spheres of radius  $8h^{-1}\text{Mpc}$  at the present day,  $\sigma_8 = 0.9$ , the primordial spectral index  $n_s = 1$ , the dark energy equation of state,  $w = -1$ , and the Hubble constant,  $h = H_0/(100\text{kms}^{-1}\text{Mpc}^{-1}) = 0.73$ . The simulations were started from realizations of a Gaussian density field set up using the Zel'dovich approximation (Zel'dovich 1970). Particles were perturbed from a glass-like distribution (White 1994; Baugh et al. 1995). The starting redshift for both sets of simulations was  $z = 63$ . The linear perturbation theory power spectrum used to set up the initial density field was generated using the Boltzmann code **CAMB** (Lewis et al. 2000). The initial density field was evolved to the present day using a memory efficient version of **GADGET-2** (Springel 2005).

Outputs of the particle positions and velocities were stored from the simulations at

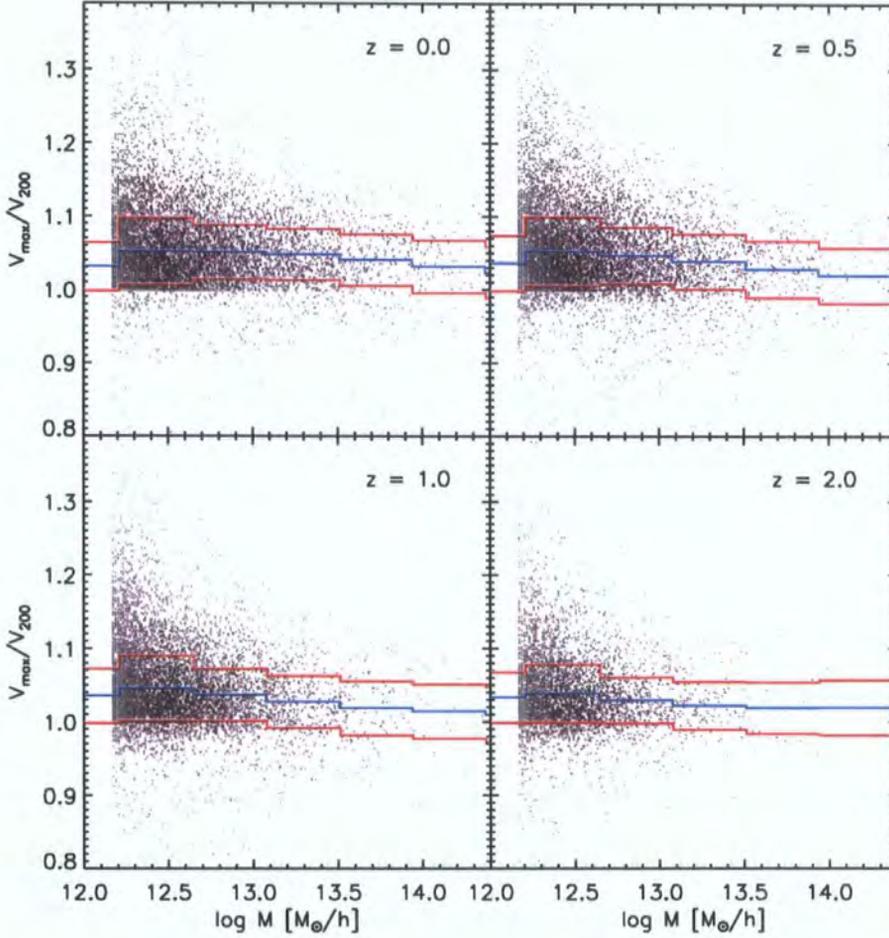


Figure 4.1: The ratio  $V_{\max}/V_{200}$  as a function of halo mass for gravitationally bound haloes in the BASICC simulation, which have a minimum of 26 particles.  $V_{\max}$  is the maximum effective circular velocity of the largest substructure within the halo and  $V_{200}$  is the effective rotation speed at the radius within which the mean density is 200 times the critical density, computed using all of the particles within this radius. Each panel shows the relation at a different redshift as indicated by the legend. The red lines show the 20-80 percentile range of the distribution of  $V_{\max}/V_{200}$  values, and the blue lines show the mean.

selected redshifts. Dark matter haloes were identified using a Friends-of-Friends (FOF) percolation algorithm (Davis et al. 1985) and substructures within these were found using a modified version of **SUBFIND** (Springel et al. 2001). Our default choice is to use the number of particles in a structure as returned by the FOF group finder to set the mass of the halo; at the end of Section 4.3.4 we discuss a variation on this to assess the sensitivity of our results to the group finder. The position of the halo is the position of the most bound particle in the largest substructure, as determined by **SUBFIND**. In this chapter, only gravitationally bound groups with more than 26 particles are considered. The **SUBFIND** algorithm also computes several halo properties such as the circular velocity profile  $V_c(r) = (GM(r)/r)^{1/2}$ ,  $V_{\max}$ , the maximum value of  $V_c$  for the largest substructure, and  $V_{200} = V_c(r_{200})$ , where  $r_{200}$  is the radius of a sphere enclosing a volume of mean density 200 times the critical density. These properties are calculated using only the particles which are bound to the main subhalo of the FOF halo; i.e. ignoring all of the other substructure haloes within the FOF halo. In the best resolved haloes, substructures other than the largest substructure account for at most 15% of the total halo mass (Ghigna et al., 1998). Later on in the chapter we will present results for the clustering of haloes as a function of mass and a second parameter. We have a limited number of output times available to us, so it is not feasible to use the formation time of the halo as the second parameter. Instead, we will use the ratio  $V_{\max}/V_{200}$ . Fig. 4.1 shows  $V_{\max}/V_{200}$  as a function of halo mass at different epochs in the **BASICC** simulation. There is a trend of declining  $V_{\max}/V_{200}$  with increasing halo mass. In cases where the density profile of the dark matter halo matches the universal profile advocated by Navarro et al. (1997),  $V_{\max}/V_{200}$  depends on the concentration parameter which characterizes the profile. Haloes in the extreme parts of the distribution of  $V_{\max}/V_{200}$  also have extreme values of the concentration parameter (Navarro, Frenk, and White 1997). More massive haloes tend to have lower values of the concentration parameter and lower values of the velocity ratio  $V_{\max}/V_{200}$ . The ratio  $V_{\max}/V_{200}$  is easier to extract from the simulation, as it does not require a parametric form to be fitted to the density profile. There is a correlation between formation time and concentration parameter, and hence the ratio  $V_{\max}/V_{200}$ , albeit with scatter (Navarro, Frenk, and White 1997; Zhao et al. 2003).

## 4.3 Results

Our ultimate goal is to measure the higher order bias of dark matter haloes. As described in Section 4.2, we follow a novel approach to do this, employing cross moments between haloes and the dark matter. The first step in this process is to compute the densities of haloes and dark matter on grids of cubical cells of different sizes<sup>2</sup>. A natural by-product of this procedure is the higher order clustering of the dark matter and haloes in terms of the auto-correlation functions. We first present the hierarchical amplitudes estimated for the dark matter (§4.3.1) and haloes (§4.3.2) using the autocorrelation function higher order moments. In §4.3.3 we show the measurements of the cross moments and in §4.3.4 we present the interpretation of these results in terms of the bias parameters.

### 4.3.1 Hierarchical amplitudes for the dark matter

Fig. 4.3.1 shows the hierarchical amplitudes  $S_N$  measured for the dark matter at different redshifts. The upper panels show the results in real space and the lower panels include the effects of redshift-space distortions using the distant observer approximation. The points indicate the median value of the hierarchical amplitudes measured in the L-BASICC ensemble and the error bars indicate the variance in these measurements. The lines show the hierarchical amplitudes predicted by perturbation theory (Bernardeau 1994; Juszkiewicz, Bouchet, and Colombi 1993). At the highest redshift plotted,  $z = 4$ , the agreement between the measurements made from the simulations and the predictions of perturbation theory is impressive, covering scales from  $5h^{-1}\text{Mpc}$  to  $100h^{-1}\text{Mpc}$  for  $S_3$  and  $S_4$ . As redshift decreases, the simulation results for  $S_5$  and  $S_6$  are slightly higher than the perturbation theory predictions. The measurements of  $S_3$  from the simulations continue to agree with the perturbation theory predictions, but over a narrower range of scales. For smoothing scales on which the variance is less than unity, the hierarchical amplitudes are expected to be independent of epoch, depending only on the shape of the linear perturbation theory power spectrum of density fluctuations (Juszkiewicz, Bouchet, and Colombi 1993; Bernardeau 1994; Gaztanaga and Baugh 1995). Fig. 4.3.1 confirms that this is the case. As the density field evolves, the measured hierarchical amplitudes change remarkably little, particularly when one bears in mind that the higher order correlation

<sup>2</sup>Tests show that density fluctuations in cubical cells can be readily translated into counts in spherical cells by simply setting the volume of the spherical cell equal to that of the cube. We use cubical cells for speed. The counts are regridded to improve the measurement of the rare event tails of the count distribution.

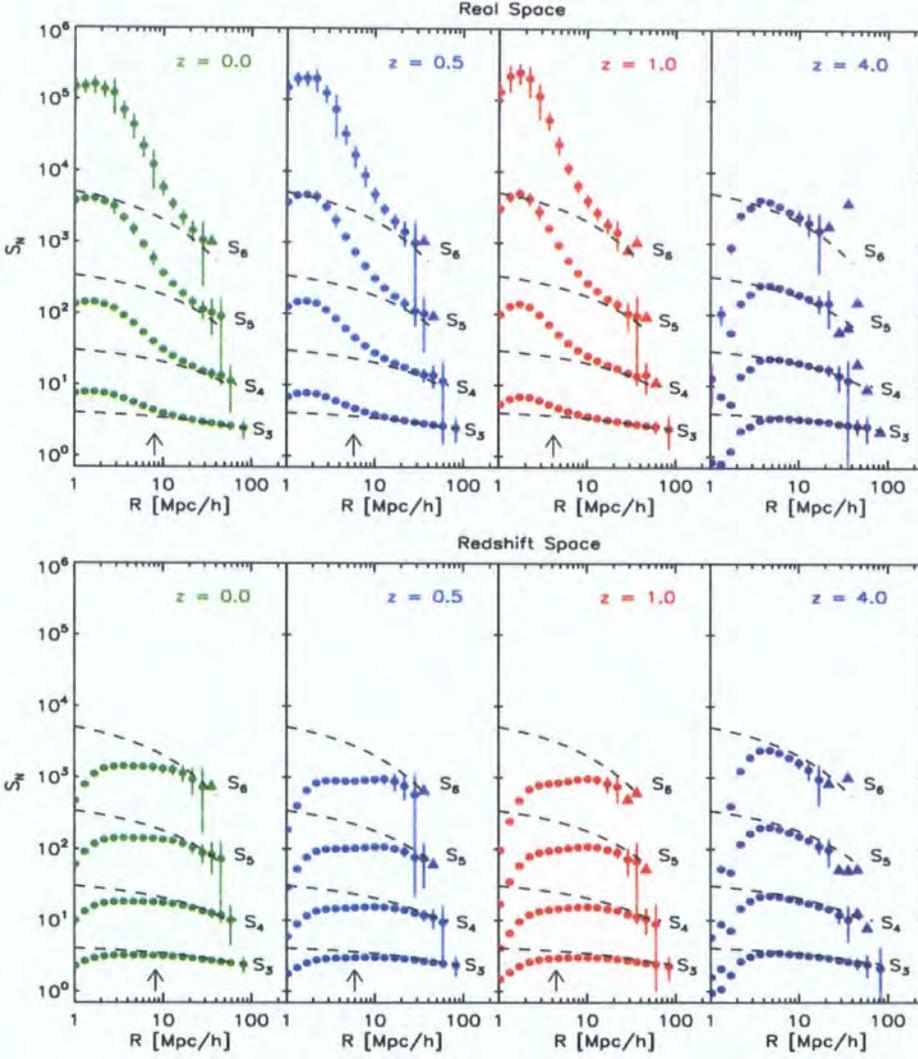


Figure 4.2: The hierarchical amplitudes ( $S_N$ ) measured for the dark matter as a function of the smoothing radius. The points show the amplitude for the  $S_N$  obtained from the L-BASICC ensemble, after taking the ratio of the median correlation functions, as defined by Eq. 4.43. The error bars show the scatter in the measurements over the ensemble, obtained by computing  $S_N$  for each simulation from the ensemble. Error bars are plotted at smoothing scales for which the fractional error is less than unity; triangles show scales on which the error exceeds unity. In both sets of panels, the dashed lines show the predictions of perturbation theory in real space. Note that no correction for shot noise has been applied to the measured amplitudes. The arrows indicate the cell radius for which the variance in the counts in cells for the dark matter is equal to unity, which is roughly the scale down to which perturbation theory should be valid; at  $z = 4$ , this scale is below  $R = 1h^{-1}\text{Mpc}$ .

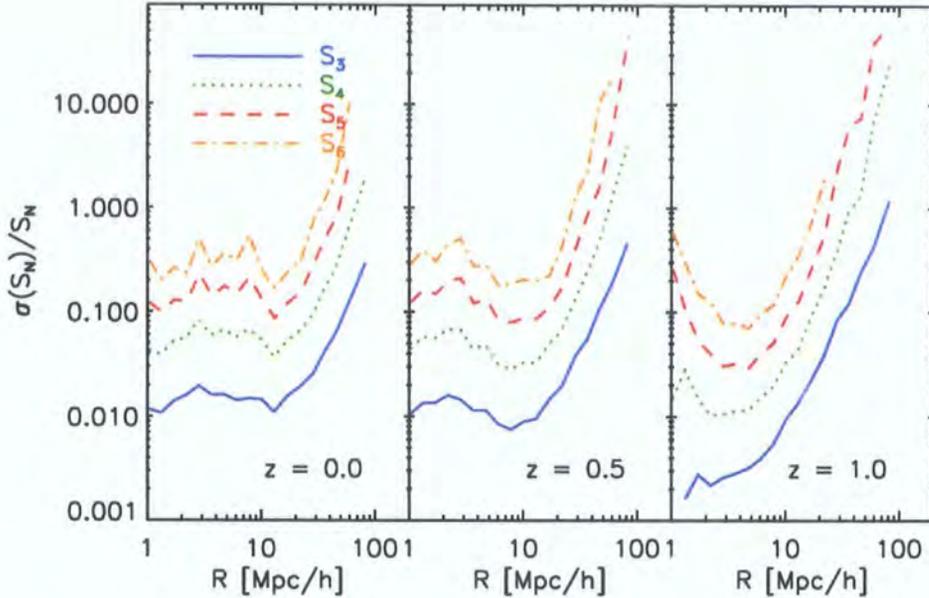


Figure 4.3: The fractional scatter,  $\sigma(S_N)/S_N$ , in the measured hierarchical amplitudes, as estimated from the 50 simulations in the L-BASICC ensemble. Different lines show the scatter for different orders as indicated by the legend.

functions change substantially between  $z = 4$  and  $z = 0$ . For example, for a cell of radius  $50h^{-1}\text{Mpc}$ , the two-point volume averaged correlation function increases by a factor of 14 over this redshift interval, and the three-point function by a factor of 197.

Nevertheless, the simulation results do tend to exceed the perturbation theory predictions on all scales at all orders as the density fluctuations grow.

The hierarchical amplitudes measured on small scales differ significantly from the predictions of perturbation theory. At  $z = 4$ , the simulation results are below the analytical predictions for cell radii smaller than  $R \sim 5h^{-1}\text{Mpc}$ . This behaviour is sensitive to the arrangement of particles which is perturbed to set up the initial density field. At later times, the memory of the initial conditions is erased on small scales and the measured amplitudes greatly exceed the expectations of perturbation theory. On these scales, the dominant contribution to the cross correlation moments is from particles within common dark matter haloes. Note that in Fig. 4.3.1 we do not correct the measured higher order correlation functions for Poisson noise, since the initial density field was created by perturbing particles distributed in a glass-like configuration which is sub-Poissonian. Hence, the dark matter density field is not a random sampling of a continuous density field (see

Angulo et al. (2008) for an extended discussion of this point). The turnover in the hierarchical amplitudes seen at small cell radii (e.g. for  $R < 2h^{-1}\text{Mpc}$ ) is due to the finite resolution of the L-BASICC simulations; the hierarchical amplitudes continue to increase in amplitude on smaller smoothing scales in the BASICC run.

The lower panels of Fig. 4.3.1 show the impact of gravitationally induced peculiar motions on the hierarchical amplitudes. We model redshift-space distortions using the distant observer approximation, in which peculiar motions perturb the particle position parallel to one of the co-ordinate axes. Virialized structures appear elongated when viewed in redshift space. On large scales, coherent bulk flows tend to increase the amplitude of correlation functions. There is a modest reduction in the amplitude of the hierarchical amplitudes on large scales. On small scales, there is a dramatic reduction in the magnitude of the  $S_N$ . The overall impact of the redshift-space distortions is to greatly reduce the dependence of the hierarchical amplitudes on smoothing scale (see Hoyle et al. 2000).

The estimated error on the measured hierarchical moments is shown in Fig. 4.3, in which we plot the fractional error on  $S_N$  obtained from the scatter in the measurements from the L-BASICC ensemble. The plot suggests that the skewness of the dark matter can be well measured on all smoothing scales considered from a volume of the size of the L-BASICC simulation cube. The range of scales over which robust measurements can be made of the hierarchical amplitudes becomes progressively narrower with increasing order. For example, at  $z = 0$ , reliable measurements of  $S_6$  are limited to smoothing radii smaller than  $R \sim 30h^{-1}\text{Mpc}$ .

### 4.3.2 The hierarchical amplitudes of dark matter haloes

The hierarchical amplitudes of dark matter haloes are more complicated than those of the dark matter. In addition to a term arising from the evolution of the density field under gravitational instability, there is a contribution which depends upon height of the peak in the initial density field which collapses to form the halo (Mo, Jing, and White, 1997). For example, if we consider the second and third order auto-correlation functions of haloes given by Eqs. 4.46 and 4.49, then the skewness for dark matter haloes,  $S_3^{\text{H}}$ , is given by:

$$S_3^{\text{H}} = \frac{\bar{\xi}_{3,0}}{(\bar{\xi}_{2,0})^2} \quad (4.59)$$

$$= \frac{3b_2}{b_1^2} + \frac{S_3}{b_1}. \quad (4.60)$$

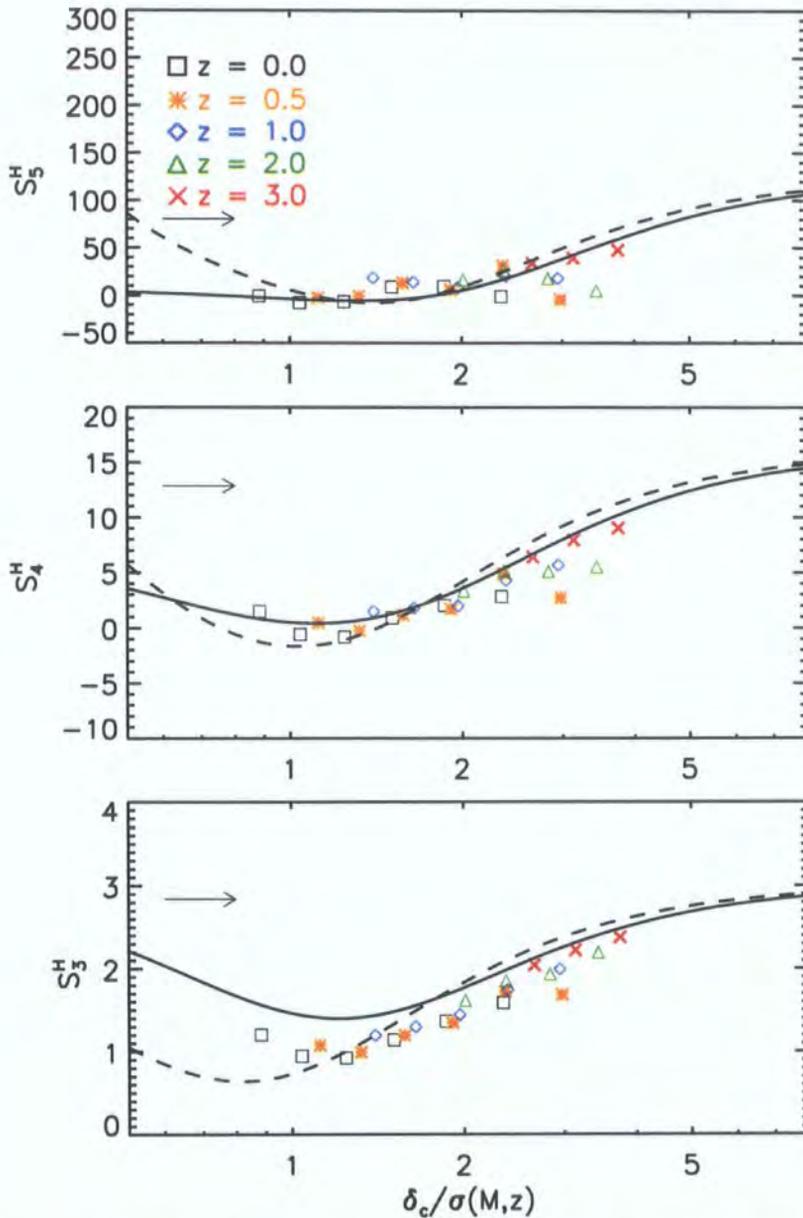


Figure 4.4: The hierarchical amplitudes of dark matter haloes, plotted as a function of the peak height corresponding to the halo mass. In this plot, the hierarchical amplitudes are averaged over cell sizes of  $20h^{-1}\text{Mpc} < R < 50h^{-1}\text{Mpc}$ . The dashed curve shows a theoretical prediction based on the spherical collapse model (Mo, Jing & White 1997) and the solid line shows a revised prediction based upon an ellipsoidal collapse, by Sheth, Mo & Tormen (2001). The corresponding hierarchical amplitudes for the dark matter, averaged over the same range of cell radii, are indicated in each panel by the arrow.

The gravitational contribution to the skewness,  $S_3$ , is diluted by the linear bias factor,  $b_1$ . In the case of rare peaks, or, equivalently, haloes with masses far in excess of the characteristic mass,  $M_*$ , at a given redshift,  $S_3^H$  approaches an asymptotic value. In this limit,  $b_k \approx b_1^k$  and so  $S_3^H \approx 3$ ; similar arguments for the fourth and fifth order hierarchical amplitudes yield asymptotic values of  $S_4^H = 16$  and  $S_5^H = 125$  (Mo, Jing, and White, 1997). Massive haloes at high redshift can therefore have non-zero hierarchical amplitudes even if the dark matter distribution still has a Gaussian distribution and hence  $S_p^{\text{DM}} = 0$ .

We plot the hierarchical amplitudes of dark matter haloes in Fig. 4.4, as a function of the scaled peak height,  $\delta_c/\sigma(M, z)$ . The simulation results are averaged over smoothing radii of  $20h^{-1}\text{Mpc} < R < 50h^{-1}\text{Mpc}$ . The dashed line shows the prediction obtained assuming the mass function of Press and Schechter (1974) and the spherical collapse model (see Mo, Jing, and White 1997). The solid line shows an improved calculation which uses the ellipsoidal collapse model and the mass function derived by Sheth et al. (2001). There is some dispersion between the simulation results at different redshifts. The measurements are in reasonable agreement with the theoretical predictions for large values of  $\delta_c/\sigma(M, z)$ . For more modest peaks, the hierarchical amplitudes of haloes averaged on large smoothing scales show a dip and are significantly smaller than the amplitude recovered for the dark matter. The strength of this dip is more pronounced in the measurements from the simulations than it is in the theoretical predictions. This discrepancy suggests that the theoretical models do not reproduce the trend of bias with halo mass for such objects, as we shall see in §4.3.4.

### 4.3.3 Cross-correlation estimates of higher order clustering

We now switch to estimating cross-correlation functions instead of auto-correlation functions. To recap §4.2, to reduce the impact of discreteness noise on our measurement of halo clustering, we cross-correlate fluctuations in the spatial distribution of haloes with the fluctuation in the dark matter density within the same cell. As the order of the correlation function increases, the number of possible permutations of halo fluctuations and dark matter fluctuations increases. For a given order of correlation function, the relation between these permutations can be understood using the expressions for the cross-moments given in §4.2.3. The relationship at second order is particularly straightforward. The halo autocorrelation function,  $\bar{\xi}_{2,0}$  (recall the first index gives the order of the halo density contrast and the second index gives the order of the dark matter density contrast) is related to the auto-correlation of the dark matter,  $\bar{\xi}_{0,2}$ , by  $\bar{\xi}_{2,0} = b_1^2 \bar{\xi}_{0,2}$ . The

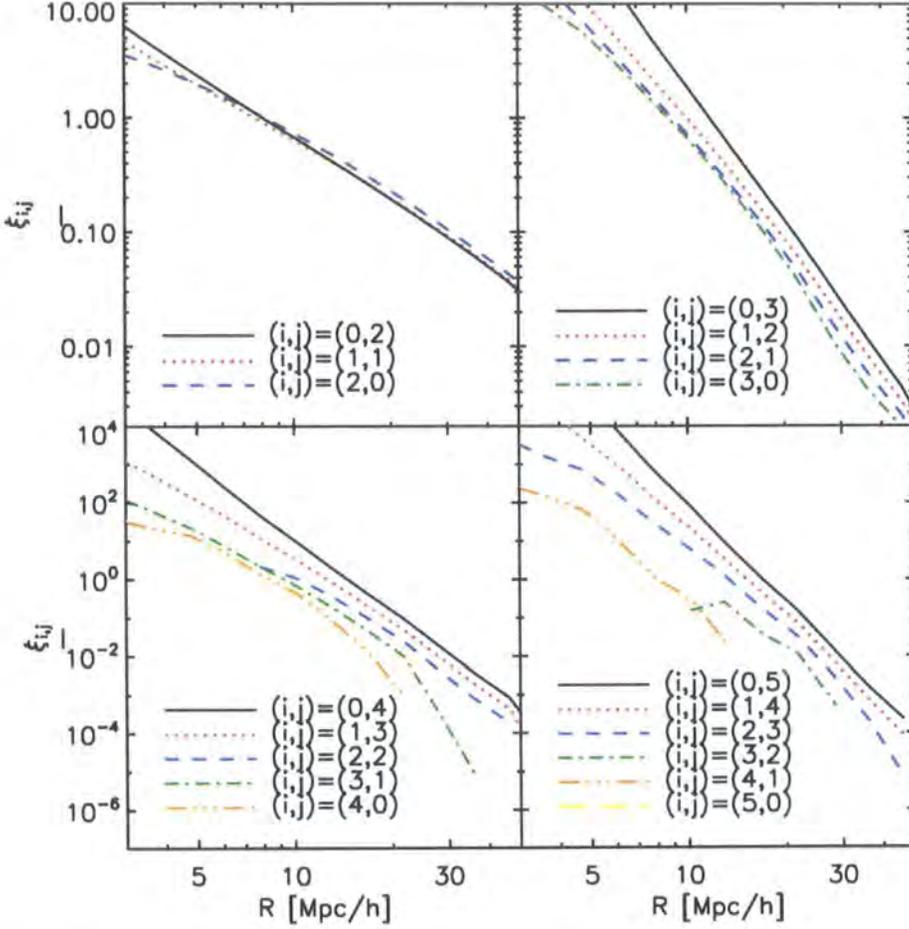


Figure 4.5: Volume-averaged  $i + j$ -point cross-correlation functions,  $\bar{\xi}_{i,j}$ , measured for haloes of mass  $1.1 \times 10^{13} < (M/h^{-1}M_{\odot}) < 2.8 \times 10^{13}$  (619386 objects) and the dark matter at  $z = 0$  in the BASICC simulation. The auto-correlation function of haloes is denoted  $\bar{\xi}_{i+j,0}$  and the auto-correlation of dark matter by  $\bar{\xi}_{0,i+j}$ . Each panel shows a different order of cross-correlation. The key shows the different permutations of cross-correlation function in each case. The moments have been corrected for Poisson noise due to the finite number of haloes.

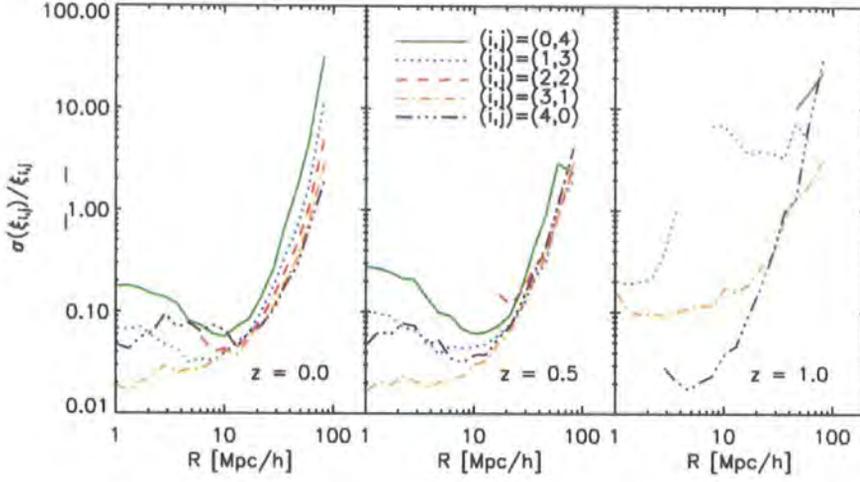


Figure 4.6: The fractional error on the four-point cross correlation functions, estimated from the scatter over the L-BASICC runs. Each panel shows the results for a different redshift, as shown by the key. The legend shows the different permutations of cross-correlation moment. To improve the statistics, all the haloes in the L-BASICC runs have been used in this case.

second-order cross correlation function,  $\bar{\xi}_{1,1}$ , is related to the autocorrelation function of dark matter by  $\bar{\xi}_{1,1} = b_1 \bar{\xi}_{0,2}$ . The primary difference between  $\bar{\xi}_{2,0}$  and  $\bar{\xi}_{1,1}$  is therefore a factor of  $b_1$ . This basic trend is approximately replicated for any order of correlation function: as fluctuations in the halo density are substituted by fluctuations in the dark matter, the amplitude of the cross-correlation is reduced by a factor which depends on  $b_1$ . Above second order, this factor is modulated by higher order bias terms and the hierarchical amplitudes of the dark matter (see §4.2.2). The precise relation between the different permutations of cross-correlation functions depends upon the values of the bias parameters and therefore on the halo mass under consideration.

We show illustrative examples of volume-averaged cross-correlation functions,  $\bar{\xi}_{i,j}$ , estimated from the BASICC simulation in Fig. 4.5. Each panel shows a different order of clustering, starting with the second moment in the top left panel and ending with the fifth order correlation function in the bottom right panel. In this plot, the haloes used have masses in the range  $1.1 \times 10^{13} < (M_{\text{halo}}/h^{-1}M_{\odot}) < 2.8 \times 10^{13}$  and the clustering is measured at  $z = 0$ . The top-left panel of Fig. 4.5 shows that there is little difference in the amplitude of the second-order correlation function on large smoothing scales between the different permutations of  $i, j$ . This implies that for these haloes, the linear bias

term  $b_1 \approx 1$ . The correlation functions are, however, different on small scales. The autocorrelation function of the dark matter ( $\bar{\xi}_{0,2}$ ) is steeper than the autocorrelation of haloes ( $\bar{\xi}_{0,2}$ ). The cross-correlation functions are different on large scales for third, fourth and fifth orders. The difference in amplitude is fairly independent of scale for cells with radii  $R > 10h^{-1}\text{Mpc}$ . Since the linear bias of this sample of haloes is close to unity, this difference is driven by the higher order bias terms and the hierarchical amplitudes of the dark matter.

One might be concerned that replacing fluctuations in halo density by fluctuations in dark matter in the higher order correlation functions leads to a reduction in the clustering amplitude (as is indeed apparent in Fig. 4.5). However, this is more than offset by a reduction in the noise or scatter of the measurement. The fractional error on the measurements of the cross-correlation functions is plotted in Fig. 4.6. The scatter is estimated using the L-BASICC ensemble. Each panel shows the scatter at a different redshift. The cross correlation  $\bar{\xi}_{1,i+j-1}$  (i.e. one part halo fluctuation,  $i + j - 1$  parts dark matter fluctuation) gives the optimal error estimate, with a performance comparable to the auto-correlation of the dark matter. At  $z = 1$ , it is not possible to measure the four-point autocorrelation function of this sample of haloes, even with a box of the size of the L-BASICC runs. Nevertheless, it is possible to measure the bias factors relating the four-point functions of haloes and mass using the cross-correlation. Our use of a cross-correlation estimator therefore allows us to extend the measurements of the higher order clustering of haloes to orders and redshifts that would not be possible using auto-correlations.

#### 4.3.4 The bias parameters of dark matter haloes

We now use the cross correlation functions to estimate the linear and higher order bias parameters of dark matter haloes. As we demonstrated in the previous section, the best possible measurement of the  $i + j^{\text{th}}$  order correlation function is obtained when the cross correlation function is made up of one part fluctuation in halo density and  $i + j - 1$  parts dark matter fluctuation: i.e. in our notation  $\bar{\xi}_{1,i+j-1}$ . This approach, combined with the huge volume of our simulation, makes it possible, for the first time, to measure the third and fourth order bias parameters, and to do so using narrow mass bins.

In this section, we use the higher resolution BASICC run, which can resolve the largest dynamic range in halo mass. We use the higher order correlation function measurements over the range of smoothing radii  $15 < (R/h^{-1}\text{Mpc}) < 50$  to estimate the halo bias

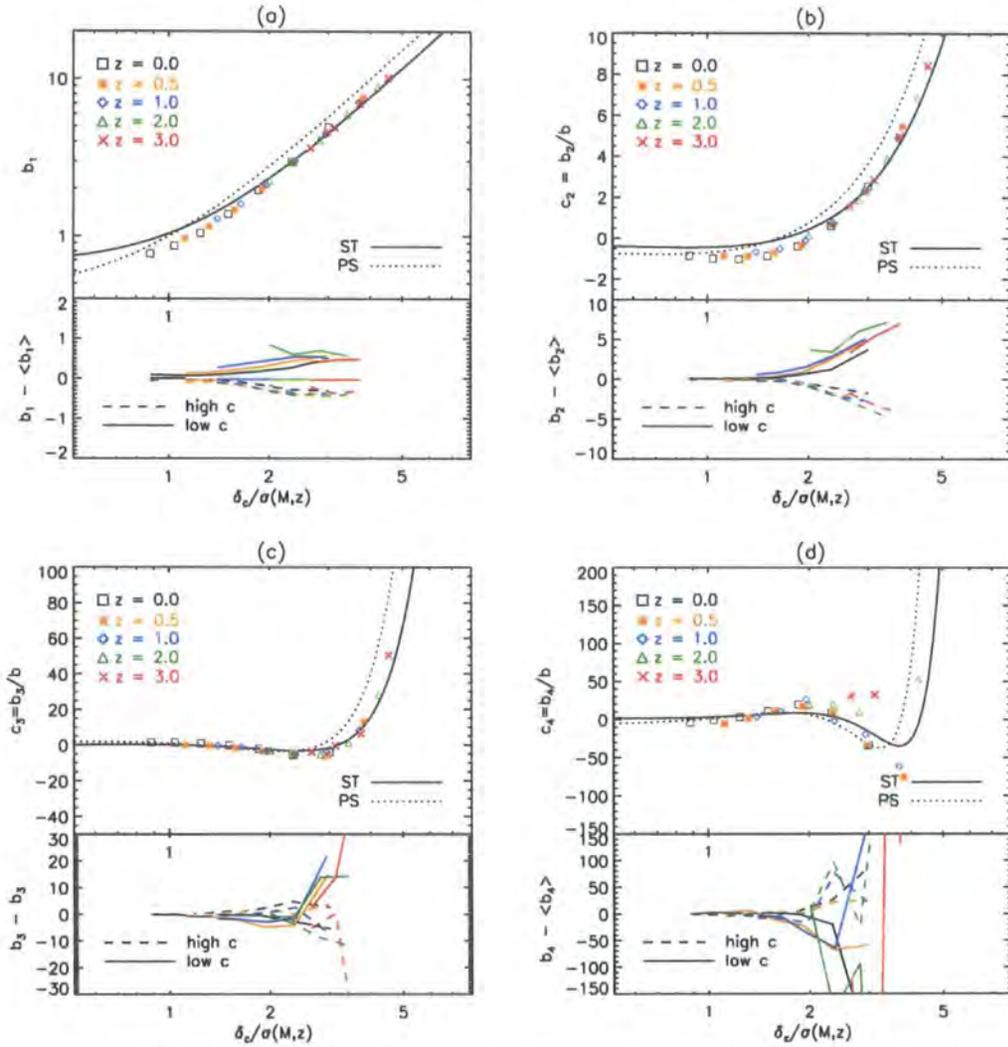


Figure 4.7: The bias parameters as a function of halo mass parametrized by  $\nu = \delta_c/\sigma(M, z)$ . Each plot shows a different order of bias parameter: a) linear bias  $b_1$ , b) the ratio of the 2nd order bias,  $b_2/b_1$ , c) the ratio of the 3rd order bias,  $b_3/b_1$  and d)  $b_4/b_1$ . In the lower panel of each plot, the residual bias parameters for the 20% of haloes with the highest or lowest values of  $V_{\max}/V_{200}$ , a proxy for concentration, are plotted. In the upper panels, symbols show the measurements for different output redshifts, as indicated by the key. The same line colours are used to show the results for different redshifts in the lower panels. In the upper panel of each plot, we plot two theoretical predictions for the bias parameters, given by Mo, Jing & White (1997) and Scoccimarro et al. (2001).

parameters. The large volume of the BASICC simulation means that we can make robust measurements of the higher order correlation functions out to larger smoothing radii than is possible with the smaller Millennium simulation. The smallest scale we use is set by the requirement that the expansion relating the overdensity in haloes to the overdensity in dark matter (Eq. 4.44) is a good approximation, i.e. when  $\bar{\xi} \ll 1$ . The scales we use to extract the halo bias parameters are considerably larger than those Gao, Springel & White (2005) and Gao & White (2007) were able to use in the Millennium. We use the simulation outputs at redshifts of  $z = 0, 0.5, 1, 2$  and 3 to measure the clustering of haloes.

The results for the first, second, third and fourth order bias parameters of dark matter haloes are presented in Fig. 4.7. Each panel corresponds to a different order. The upper half of each panel shows the respective order of bias parameter as a function of halo mass, expressed in terms of the peak height corresponding to the halo mass,  $\delta_c/\sigma(M, z)$ . The lower half of each panel shows the deviation from the bias parameter extracted for a given mass for samples of the 20% of haloes in the mass bin with the highest and lowest values of  $V_{\max}/V_{200}$ , which we are using as a proxy for halo concentration. Different symbols in the upper panels show the measurements at different output redshifts in the BASICC run, as indicated by the key; the same colours are used to draw the lines showing results for samples defined by different  $V_{\max}/V_{200}$  values at the same output redshifts in the lower panels.

In Fig. 4.7, there is remarkably little scatter between the results obtained from the different output redshifts for the case of the overall bias as a function of mass. This is encouraging, as it shows that our results are not affected by resolution (haloes with similar values of  $\delta_c/\sigma(M, z)$  at different output times are made up of very different numbers of particles). Gao, Springel & White (2005) were able to measure the linear bias parameter up to haloes corresponding to peak heights of  $3\sigma$ ; we are able to extract measurements for haloes corresponding to  $5\sigma$  peaks.

In the upper sub-panels of Fig. 4.7, we show two theoretical predictions for the bias parameters of dark matter haloes. The dotted lines show the predictions from Mo, Jing, and White (1997), based on an extension of Press & Schechter's (1974) theory for abundance of dark matter haloes and the spherical collapse model. The solid lines show the calculation from Scoccimarro et al. (2001) which uses the mass function of Sheth and Tormen (1999). Our results tend to agree best with the later, although the measurements favour a steeper dependence of bias on peak height at all orders. For less rare peaks, neither theoretical model gives a particularly good fit to the simulation results. A similar

trend, albeit with more scatter between the results at different output redshifts was found by Gao, Springel, and White (2005) (see also Wechsler et al. 2006; Jing, Suto, and Mo 2007).

Previous studies have reported a dependence of clustering strength on a second halo property besides mass, such as halo formation time or concentration (Wechsler et al. 2006; Gao & White 2007). We do not have sufficient output times to make a robust estimate of formation time so we use a proxy for halo concentration instead,  $V_{\text{max}}/V_{200}$ . We find that the clustering of high peak haloes is sensitive to whether the halo has a high or low value of  $V_{\text{max}}/V_{200}$ . The 20% of haloes with the lowest values of  $V_{\text{max}}/V_{200}$  within a given mass bin (i.e. those with the lowest concentrations) have the largest linear and second order bias terms. This result agrees with previous estimates of the dependence of the linear bias term on halo concentration (Wechsler et al. 2006).

The peak height dependence of the third and fourth order bias terms for haloes split by  $V_{\text{max}}/V_{200}$  is more complicated. Fig. 4.7 shows that the third order bias depends on our concentration proxy in a non-monotonic fashion. The trend for the fourth order bias is reversed compared with the results for the first and second order bias parameters: low concentration haloes have a negative value of the fourth order bias. We note that it would not be possible to measure a fourth order bias at all using halo auto-correlation functions.

One might be concerned that our results could be sensitive to the operation of the group finder. In particular, it is well known that the FoF algorithm can sometimes spuriously link together distinct haloes into a larger halo, through bridges of particles (e.g. Cole and Lacey 1996). We therefore carried out the exercise of relabelling the mass of each halo by the mass of the largest substructure as determined by SUBFIND. In the rare cases in which haloes are incorrectly linked into a larger structure, using instead the SUBFIND mass would result in a significant shift in the mass bin to which the halo is assigned. Moreover, one would expect that low concentration FoF haloes would be more prone to being broken up in this way. However, we found no change in our results upon following this procedure, demonstrating that the trends we find for the dependence of bias on mass and concentration are robust.

## 4.4 Summary and Discussion

In this chapter, we have combined ultra-large volume cosmological simulations with a novel approach to estimating the higher order correlation functions of dilute samples of

objects. The large simulation volume allows us to extract bias parameters on large scales, which follow linear perturbation theory more closely, and provides us with large samples of high mass haloes from which robust clustering measurements can be made. The cross-moment counts-in-cells technique we use to estimate the higher order clustering of dark matter haloes has superior noise performance to traditional autocorrelation functions, allowing us to probe clustering to higher orders. These improvements made it possible to extend previous work on the assembly bias of dark matter haloes in a number of ways. We have been able to extract measurements of halo clustering for objects corresponding to  $5\sigma$  peaks, almost twice as high as in earlier studies. We have also presented, for the first time, estimates of the higher order bias parameters of haloes, up to fourth order, and using narrow mass bins.

Our results are in qualitative agreement with those in the literature where they overlap. We find that the linear bias factor,  $b_1$ , is a strong function of mass, varying by an order of magnitude for peaks ranging in height from  $\delta_c/\sigma(M, z) = 1$  to 5. We use the ratio of the maximum of the effective halo rotation speed to the speed at the virial radius,  $V_{\max}/V_{200}$  as a proxy for halo concentration. High mass, high  $V_{\max}/V_{200}$  haloes are less strongly clustered than the same mass haloes with low values of  $V_{\max}/V_{200}$ ; haloes with  $\delta_c/\sigma(M, z) \sim 4$  display second order clustering that differs by  $\approx 25\%$  between the 20% with the lowest values of  $V_{\max}/V_{200}$  and the 20% of the population with the highest values of this ratio.

It is reassuring that we recover a similar dependence of the linear bias on halo mass when labelling haloes by  $V_{\max}/V_{200}$  as other authors found using the concentration parameter (Wechsler et al. 2006). This trend is the opposite to that recovered when halo samples are split by formation time. Gao, Springel, and White (2005) found no dependence of the clustering signal on halo formation time for massive haloes. This is puzzling since formation time and concentration are correlated, albeit with scatter (e.g. Neto et al. 2007). Croton, Gao, and White (2007) have argued that this suggests that an as yet unknown halo property is a more fundamental property in terms of determining the clustering strength (for theoretical explanations of the physical basis of assembly bias see e.g. Zentner 2007; Ariel Keselman and Nusser 2007; Dalal et al. 2008 ).

The second order bias parameter,  $b_2$ , displays qualitatively similar dependences on mass and  $V_{\max}/V_{200}$  to  $b_1$  with the difference that  $b_2$  is negative around  $\delta_c/\sigma(M, z) \sim 1$ . The third and fourth order bias parameters are more complicated, being essentially independent of mass until peaks  $\delta_c/\sigma(M, z) \sim 2 - 3$  are reached, where there is a dip in

bias before a rapid increase for rarer peaks. The dependence on  $V_{\max}/V_{200}$  is also different at third and fourth order.

We compared our measurements for the bias parameters with analytic predictions. For haloes corresponding to rare peaks, the trend in linear bias versus peak height is intermediate between the predictions of Mo, Jing, and White (1997), which are based on Press & Schechter's (1974) theory for the abundance of haloes and the spherical collapse model, and the calculation of Sheth, Mo & Tormen (2001) and Scoccimarro et al. (2001), based on ellipsoidal collapse and an improved estimate of the halo mass function. Both analytic calculations predict a weaker dependence of  $b_1$  on peak height around  $\delta_c/\sigma(M, z) \sim 1$  than we find in the simulation. The comparison between the simulation measurements and the analytic predictions is similar for  $b_2$ . For the third and fourth order bias parameters, the simulation results are in good agreement with the analytic predictions for modest peaks. For rare peaks, the bias parameters measured from the simulation are again in between the two analytic predictions.

Observations of clustering are already entering the regime in which our simulation can play an important role in interpreting the measurements. Existing observations of high redshift quasar clustering suggesting that these objects live in haloes corresponding to  $\sim 5 - 6$  sigma peaks in the matter distribution at  $z = 4$  (White, Martini & Cohn 2007). Future galaxy surveys, due to the volume covered and number of galaxies targeted, will yield measurements of clustering with unprecedented accuracy, to higher orders than the two-point function. The measurements presented in this chapter will provide invaluable input to future models of galaxy clustering based on halo occupation distribution models, which have been modified such that galaxy clustering is a function of mass and a second halo property.



# Chapter 5

## *The Fate of Substructures in Cold Dark Matter Haloes*

### 5.1 Introduction

The presence of substructures within dark matter haloes is a distinctive signature of a universe where structures grow hierarchically. Low mass objects collapse at high redshift, and then increase their mass by smooth accretion of dark matter or by merging with other haloes. Once a halo is accreted by a larger one, its diffuse outer layers are rapidly stripped off by tidal forces. However, the core, which is much denser, generally survives the accretion event and can still be recognized as a self bound structure or subhalo within the host halo for some period of time afterwards.

In early N-body simulations, haloes appeared as fairly smooth objects (Frenk et al., 1985, 1988). However, as the attainable mass and force resolution has increased, subhaloes have been identified and their properties studied in detail by many authors over the past decade (e.g. Ghigna et al., 1998, 2000; Tormen et al., 1998; Moore et al., 1999; Klypin et al., 1999b,a; Springel et al., 2001; Stoehr et al., 2002; De Lucia et al., 2004; Gao et al., 2004; Nagai and Kravtsov, 2005; Shaw et al., 2007; Diemand et al., 2008; Springel et al., 2008). The properties of the subhalo population have important implications for galaxy formation, dark matter detection experiments and weak lensing. For instance, subhaloes are expected to host satellite galaxies within groups and clusters and their evolution once inside the host could give rise to observable changes. In particular, a merger between two substructures could trigger an episode of star formation or a morphological transformation (e.g. Somerville and Primack, 1999).

In spite of this, the merger history of subhaloes remains relatively unexplored. This is a challenging problem which demands a simulation with high mass and force resolution.



In particular, obtaining a statistical sample of mergers involving the largest substructures requires a large sample of host haloes. Most studies of substructure in halos have focused on resimulating, at very high resolution, a small number of halos selected from a larger, lower resolution simulation. However, by studying only a few haloes, important aspects related to variations produced by differences in the accretion and merger histories of haloes, as well as any influence of the environment could remain hidden. This approach may also introduce systematic biases arising from the criteria used to select the haloes to be resimulated.

In this chapter, we overcome these problems by using the largest dark matter simulation published to date, the Millennium Simulation (MS, Springel et al. 2005). The MS provides a large cosmological sample of dark matter haloes and associated substructures spanning a considerable range in mass, allowing us to assess robustly the properties and fate of the subhalo population. We complement our results with a higher resolution simulation of a smaller volume (hereafter HS) which has a particle mass almost ten times smaller than that used in the MS (Jenkins et.al, in prep).

The layout of this chapter is as follows. In Section 5.2, we briefly describe the simulations used in this work along with the properties of our halo and subhalo catalogues. In Section 5.3 we investigate some general properties of subhaloes, namely their mass function, radial distribution and spatial orientation with respect to their host halo. The exploration of substructure mergers and destruction is presented in Section 5.4. Finally, we summarize our findings in Section 5.5.

## 5.2 Method

In this section we describe the N-body simulations we have used analyzed in this work. We also discuss the identification and characterization of the halo and subhalo catalogues.

### 5.2.1 N-body Simulations

The main simulation on which our analysis is based is the Millennium Simulation (Springel et al., 2005). The MS covers a comoving volume of  $0.125 h^{-3} \text{Gpc}^3$  of a  $\Lambda$ CDM universe in which the dark matter component is represented by  $2160^3$  particles. The assumed cosmological parameters are in broad agreement with those derived from joint analyses of the 2dFGRS galaxy clustering (Percival et al., 2001) and WMAP1 microwave background data (Spergel et al., 2003; Sánchez et al., 2006), as well as with those derived from WMAP5

data (Komatsu et al., 2008). In particular, the total mass-energy density, in units of the critical density, is  $\Omega_m = \Omega_{\text{dm}} + \Omega_b = 0.25$ , where the two terms refer to dark matter and baryons, with  $\Omega_b = 0.045$ ; the amplitude of the linear density fluctuations in  $8h^{-1}\text{Mpc}$  spheres is  $\sigma_8 = 0.9$ ; and the Hubble constant is set to  $h = H_0/(100 \text{ kms}^{-1}\text{Mpc}^{-1}) = 0.73$ . The particle mass is  $m_p = 8.6 \times 10^8 h^{-1}M_\odot$  and the Plummer-equivalent softening of the gravitational force is  $\epsilon = 5 h^{-1}\text{kpc}$ .

To complement our results and to test for numerical effects we have also employed another simulation with better mass resolution to which we refer as HS. This simulation follows  $900^3$  dark matter particles in a  $\Lambda\text{CDM}$  cube of side  $100 h^{-1}\text{Mpc}$ . The HS assumes the same cosmological parameters as the MS. However, the smaller box yields a smaller particle mass,  $m_p = 9.5 \times 10^7 h^{-1}M_\odot$ , so objects of a given mass are resolved with almost 10 times more particles than in the MS. Finally, in the HS the softening length is  $\epsilon = 2.4 h^{-1}\text{kpc}$ .

The MS and HS were carried out using a memory efficient version of the `Gadget-2` code (Springel, 2005).

### 5.2.2 Halo and Subhalo catalogues

In both simulations, particle positions and velocities are written at 64 output times which, for  $z < 2$ , are roughly equally spaced in time by 300 Myr. In each of these outputs we have identified dark matter haloes using the friends-of-friends (FoF) algorithm (Davis et al., 1985), with a linking length of 0.2 times the mean interparticle separation. The volume and particle number of the MS provide a unique resource of well resolved haloes to study. By way of illustration, there are 90891 haloes at  $z = 0$  with mass in excess of  $5.4 \times 10^{12} h^{-1}M_\odot$  (one of the bins we use below), which corresponds to 6272 particles; at  $z = 1$  the number of haloes in excess of this mass is still 61481. On the cluster-mass scale, for example, there are 356 haloes at  $z = 0$  which are more massive than  $4 \times 10^{14} h^{-1}M_\odot$ , corresponding to 464576 particles.

Well resolved FoF haloes are not smooth, but contain a considerable amount of mass in the form of substructures. These dark matter clumps or “subhaloes” are identified and catalogued using a modified version of the subhalo finder algorithm, `SUBFIND`, originally presented in Springel et al. (2001). The statistics of the subhalo catalogue are impressive. At  $z = 0$  `SUBFIND` lists 339840 structures with more than 200 particles in the MS within haloes of at least  $5.4 \times 10^{12} h^{-1}M_\odot$ . At  $z = 1$  there are 194629 substructures with the same characteristics. Note that `SUBFIND` not only finds substructures within a FoF halo,

but it is also capable of identifying substructures within substructures.

An important issue for studies of substructures is the definition of the boundary and position of the host halo. In our analysis, the centre of the halo is defined as the position of the particle possessing the minimum gravitational potential. This choice for the halo position agrees, within the softening length, with that found by a shrinking sphere algorithm (Power et al., 2003) for 93% of the haloes that are resolved with 450 or more particles. As shown by Neto et al. (2007), the 7% of cases in which the two methods disagree are due to the FoF algorithm artificially linking multiple structures. In these cases the position of the most bound particle provides a more robust identification of the centre, as noted by Neto et al. (2007).

We define the halo boundary as the sphere of radius  $r_{200}$  which contains a mean density of 200 times the critical density,  $\rho_{\text{crit}}$ . Therefore, the mass of the halo is:

$$M_{200} = \frac{4}{3}\pi 200\rho_{\text{crit}}r_{200}^3. \quad (5.1)$$

We keep in our catalogues only subhaloes within this sphere. Although the choice of the factor of 200 is motivated by the spherical collapse model in a Einstein-de-Sitter universe, it is somewhat arbitrary for our  $\Lambda$ CDM simulations. However, the  $r_{200}$  definition has the advantage of being independent of both redshift and cosmology. Moreover, it has become a de facto standard in studies of substructures. Nevertheless, we have tested our results against other definitions of the halo boundary without finding any qualitative differences. In the following, when we refer to the mass and radius of a host halo, we always mean  $M_{200}$  and  $r_{200}$ .

Finally, we build merger trees using an algorithm similar to that described by Springel et al. (2005) which follows the evolution of subhaloes. In this way, we can identify the haloes and subhaloes that will be involved in a merger during a subsequent snapshot. Note that these merger trees are constructed using only the information contained in the FoF and SUBFIND catalogues, and there is no attempt to force mass conservation, as would be required if the merger trees were to be used in a galaxy formation code (see Harker et al., 2006). The descendant of a subhalo is defined as the structure that contains the majority of the 10 percent most bound particles from the subhalo. When two satellite subhaloes have the same descendant in a following snapshot, we tag such an event as a substructure merger.

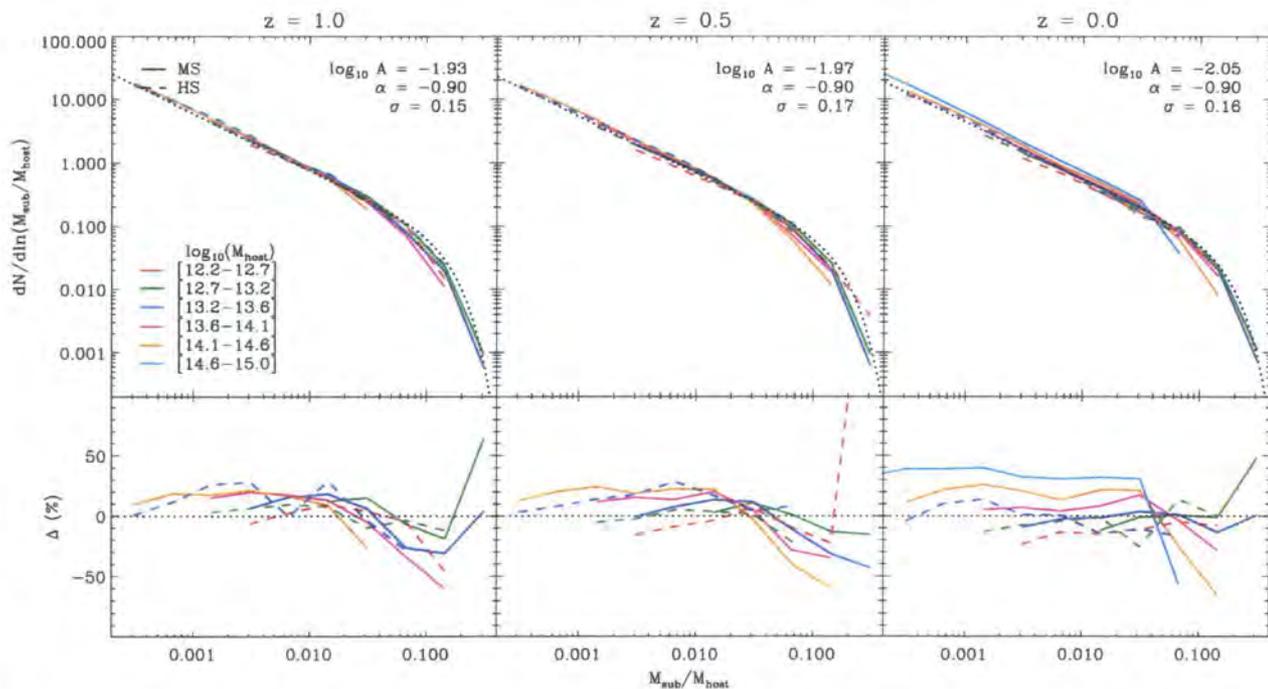


Figure 5.1: *Top row*: Differential number of substructures per host halo as a function of their mass relative to that of the host halo,  $M_{\text{sub}}/M_{\text{host}}$ . Note this is the mass of the subhalo at the redshift labelled, in some cases after substantial stripping of mass has taken place. Solid lines show the results from the MS while dashed lines show the results from the HS. In both cases, lines of different colours show the subhalo mass function in host haloes of different masses (as indicated by the legend). Each column shows a different redshift, as labelled. At each redshift, the dotted lines display the overall best fit of our model, Eq. 5.2, with the parameter values given in the legend. Parameters of the fits to individual mass bins are listed in Table 1. *Bottom row*: Relative difference between the overall best fit and measurements of the subhalo mass function for different host masses. The dot-dashed line shows the difference between our model, Eq. 5.2, and a power-law fit. Only results for subhaloes which are resolved with more than 50 particles are shown.

### 5.3 SubHalo properties

Before presenting our results regarding subhalo mergers, we consider some general properties of the subhalo population. Although some of these properties have been studied by previous authors, the large volume and high resolution of the MS and HS reveal some features which were inaccessible in earlier work. Furthermore, the knowledge of the subhalo properties will help us to understand the results presented in the next section.

#### 5.3.1 Subhalo mass function

We first consider the distribution of subhalo masses - the subhalo mass function. The top panels of Fig. 5.1 show the mean number of substructures within dark matter haloes, per host halo, per logarithmic interval in subhalo mass. The results are displayed as a function of subhalo mass relative to the mass of the halo in which it resides,  $M_{\text{sub}}/M_{\text{host}}$ . In this way we can easily compare results across a range of halo masses. In the ranges of overlap, the results from the MS and HS agree well; this provides a useful, but limited, test of convergence of our results.

For the redshifts plotted in Fig. 5.1 there is only a small variation of the subhalo mass function with host halo mass. Indeed, a universal function describes the behaviour reasonably well over the range of subhalo mass resolved by our simulations:

$$\frac{dN}{d \ln(M_{\text{sub}}/M_{\text{host}})} = A \left( \frac{M_{\text{sub}}}{M_{\text{host}}} \right)^\alpha \exp \left[ -\frac{1}{\sigma^2} \left( \frac{M_{\text{sub}}}{M_{\text{host}}} \right)^2 \right], \quad (5.2)$$

where  $N$  is the number of subhaloes per host halo. The values of  $A$ ,  $\alpha$  and  $\sigma$  in this overall fit at each redshift are given in the legend of Fig. 5.1. For this overall fit, we have forced the slope  $\alpha$  to have the same value independently of redshift. In general, we find that  $\alpha = -0.9$  is a good approximation to the best fit from  $z = 0$  to  $z = 2.5$ . It is also important to note that the power-law fit widely used in the literature, (e.g. Gao et al. 2004) is only valid over a limited range of fractional subhalo masses,  $M_{\text{sub}}/M_{\text{host}} < 0.04$ . We also see that the maximum subhalo mass for which a power-law successfully describes the mass function decreases at higher redshifts,  $M_{\text{sub}}/M_{\text{host}} \sim 0.015$  at  $z = 1$  and  $M_{\text{sub}}/M_{\text{host}} \sim 0.04$  at  $z = 0$ . The bottom panels of Fig. 5.1 show the relative difference between the fit given by Eq. (5.2) and the mass function of subhaloes measured in host haloes of different masses.

We have also fitted Eq. 5.2 to the subhalo mass functions in each halo mass bin, this time letting the slope  $\alpha$  vary; we list the best-fit parameters for  $z = 0$  in Table 5.1. At the low fractional mass end, where the subhalo mass function behaves as a power-law,

	$M_{\text{host}} [h^{-1}M_{\odot}]$	$\log_{10} A$	$\alpha$	$\sigma$	$\log_{10}(M_{\text{sub}}/M_{\text{host}})$
	$9.2 \times 10^{12}$	-2.05	-0.87	0.17	-1.8
	$2.7 \times 10^{13}$	-2.06	-0.89	0.16	-2.5
MS	$7.9 \times 10^{13}$	-1.98	-0.88	0.13	-2.8
	$2.3 \times 10^{14}$	-2.00	-0.90	-	-3.5
	$6.8 \times 10^{14}$	-1.86	-0.87	-	-3.8
	$3.1 \times 10^{12}$	-2.00	-0.83	0.16	-2.5
HS	$9.2 \times 10^{12}$	-2.05	-0.88	0.17	-2.8
	$2.7 \times 10^{13}$	-2.15	-0.93	0.14	-3.5

Table 5.1: The best-fit parameters to the mass function of subhaloes residing in haloes of different mass at  $z = 0$ , using Eq. 5.2. The columns are as follows: (1) The N-body simulation from which the halo sample was extracted. (2) The mean mass of the host haloes. (3) The logarithm of the amplitude. (4) The power-law index. (5) The damping strength. (6) The minimum fractional subhalo mass included in the fitting.

we generally find slopes that are lower than the critical value,  $\alpha = -1$  (which separates divergent from convergent mass functions). The slopes we find are in broad agreement with previous estimates of the power-law index of the subhalo mass function, which range from  $-0.8$  to  $-1.0$  (Moore et al., 1999; Ghigna et al., 2000; De Lucia et al., 2004; Gao et al., 2004; Diemand et al., 2004; Shaw et al., 2007; Diemand et al., 2007). In particular, our results agree with those from the much higher resolution simulations of individual galactic halos of Springel et al. (2008), but are inconsistent with the steeper slope advocated, also for galactic halos, by Diemand et al. (2008).

At the high mass end, the subhalo mass function departs from a power-law and decreases exponentially. This behaviour was previously detected in N-body simulations (at lower significance) by Giocoli et al. (2008) (and predicted analytically by van den Bosch et al. (2005)). However, this feature was not apparent in earlier studies which used resimulations of individual haloes. Resimulations of single objects have the advantage that computational effort can be focused. A halo can be resolved with a vast number of particles and its substructures identified over a large range of masses. Unfortunately, this approach comes at the price of losing the rich information contained in the variety of assembly histories, relaxation states and, more importantly, the population of high mass

subhaloes. As can be seen from Fig. 5.1, the abundance of these objects is much lower than that of smaller subhaloes – usually we would find just a few in each halo. Because these haloes are so rare, the damping of the power-law at high  $M_{\text{sub}}/M_{\text{host}}$  is missed in individual resimulations. By contrast, with the huge sample of haloes and their massive subhaloes in our analysis, we can robustly probe this subhalo mass range.

Even though the subhalo mass function appears roughly universal (e.g. Moore et al. 1999), we have detected at every redshift a small dependence on the mass of the host system. Small substructures of the same fractional mass are more abundant in high mass haloes than in low mass haloes. This correlation has also been seen in a number of other studies (e.g. Gao et al., 2004; Shaw et al., 2007; Diemand et al., 2007). However, we also find evidence that this trend holds only in the power-law region of the subhalo mass function and actually reverses at the high mass end - low mass haloes seem to host relatively more massive subhaloes than do high mass haloes.

Perhaps surprisingly, the variety of features present in the mass function of subhaloes is consistent with a relatively simple picture. There are two key ingredients that shape the subhalo mass function: (i) the mass function of infalling objects and (ii) the dynamical evolution of subhaloes orbiting within the host halo due to dynamical friction and tidal stripping. The first of these is responsible for the universality described above and sets the subhalo mass function to first order. As first found by Lacey and Cole (1993) using the extended Press-Schechter formalism, and confirmed by Giocoli et al. (2008) using N-body simulations, the mass function of subhaloes at infall is almost independent of host halo mass and redshift when expressed as a function of  $M_{\text{sub}}/M_{\text{host}}$ , and can be described as a power-law with a high mass cut-off.

After subhaloes fall into a host halo, their orbits sink due to dynamical friction and, at the same time, the subhaloes lose mass due to tidal stripping. These processes cause the subhalo mass function to evolve away from its form at infall. The rates for these processes depend on the fractional mass of the subhalo,  $M_{\text{sub}}/M_{\text{host}}$ , and on the dynamical timescale of the host halo. Therefore, if all haloes had identical structure and assembly histories, these processes would preserve a universal form for the subhalo mass function, independently of  $M_{\text{host}}$ . However, haloes of different masses on average assemble at different redshifts in spite of the similar mass function of subhaloes at infall, and this breaks the universal shape of the subhalo mass function, as discussed by van den Bosch et al. (2005) and Giocoli et al. (2008). On average, massive haloes are younger than their less massive counterparts and they are more likely to have experienced recent mergers

(Lacey and Cole, 1993). These provide a fresh source of substructures which have had less time for orbit decay due to dynamical friction and to be tidally stripped. High mass haloes are therefore expected to have more substructures than low mass haloes. Another effect which acts in the same direction is that small haloes tend to accrete their subhaloes at higher redshifts when dynamical timescales are shorter. As a result, they strip out mass from the substructures more quickly than large haloes, where massive substructures can survive for longer.

### 5.3.2 Most massive subhaloes

The high-mass tail of the distribution of substructure is examined in greater detail in Fig. 5.2. The three panels in this plot display the distribution of the fractional mass,  $M_{\text{sub}}/M_{\text{host}}$ , for the first, second and third largest substructures within haloes of different mass at  $z = 0$ . As before, results from the MS and HS agree very well.

In contrast to the results presented in the previous subsection, the distributions of fractional masses seem to be independent of the host halo mass. (We have also checked that they are independent of redshift.) In particular, in every halo, the fractional masses follow a log-normal distribution with mean  $\langle \log_{10}(M_{\text{sub}}/M_{\text{host}}) \rangle = -1.42, -1.79$  and  $-1.99$ , and standard deviation  $\sigma_{\log_{10}(M_{\text{sub}}/M_{\text{host}})} = 0.517, 0.382$  and  $0.348$  for the each of the three largest subhaloes respectively. Albeit with considerable scatter, these values imply that the most massive substructure contains typically 3.7% of the total mass of the halo while the second and third most massive subhaloes contain 1.6% and 1% of the mass respectively.

Due to the large dispersions, the distributions can only be measured reliably in haloes resolved with a large number of particles. For instance, the mean fractional mass of subhaloes is overestimated for haloes resolved with fewer than  $\sim 1000$  particles (the exact limit depends on the scatter and mean of the true distribution), i.e.  $\sim 1 \times 10^{12} h^{-1} M_{\odot}$  in the MS and  $\sim 1 \times 10^{11} h^{-1} M_{\odot}$  in the HS. The upward bias is caused by the finite resolution of the simulations (there is a limit on the smallest subhalo that we can identify) which truncates the low mass tail of the distribution of fractional masses.

Hints of a universal behaviour of the fractional masses of the largest subhaloes were already detected by De Lucia et al. (2004) (although they claim a weak dependence with host halo mass). Our results are broadly consistent with theirs but, with the large halo catalogues from the MS and HS, we are able to probe the full probability distribution function robustly.

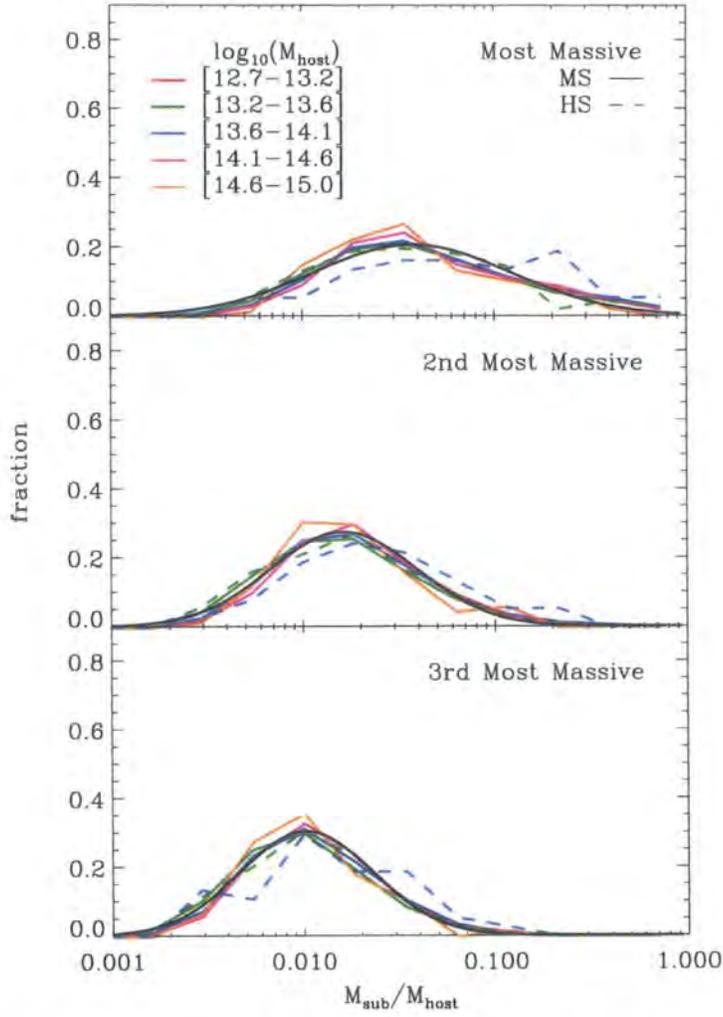


Figure 5.2: The distribution of the fractional mass,  $M_{\text{sub}}/M_{\text{host}}$ , of the 1st, 2nd and 3rd largest substructures in haloes of different mass at  $z = 0$ . The solid lines show the results from the MS while the dashed lines show the results from the HS. In each panel, the black solid lines indicate the log-normal function that best fits our results. Note that only substructures resolved with 20 particles or more are displayed.

The apparently universal shape of these distributions could, in principle, be understood within the broad picture just discussed. Presumably it reflects the distribution of masses of the infalling haloes which, as we have seen, is independent of the host halo mass (Lacey and Cole, 1993; Giocoli et al., 2008). The large scatter must then result from the large range of accretion histories at a given host halo mass. We leave further investigation of these ideas for future work.

### 5.3.3 Radial distribution of subhaloes

Fig. 5.3 shows the number density of subhaloes as a function of radius, relative to the mean number density of substructures within  $r_{200}$  in the same fractional mass range. Each panel focuses on substructures of different masses, from small subhaloes ( $10^{-4} < M_{\text{sub}}/M_{\text{host}} < 10^{-3}$ ) in the leftmost panel to large ones ( $10^{-2} < M_{\text{sub}}/M_{\text{host}} < 1$ ) in the rightmost panel. As in previous plots, lines of different colours show results for subhaloes that reside in haloes of different mass, and the different line types (solid and dashed) indicate the results for the two simulations. We also plot the radial profile of the dark matter as a black dotted line in each panel.

Comparison of the MS and HS indicates that our results are insensitive to the mass resolution (although the overlap between the two simulations is limited). As in previous studies (e.g. Gao et al. 2004), we find that the radial distribution has little dependence on the host halo mass at a given  $M_{\text{sub}}/M_{\text{host}}$ . This is quite remarkable since each panel mixes subhaloes that: (i) are resolved by numbers of particles that differ by orders of magnitude and (ii) occupy haloes which are in a variety of dynamical states (age, relaxation, etc). We also see that in all cases, the radial distribution of subhaloes is less centrally concentrated than the dark matter, as was also found in previous studies (e.g. Ghigna et al., 1998, 2000; Gao et al., 2004; Diemand et al., 2004; Nagai and Kravtsov, 2005; Shaw et al., 2007; Springel et al., 2008).

In addition, we see a significant difference between the distribution of massive subhaloes ( $M_{\text{sub}} > 10^{-2} M_{\text{host}}$ ) and that of small ones ( $M_{\text{sub}} < 10^{-3} M_{\text{host}}$ ). While the overall radial profiles seem to be fairly independent of subhalo mass, the more massive subhaloes tend to avoid the central regions of the host halo, while the less massive ones have a more centrally concentrated distribution (see also De Lucia et al., 2004). However, the distributions agree in the outer parts of the halo.

Springel et al. (2008) found a similar effect to ours in the Aquarius set of simulations of galactic halos which, although limited in number, span a huge dynamic range in subhalo

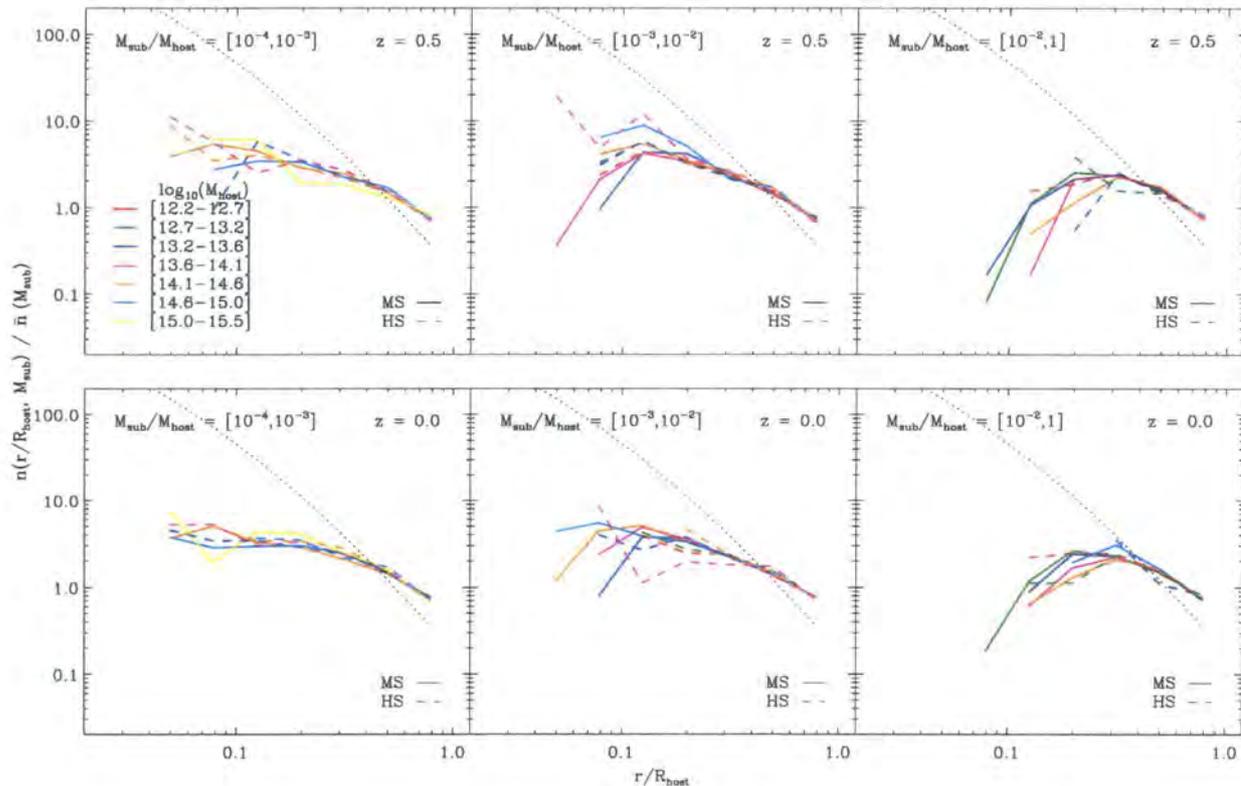


Figure 5.3: The number density of subhaloes relative to the mean within  $r_{\text{host}}$ , as a function of the distance to the centre of their host halo, in units of the radius of the host halo,  $r_{\text{host}}$ . Each column corresponds to a different fractional mass range for subhaloes, while the rows display the results at two separate redshifts. The solid and dashed lines show the density profiles from subhaloes in the MS and HS respectively. In both cases, the different colours correspond to subhaloes residing in haloes of different mass as shown in the legend. The black dotted lines in each panel indicate the mean dark matter density profile of haloes in our simulations. Results are shown only for subhaloes resolved with at least 200 particles. The top row shows results for  $z = 0.5$  and the bottom row for  $z = 0$ .

mass.

These dissimilar density profiles for different subhalo masses have a simple dynamical explanation (e.g. Tormen et al., 1998; Nagai and Kravtsov, 2005). Once a halo falls into a more massive system, dynamical friction and tidal stripping start to act. The accreted subhalo will rapidly be stripped of its outer layers and will lose a significant fraction of its mass during the first pericentric passage. This mechanism naturally differentiates the radial distribution of substructures of different masses: massive structures sink more rapidly due to dynamical friction and, as a result, also lose mass more rapidly by tidal stripping. Therefore they do not survive long in the central regions, in contrast to small subhaloes. The massive subhaloes which are present in the halo most have been accreted more recently than the average low mass subhalo. The timescale for dynamical friction depends on the relative mass of the subhalo and its host halo, not on their absolute values, which would explain the approximate independence of the distribution on the host halo mass.

### 5.3.4 Angular distribution of subhaloes

To end this section we investigate the angular distribution of subhaloes within dark matter haloes. Previous work has examined the relationship between the angular distribution of substructures and the *shape* of the host halo (Tormen, 1997; Libeskind et al., 2007; Knebe et al., 2008a,b). Here, we examine instead the orientation relative to the *spin axis* of the host halo. Fig. 5.4 shows the probability distribution function of the cosine of the angle between the angular momentum vector of the host halo and the vector joining its centre with that of the subhalo. We show results for two separate ranges of subhalo mass: subhaloes with mass smaller than 2% of the host halo mass (dashed lines) and those with masses greater than 2% (solid lines). We distinguish different host halo masses by different colours, and show different redshifts in different panels. Note that we only display results for the MS simulation for clarity.

As shown by Bett et al. (2007), the accuracy of the measurement of spin direction in the MS degrades significantly (uncertainty  $> 15$  deg) for haloes resolved with fewer than 1000 particles or for those where the spin magnitude,  $|j|$ , is such that:

$$\frac{|j|}{\sqrt{G M_{\text{host}} r_{\text{host}}}} < 10^{-1.4}, \quad (5.3)$$

where  $G$  is Newton's gravitational constant. Although the inclusion of haloes that do not

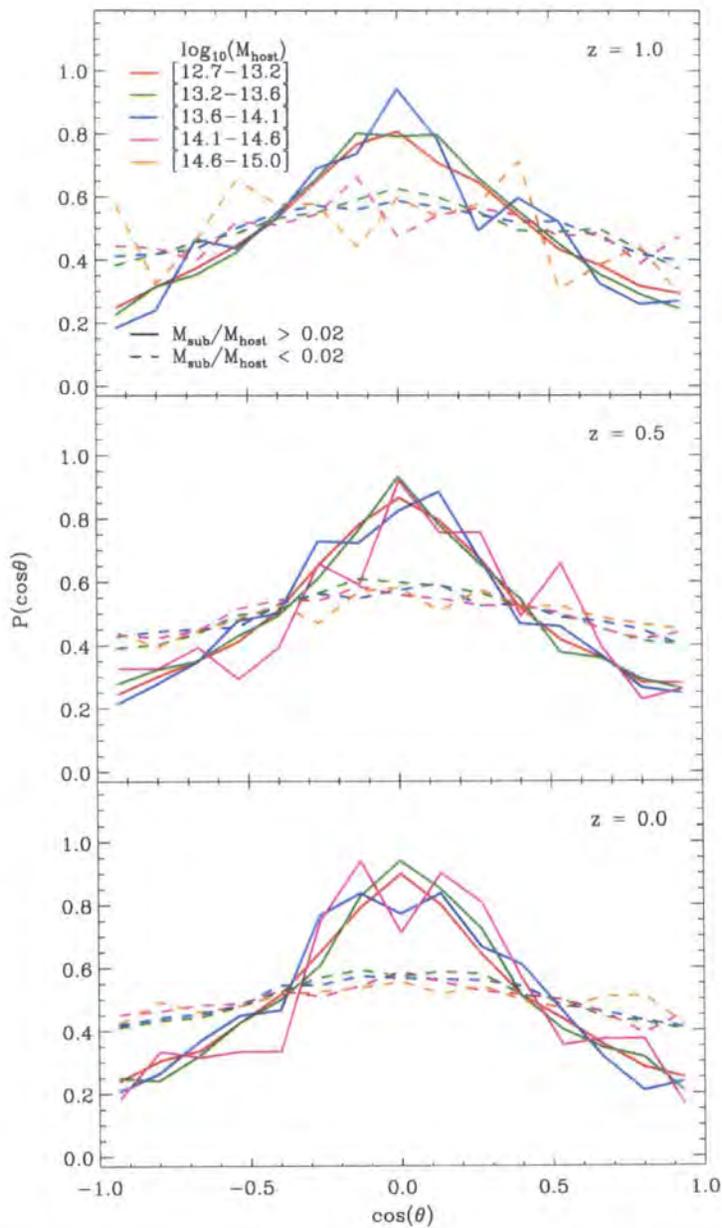


Figure 5.4: The probability density distribution of the cosine of the angle  $\theta$  between the angular momentum vector of the host halo and the vector joining its centre with that of the subhalo. The lines in each panel display the distribution for subhaloes in two different mass bins:  $0.02 < M_{\text{sub}}/M_{\text{host}} < 1$  (solid lines) and  $0.004 < M_{\text{sub}}/M_{\text{host}} < 0.02$  (dashed lines). Lines of different colour indicate subhaloes residing in host haloes of different masses, as shown in the legend. An isotropic angular distribution corresponds to a horizontal line.

satisfy these criteria does not seem to affect our results quantitatively, we have chosen to show only those haloes that met these requirements, so that the angle relative to the spin axis can be reliably determined.

We see from Fig. 5.4 that the angular distribution of subhaloes tends to be aligned perpendicular to the spin axis of the host halo. (We remind the reader that in this plot, an isotropic angular distribution would correspond to a horizontal line, while a distribution aligned at 90 deg to the spin axis will peak around  $\cos\theta \sim 0$ .) The strength of this alignment effect depends on the fractional subhalo mass,  $M_{\text{sub}}/M_{\text{host}}$ , being much stronger for higher mass subhaloes. We also see that the angular distribution for a given  $M_{\text{sub}}/M_{\text{host}}$  is almost independent of the host halo mass and the redshift (Kang et al., 2007, see also).

We can understand this behaviour qualitatively as reflecting the growth of haloes by the accretion of dark matter (in halos or more diffuse form) along filaments. The central regions of haloes acquire most of their angular momentum at a relatively late stage from the orbital angular momentum of this infalling material, and so will tend to have spin axes perpendicular to the current filament (e.g Shaw et al., 2006; Aragón-Calvo et al., 2007). On the other hand, insofar as the subhaloes “remember” the direction from which they fell in once they are orbiting inside the host halo, then their spatial distribution will tend to be aligned with the filament from which they were accreted, and so will be perpendicular to the spin axis. We can also understand the dependence of the strength of this alignment on subhalo mass in this picture. Subhaloes with large  $M_{\text{sub}}/M_{\text{host}}$  on average have been orbiting in the host halo for less time than haloes of lower  $M_{\text{sub}}/M_{\text{host}}$ , due to the combined effects of dynamical friction (which causes higher mass subhaloes to sink faster) and tidal stripping (which converts high-mass subhaloes to low mass). We expect subhaloes increasingly to lose memory of their initial infall direction the longer they have orbited in the host halo (which in general is lumpy and triaxial). Since high  $M_{\text{sub}}/M_{\text{host}}$  subhaloes have undergone fewer orbits, their current angular distribution should be more closely aligned with their infall direction, and therefore with the current filament, compared to subhaloes of lower mass.

Our results seem generally consistent with previous simulation results on the alignment of the subhalo distribution with the shape of the host halo, and the relationship between the shapes and the spin axes of halos. Tormen (1997) found that the angular distribution of subhaloes as they fall into a host halo (crossing through  $r_{200}$ ) is anisotropic, and tends to be aligned along the major axis of the host halo. Previous studies (e.g. Knebe et al.,

2004; Zentner et al., 2005; Libeskind et al., 2007) found that the angular distribution of subhaloes within a host halo is aligned along the major axis of the host halo. On the other hand, Bett et al. (2007) showed that the angular momentum of a halo is generally aligned with its minor axis and perpendicular to its major axis. Putting these results together, we would expect the subhalo distribution to be aligned perpendicular to the spin axis of the host halo, but ours is the first study to demonstrate this directly, and also to demonstrate that the strength of the alignment depends on subhalo mass.

## 5.4 Mergers between subhaloes

As we have seen, once a halo is accreted by a larger one, its outer layers are rapidly stripped by tidal forces. However, the core generally survives the accretion event and can still be recognized as a substructure or satellite subhalo within the host halo for some time afterwards. Furthermore, not only may the main infalling halo survive, but also substructures within it. In this case, there are substructures inside substructures.

While orbiting inside the halo, dynamical friction causes the orbit of a subhalo to lose energy and to sink towards the centre of the host halo. As the subhalo sinks, it suffers further tidal stripping. Eventually, the subhalo may be totally disrupted: there is a merger between the satellite subhalo and the central subhalo. Nevertheless, on its way to destruction, a subhalo can survive for several orbits during which it may experience a merger with another satellite subhalo. In the following subsections we will investigate the merging of these substructures.

The interaction between subhaloes was previously investigated in cosmological simulations by Tormen et al. (1998), who studied the rate of penetrating encounters between satellite subhaloes, but not the merger rate. Makino and Hut (1997) derived an expression for the merger rate between subhaloes in galaxy clusters based on an entirely different approach, motivated by the kinetic theory of gases. In this case, the merger rate per unit volume between halos of mass  $M_1$  and  $M_2$  is  $R_{\text{merge}} = n_1 n_2 \sigma(v_{12}) v_{12}$ , where  $n_1$  and  $n_2$  are the respective number densities,  $v_{12}$  is the relative velocity, and  $\sigma(v_{12})$  is the merger cross-section. They used N-body simulations of isolated spherical halos to derive merger cross-sections for equal-mass halos as a function of their relative velocity, and then assumed that mergers in clusters occurred between pairs of subhaloes drawn from random uncorrelated orbits, with a Maxwellian distribution of relative velocities. The Makino and Hut expression was then extrapolated to the case of unequal subhalo masses and

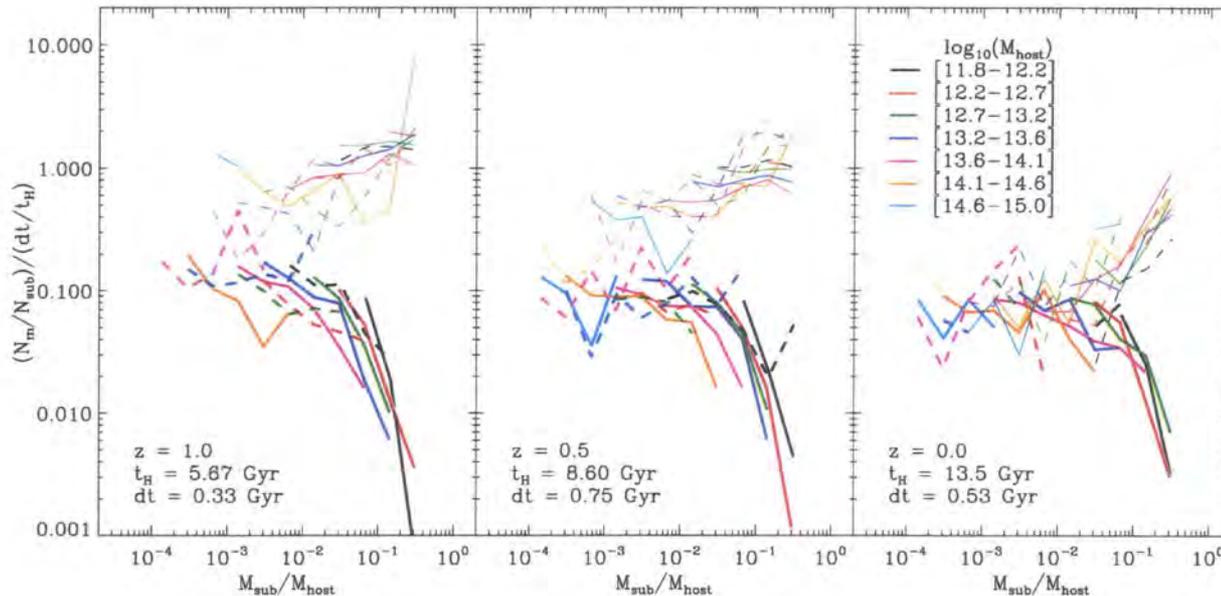


Figure 5.5: The mean number of satellite mergers per subhalo and per unit of time relative to the age of the universe, as a function of the mass of the progenitor of the less massive object involved in the merger. Two cases are displayed: the number of satellites destroyed or merging with the main substructure (top thin lines) and the number of mergers between two satellites (thick bottom lines). The solid lines show results from the MS while the dashed lines show results from the HS. As indicated by the legend, in both cases, coloured lines represent results for haloes of different mass. The three panels are for three different redshifts:  $z = 1.0$ ,  $z = 0.5$  and  $z = 0$ . Note that in the case of the merger between a satellite and a central structure, we show examples involving subhaloes of at least 200 particles, but we reduce the limit to 50 particles in the case of mergers between two satellites. In each panel, the legend states the redshift, the age of the universe,  $t_H$ , and the time interval,  $dt$ , over which we measure the rates.

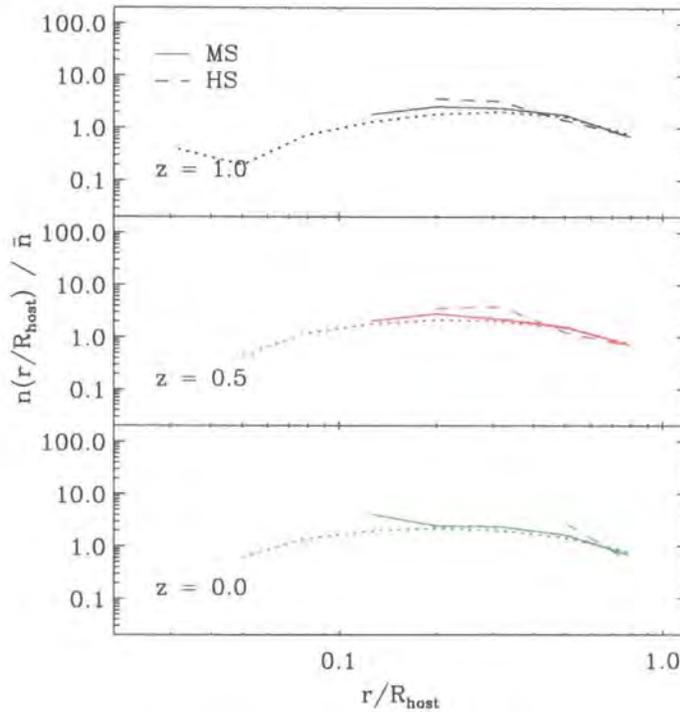


Figure 5.6: The number density of subhalo-subhalo mergers relative to the mean density of subhaloes within  $r_{200}$  as a function of the distance to the centre of the host halo. The results from the MS are shown by solid lines while the results from the HS are shown by dashed lines. In each subpanel the dotted lines show the radial distribution of all subhaloes (regardless of whether they are merging or not) in the MS. Mergers involving subhaloes resolved with at least 50 particles are included in the plot.

incorporated into a semi-analytical model of galaxy formation by Somerville and Primack (1999) and Hatton et al. (2003). We will investigate below whether the Makino and Hut (1997) kinetic theory approach has any applicability to subhalo mergers in a realistic cosmological context.

#### 5.4.1 Subhalo merger rate

Fig. 5.5 shows the mean merger rate of satellite subhaloes, plotted against the fractional mass of its progenitor. This is the mass of the satellite before accretion divided by the mass of the host halo at the time of the merger. The rate is normalized per subhalo, with time in units of the age of the universe at that redshift. This normalized rate is thus roughly equal to the probability that a satellite subhalo will merge over one Hubble expansion time. A rate higher than unity indicates that the process happens on a timescale shorter than a Hubble time. There are two sets of curves in this figure: (i) the thinner, higher amplitude lines which show mergers between a satellite and a central subhalo, as a function of the subhalo mass, and (ii) the thick lines which correspond to satellite-satellite mergers, plotted as a function of the mass of the smaller subhalo. As in previous plots, different line colours show different host halo masses, and different line styles (solid and dashed) show the two simulations used.

We see from Fig. 5.5 that over most of the subhalo mass range resolved by our simulations (for  $M_{\text{sub}}/M_{\text{host}} \gtrsim 10^{-3}$ ), it is more likely for a satellite subhalo to merge with the central subhalo than with another *more massive* satellite subhalo. For instance, at  $z = 1$ , taking into account all host haloes, there are 17155 satellites which merge with a central subhalo over one timestep, while the number of satellites involved in a merger with another satellite over the same period is 509, a ratio of 40 : 1. The situation is similar at  $z = 0$  even though the ratio decreases to 6 : 1 (1645 vs 290). In general, the likelihood of both merger rates slightly decreases at lower redshifts. This may reflect the slower build-up of structure (relative to the Hubble time) as the universe becomes dominated by vacuum energy.

As we consider smaller subhalo masses, we see a decrease in the destruction rate (see the appendix for a discussion of overmerging effects due to insufficient mass resolution). This may be due to the inefficiency of dynamical friction for low mass structures. On the other hand, there is an increase in the satellite-satellite merger rate as the subhalo mass decreases. Presumably this is due to the increasing number of potential merger partners, reflecting the form of the subhalo mass function. Additionally, the abundance of both

types of mergers is similar in the range  $10^{-3} < M_{\text{sub}}/M_{\text{host}} < 10^{-2}$ . Unfortunately, at this point our results from low mass haloes become limited by resolution (i.e. we cannot identify smaller substructures) and the results from high mass haloes become dominated by Poisson noise (i.e. less than one merger event in the whole simulation). Over the range that is reliably covered, we can see no strong systematic differences in Fig. 5.5 between the results derived from host haloes of different masses. This agreement is quite remarkable given the relatively large dynamical range resolved in the simulations.

A merger between two objects is not always a straightforward quantity to define in numerical simulations. The problem originates from the fact that any definition is intrinsically linked to the mass and time resolution of the simulation. For instance, if in a higher resolution simulation we identify the remnant of a subhalo down to a smaller mass threshold, then the mass ratio of the merger, as well as the time at which it happens, could, in principle, disagree with the values measured in a lower resolution simulation. Similarly, with better time resolution, one could follow the mass loss of a subhalo more accurately which, in principle, could also change the measured mass ratio of the merger. To avoid these problems, we have chosen to use in Fig 5.5 the mass of the satellite before accretion, rather than the mass at the moment of the merger.

For all these reasons it is very important to note the agreement in Fig. 5.5 between the results from the MS (solid lines) and those from the HS (dashed lines). This agreement gives us confidence that our results are not sensitive to mass resolution. (Note that this is not true for subhaloes resolved with fewer particles as shown in the appendix.) Furthermore, the weak dependence of the quantities plotted in Fig. 5.5 on host halo mass confirms this conclusion. In practice, a subhalo of  $M_{\text{sub}}/M_{\text{host}} = 0.1$  in a host of  $10^{12} h^{-1} M_{\odot}$  exhibits the same behaviour as a subhalo of the same fractional mass but in a halo of  $10^{14} h^{-1} M_{\odot}$  even though the latter is resolved with 100 times more particles. This is quite remarkable.

One of the reasons for the insensitivity to mass resolution comes from our definition of a merger (see §5.2.2). We do not tag an event as a merger when we cannot identify the subhalo anymore, but rather when it has lost a fixed fraction of its most bound mass. This definition responds more to dynamical processes than to numerical ones.

The implications of discrete time measurements are less clear for our definition of a merger. As an example, consider the case of very poor time resolution, and a halo that is just about to fall into a larger one. If tidal forces stripped off more than 95% of its mass before the next snapshot, then we would have identified this event as a merger. On the

other hand, if the time resolution were good enough, we could have identified the subhalo at intermediate stages, updating its mass and the corresponding most bound 10 percent. As long as stripping does not occur on a timescale much shorter than the time resolution, it is even possible to imagine that the line of descendants continues indefinitely. However, since a merger is not a discrete event, better time resolution does not necessarily imply a more accurate determination of a merger. With infinite time resolution, we would follow most of the merging process down to the point when mass resolution becomes important, i.e. every subhalo disruption would be caused by lack of mass resolution.

However, the typical timescale for dynamical friction and tidal disruption is  $T_{\text{fric}} \sim t_H$  for  $M_{\text{sub}}/M_{\text{host}} \sim 0.1 - 0.2$  (Jiang et al., 2008), i.e. much longer than the time spacing of our simulation outputs ( $\sim 300\text{Myr}$ ). Furthermore, subhalo mergers seem to take place very fast. Both these factors suggest that time resolution is not an issue for this study. In fact, we have checked that our results do not change if we choose snapshots that are twice as widely spaced as those used to build the merger trees. Nevertheless, we advise the reader to keep these limitations in mind.

#### 5.4.2 Characterization of subhalo-subhalo mergers

In most cases the subhalo-central merger occurs very close to the potential minimum of the host halo. The spatial location of satellite-satellite mergers, on the other hand, has a very distinctive distribution. In the following subsection we investigate this further.

##### 5.4.2.1 Radial distribution of satellite-satellite mergers

First, in Fig. 5.6 we look at the spherically averaged radial distribution of satellite-satellite mergers. The figure shows the number density of mergers, relative to the mean density of subhaloes within  $r_{200}$ , as a function of the distance to the centre of the halo. We also display, as dotted lines, the distribution of all the substructures from the MS<sup>1</sup>.

At every redshift plotted, the radial distribution of mergers is proportional to the radial distribution of subhaloes. This implies that most of the mergers between subhaloes

---

<sup>1</sup>At first sight, the distribution of all substructures seems to disagree with the results of Section 5.3.3. Since the subhalo population is dominated by small mass objects, one would naively expect the distribution of all substructures to follow that of the smallest subhaloes; as seen in Fig. 5.3, this has a slope which is always negative. However, in practice, the dominant effect is the high abundance of low mass host haloes in which only massive substructures can be resolved. As a result, the distribution of all subhaloes in the MS resembles the distribution of the most massive substructures

do indeed occur in the outer regions of the host halo. Note that in these regions the background density is lower than in the inner regions, making it easier to identify subhaloes. For this reason, we can follow satellite-satellite mergers down to structures resolved with 50 particles, as opposed to the minimum of 200 particles we require for central-satellite mergers.

Our results do not appear consistent with the naive expectation from a gas kinetic theory approach that the number density of mergers should be proportional to the number density of subhalo pairs, i.e.  $R_{merge} \propto n_{sub}^2$ . This discrepancy indicates that most of the satellite-satellite mergers do not occur because of random encounters between two unrelated substructures. We investigate this idea further in the following subsection, where we look back at the orbits of the subhaloes that merge.

#### 5.4.2.2 Orbits of merging satellites

Fig. 5.7 shows the distribution function of the separation angle  $\theta$  between the progenitors of subhaloes involved in a merger. The angle is measured from the centre of the host halo in which the merger is going to take place, at the last snapshot in which both subhaloes were identified outside the halo that later hosts the satellite-satellite merger. It thus represents the angle between the subhaloes at the time they fall into the host halo. The first point to note is that the distribution seems to be universal in the sense that it is roughly independent of the mass of the host halo. (We have also checked that it is roughly independent of redshift.) However, the most important feature is that the distribution is clearly dominated by small separation angles. About 65% of the mergers occur between subhaloes that were separated by less than 30 deg at the moment of accretion. (This percentage increases to 73% for an angle of 43 deg.) This demonstrates that the mergers are mostly between two or more systems that were already dynamically associated before they fell into the larger system. If the gas kinetic theory approach of Makino and Hut (1997) applied to this case, then the mergers would be between subhaloes on random orbits, and we would expect a more uniform distribution in  $\cos\theta$ . (It would not be completely uniform since the subhalo population is not isotropic, as shown in Fig. 5.4.)

More information about the orbits of merging subhaloes is given in Fig. 5.8, where we display three representative examples of the two most common configurations of a satellite-satellite merger. These examples correspond to real sequences found in the MS. The plot tracks the position of substructures up to the snapshot of the merger (which happens at the rightmost position), starting on the left, 9 snapshots earlier. We show

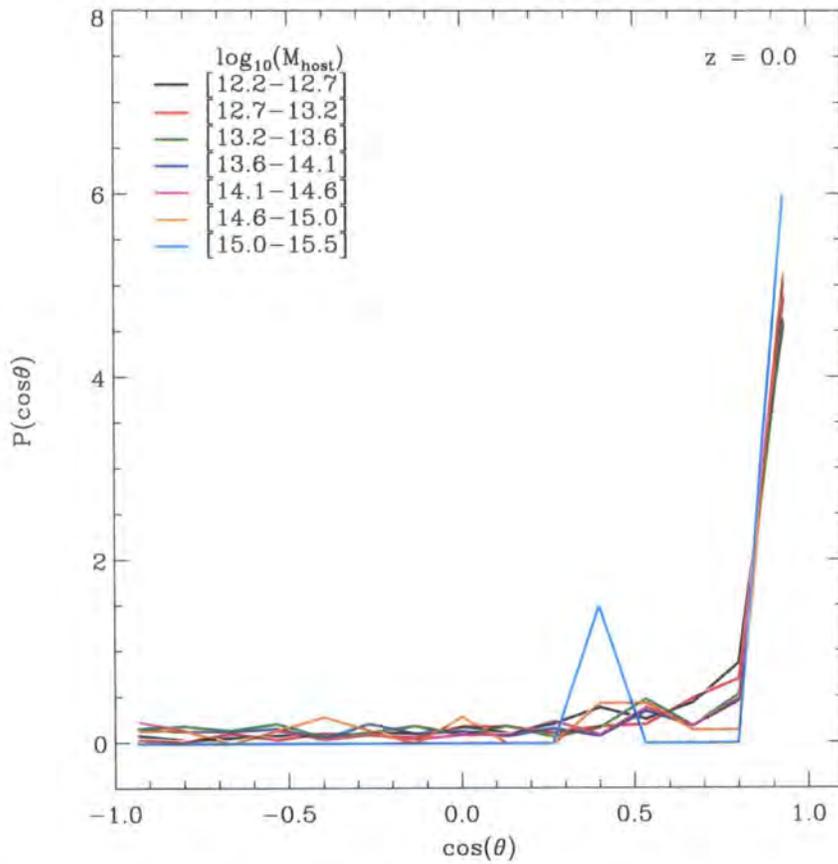


Figure 5.7: Probability distribution of the cosine of the separation angle  $\theta$  between the progenitors of two substructures that are going to merge. The separation angle is measured at the last snapshot in which the subhaloes were identified outside the halo that hosts the satellite-satellite merger. Lines of different colours indicate mergers happening in haloes of different masses as indicated in the legend.

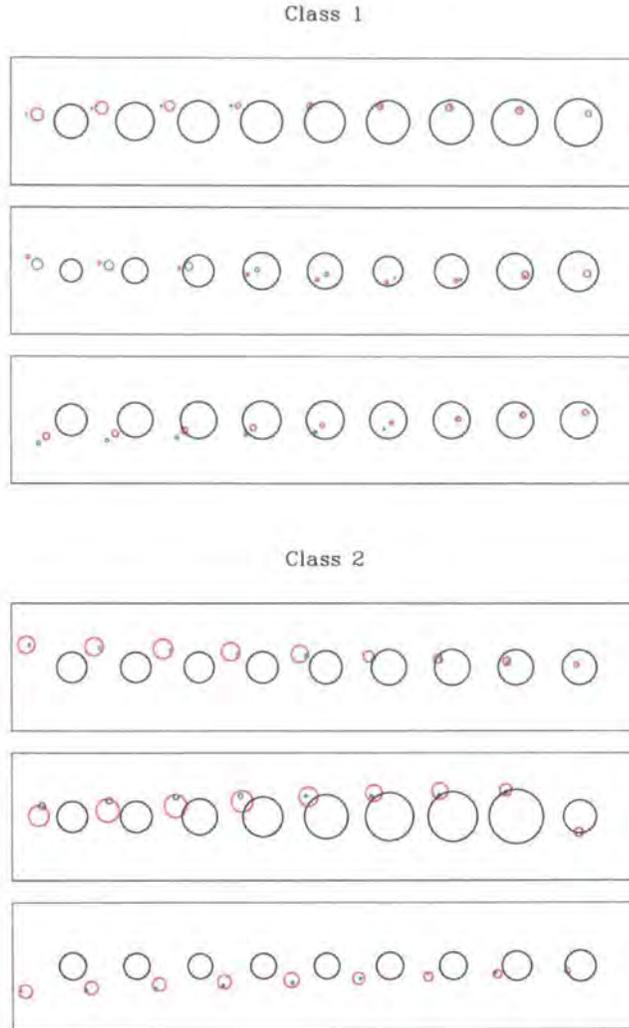


Figure 5.8: Three representative examples extracted from the MS for each of the two most common configurations between two satellite subhaloes that merge. The plots show the relative positions of the host and satellite halos in a time sequence, with time increasing from left to right. The black circles correspond to the halo that hosts the merger, while red and green circles show the positions of the satellites involved in the merger. The circles' radii are proportional to the half mass radius of each substructure. Class 1: in this case the satellites were part of two separate haloes (red and green circles) before they were accreted into a larger halo (black circles). Class 2: both substructures belonged to the same halo before it was accreted into the larger structure which hosts the merger.

as a black circle the halo that hosts the satellite-satellite merger and, as green and red circles, the progenitors of the subhaloes involved in the merger. The red circle at the end of the sequence indicates the subhalo resulting from the merger. The radii of the circles are proportional to the half-mass radius of the subhalo.

The two most common configurations are as follows: Class 1: the progenitors of the subhaloes correspond to two separate haloes which were accreted at approximately the same time. Note that, as shown by Fig. 5.7, these haloes were spatially close at the time of accretion. Class 2: the merger occurs between two substructures that were part of the same halo before it fell into the host halo. In other words, there is a halo that contains two substructures which survived the accretion and subsequently merged. The merger event which started outside the main halo is completed inside it, as a subhalo-subhalo merger.

Most subhalo mergers occur between substructures that are accreted close together both in time and location. Generally, they were already part of the same system before it was accreted into a larger one, or were part of two separate haloes that were about to merge. This is probably a requisite for a subhalo merger to occur. The potential generated by the other satellite has to be at least comparable to that of the main halo. Hence, satellites accreted at different angles will follow relatively independent dynamical histories and are much less likely to merge.

#### 5.4.2.3 The mass ratio of subhalo mergers

In Fig. 5.9, we inspect the relative masses of the satellite subhaloes which merge. The  $x$ -axis indicates the mass of the smaller subhalo and the  $y$ -axis shows the mass of the larger one. Interestingly, we find that, for the range of host halo masses plotted, the most common merger is that between two substructures of dissimilar masses,  $M_{\text{sub},1} \sim 10 \times M_{\text{sub},2}$ . Note that this trend is contrary to the naive expectation whereby the mergers are simply proportional to the abundance of substructures, in which case the maxima would be located around the line  $M_{\text{sub},1} = M_{\text{sub},2}$ . However, it is roughly consistent with the idea that substructure mergers happen between two structures that were part of the same halo before accretion. For instance, if the most common merger happens between the main subhalo and its most massive substructure, then, as we have seen, we would expect to find a mass ratio of 1:25 (see Fig. 5.2) and the maxima of Fig. 5.9 along  $M_{\text{sub},1} \sim 10 \times M_{\text{sub},2}$

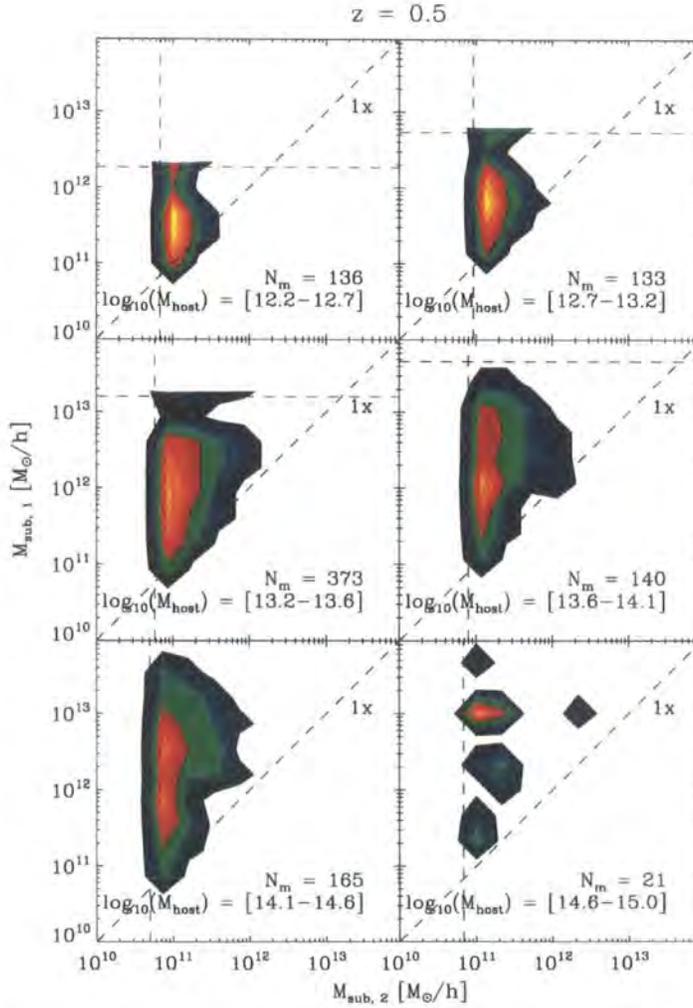


Figure 5.9: Contour plot showing the logarithm of the number of satellite-satellite mergers as a function of the masses of the merging subhaloes at  $z = 0.5$ . The  $x$ -axis indicates the mass of the smaller subhalo while the  $y$ -axis indicates the mass of the larger subhalo. The different panels show the results for host haloes of different masses as indicated on each panel. The numbers in the bottom right show the number of mergers displayed in each panel. The vertical dashed lines indicate the 200 particle limit and the diagonal lines correspond to a 1:1 ratios between the masses of the two subhaloes. The horizontal lines show the mass limit on the more massive participant imposed by the choice of mass bin.

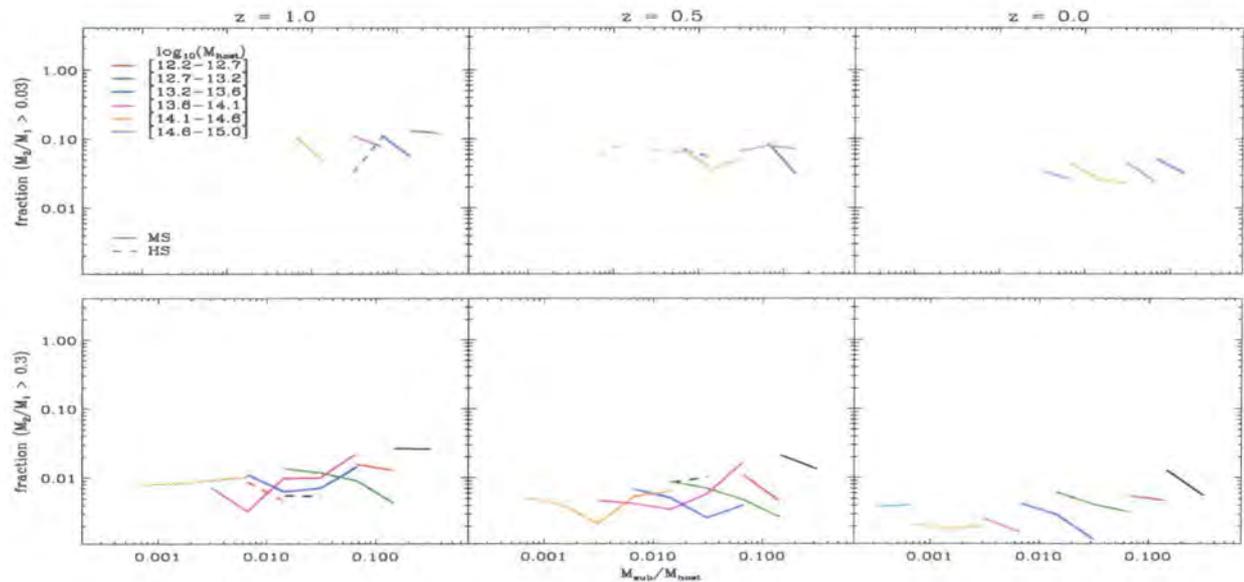


Figure 5.10: The fraction of substructures that have experienced a merger with another substructure since the time of accretion into the current host halo. The  $x$ -axis gives the subhalo mass at the redshift shown, while the ratio  $M_2/M_1$  on the  $y$ -axis is for the two progenitors of the subhalo at the time they merged. The results from the MS are shown by solid lines while the results from the HS are shown by dashed lines. The coloured lines represent the data from haloes of different masses, as indicated by the key. The two rows correspond to different mass ratios between the subhalo progenitors involved in the merger:  $M_{\text{sub},2} > 0.03 M_{\text{sub},1}$  (top row) and  $M_{\text{sub},2} > 0.3 M_{\text{sub},1}$  (bottom row) where  $M_{\text{sub},2}$  refers to the larger satellite involved in the merger. The three panels display the results for substructures identified at redshifts  $z = 1, 0.5$  and  $0$  respectively.

### 5.4.3 Merger probability since accretion

Finally, in Fig. 5.10 we plot the fraction of subhaloes at a given redshift that have had a merger with another satellite subhalo since accretion into the current host halo. The top panels show mergers between satellites with a mass ratio greater than 0.03, i.e. in which the less massive subhalo has, at least, 3% of the mass of the larger one. In the bottom panels we consider mergers between subhaloes with more similar masses: the minimum mass ratio is 0.3.

The fraction of current subhaloes which have experienced a merger in the past is a quantity strongly affected by resolution. For instance, in the history of a subhalo resolved with 1000 particles, because of our 200 particle mass cut on subhaloes, we can only record mergers with other subhaloes which account for at least one fifth of the final subhalo mass. On the other hand, if our current subhalo is resolved with 10000 particles, then a much wider range of merger mass ratios can be tracked. These considerations are further complicated by the fact that we expect the measured mass of a subhalo to be less than the mass of its progenitors at infall, due to tidal disruption and stripping; hence an object that is below our 200 particle limit at a particular redshift could have been above this mass cut when it experienced the subhalo-subhalo merger.

To improve statistics, whilst at the same time attempting to avoid building a resolution dependence into our results, we relax the particle number constraint on subhaloes for this exercise. At the redshift a subhalo is identified (i.e. the redshift plotted in Fig. 5.10), we consider subhaloes of 30 particles or more. At the redshift of the subhalo merger, the progenitors must both have 50 particles or more to be counted.

Fig. 5.10 shows that the probability of a subhalo merger is constant for subhaloes of different mass. About 1% percent of subhaloes have had a merger with another subhalo with a mass ratio  $> 0.3$ . For a mass ratio  $> 0.03$ , this fraction increases to  $\sim 10\%$ . We also note that these fractions show a weak decrease with redshift.

## 5.5 Summary and conclusions

We have used the Millennium simulation, together with a simulation which has 10 times better resolution but about 100 times smaller volume, to investigate the general properties of the substructures within dark matter haloes, including their merger rates. Our main findings can be summarized as follows:

In agreement with previous studies, we find that the mass function of low and interme-

diate mass subhaloes follows roughly a power-law. However, we also find an exponential cut-off in the mass function at high subhalo masses. We have provided an expression, Eq. 5.2, that describes this behaviour accurately. We also detect a small but systematic dependence of the number of subhaloes on the mass of the host halo. On average, at a given fractional mass,  $M_{\text{sub}}/M_{\text{host}}$ , high mass haloes contain more low and intermediate mass substructures than their less massive counterparts. In contrast, we find evidence that high mass haloes contain fewer high mass subhaloes than do low mass haloes. In spite of this, the fractional mass of the first, second and third most massive substructures is insensitive to the mass of the host halo and of the redshift.

We confirm that the radial and angular distributions of subhaloes are roughly independent of the host halo mass and redshift. However, we find that the radial distribution does depend on the subhalo mass relative to that of the host halo. The subhalo distribution is less concentrated than the dark matter, but the radial distribution of low mass subhaloes tends to be more concentrated than that of high mass subhaloes. This difference can be understood as resulting from the different efficiency of dynamical friction in subhaloes of different mass. On the other hand, these discrepancies between the radial distributions of low and high mass subhaloes disappear in the outer parts of the halo, as it have been seen in recent ultra-high resolution simulations of galactic halos (Springel et al., 2008).

The angular distribution of subhaloes tends to be aligned perpendicular to the spin axis of the host halo. This is probably due to an anisotropic mass accretion - mergers happen preferentially along filaments. The alignment is strong for the most massive subhaloes, but is much weaker for low mass substructures since, on average, they have spent a few orbital times inside the halo which would randomize their orientation.

We have found that satellite-satellite mergers do occur. Over most of the mass range resolved in our simulations, they are subdominant when compared with mergers between satellites and the central subhalo. However, we see some indication that satellite-satellite mergers dominate for the lowest mass subhaloes ( $M_{\text{sub}}/M_{\text{host}} < 10^{-3}$ ). As many other subhalo properties, the merger rates appear to be a function of the fractional subhalo mass only, and are independent of the particular host or subhalo mass.

The radial distribution of satellite-satellite subhalo mergers closely follows the radial distribution of subhaloes. This implies that most of the subhalo mergers happen in the outer layers of the halo. For the most part, these mergers involve subhaloes that are already dynamically associated before accretion into the main halo, i.e. they were either part of the same halo, or of two separate haloes that were accreted at similar times and

locations. At every redshift, most of these subhaloes which subsequently merged were closer together than  $30 \text{ deg}$  as seen from the centre of the halo that hosts the merger, at the time they fell in.

Finally, we find that a small fraction of the high-mass subhaloes has experienced a merger with another subhalo since accretion into the current host halo. The values depend on the mass ratio of the merger, but vary from a few percent for mass ratios greater than 0.3 to  $\sim 10\%$  for mass ratios greater than 0.03.

In spite of using some of the largest simulations to date, our results could still be affected to some extent by numerical resolution. Due to the rarity of the events we are trying to study, it is difficult to find a range of substructure and host halo masses where we have, at the same time, (i) enough particles to resolve substructures well, (ii) enough haloes to distinguish real trends from cosmic variance, and (iii) enough subhaloes to establish their properties and dynamics. Fortunately, as we have shown, many properties can be described as a function of only the fractional subhalo mass. In these cases we are observing the same system resolved with many different numbers of particles, so it is reassuring that we find the same trends for different host halo masses. This gives us confidence that these results are robust. On the other hand, quantities which scale with halo mass are much less reliable and could still be affected by resolution effects. Much larger simulations, currently beyond reach, will be needed to check them.

## Appendix: numerical effects

Numerical artifacts can pose serious problems in obtaining a robust estimate of different properties of the population of subhaloes. For instance, two-body encounters, particle heating, or force softening could easily dilute substructures that are not resolved with enough particles (Moore et al., 1996). These problems translate into an overestimation of the number of satellite-central subhalo mergers in each timestep.

Such a feature is clear in Fig. 5.11, which is similar to Fig. 5.5, but for satellite-central mergers only and including subhaloes with less than 200 particles. For these objects, we can see a strong disagreement between the merger rate of substructures in the simulations with different resolution which is manifest as an upturn in the curves. However, the upturn disappears for subhaloes with  $N > 200$  which is the limit set in this chapter.

An overestimation of the destruction rate also has implications for other quantities such as the abundance and radial distribution of subhaloes. For instance, the subhalo mass

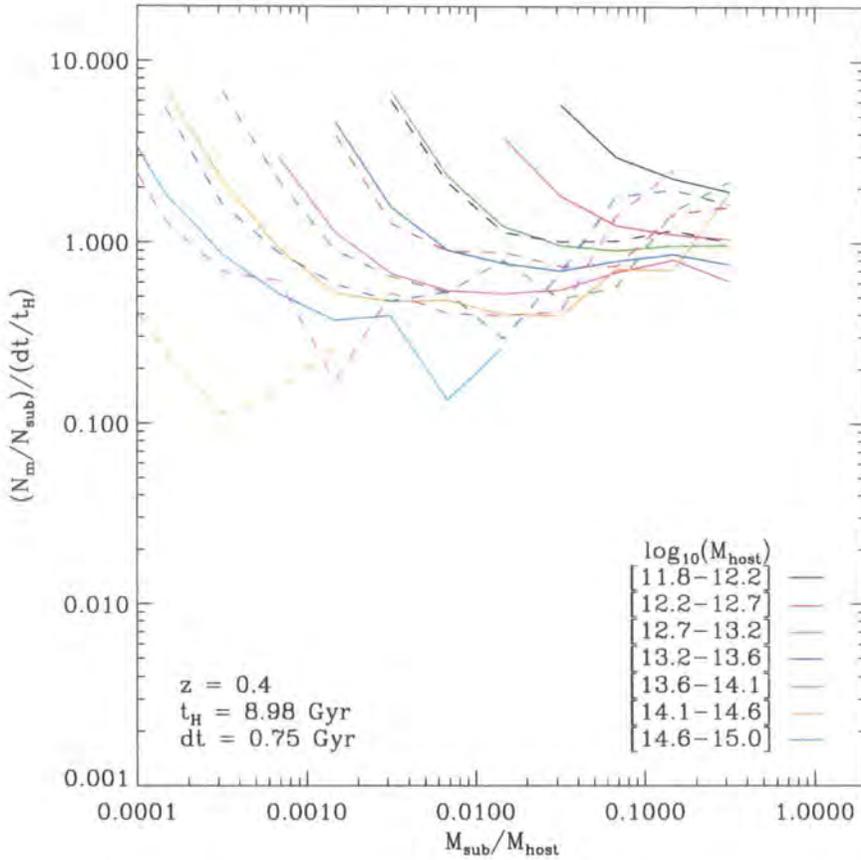


Figure 5.11: The mean number of satellite-central subhalo mergers per subhalo and per unit of time as a function of the subhalo mass. The solid lines show the results from the MS while the dashed lines show the result from the HS. The coloured lines represent the results from haloes of different mass, as indicated by the legend. Note that we display results from subhaloes with 20 particles or more. The upturn in  $N_m$  for low mass subhaloes is due to the inclusion of subhaloes resolved with fewer than 200 particles. Once the  $N > 200$  criterion is applied, the upturn disappears as shown in Fig 5.5.

function shows a cut-off at low masses compared with the expected power-law behaviour when we include subhaloes resolved with fewer than  $\sim 50$  particles. (This quantity is less affected since most of the haloes are in the outer layers of the halo.) On the other hand, the inner region of the radial distribution is more sensitive to these effects. Once subhaloes with fewer than 200 particles are included in Fig. 5.3 the distribution becomes less centrally concentrated.

Our convergence study indicates that 200 particles is the limit below which results are unduly affected by resolution. This is why we have adopted this minimum particle count throughout this chapter, except when otherwise stated explicitly. This choice should minimize finite-resolution effects.

# Chapter 6

## *Extending the dynamic range of low resolution dark matter simulations*

### 6.1 Introduction

Ever since the first measurements, the clustering of galaxies has been recognised as a powerful tool to investigate some of the most fundamental questions in cosmology and galaxy formation. After many years of advances on both observational and theoretical fronts, accurate information regarding the total matter content in the Universe, the amount of baryons, the formation and evolution of galaxies, and even the nature of gravity, has been extracted successfully from the analysis of the three dimensional distribution of galaxies.

In spite of this enormous contribution to physical cosmology, the usefulness of the galaxy distribution is far from over. New exciting windows have been proposed to explore the nature of the dark energy and primordial non-Gaussianities, such as the imprint of the Baryonic Acoustic Oscillations (BAO, e.g. Eisenstein et al., 2005; Cole et al., 2005; Gaztanaga et al., 2008), the form of redshift space distortions (Guzzo et al., 2008; Percival and White, 2008) and the clustering amplitude of dark matter haloes (Seljak, 2008; Matarrese and Verde, 2008; Carbone et al., 2008). In addition, higher order clustering could provide a fresh new view of the cosmological issues mentioned above.

A key step in achieving the proposed goals is to obtain a complete understanding of the systematic and statistical errors associated with each clustering measurement. An accurate treatment in the error estimations of not only the survey geometry and number density of objects, but of cosmic variance, nonlinear evolution, scale dependent bias, redshift-space distortions, and discreteness effects, rather than being an extra level of sophistication, become essential ingredients in modern data analysis. The importance of this is accentuated by the imminent arrival of the ever larger galaxy surveys (e.g.

Pan-STARRS, Dark Energy Survey, EUCLID).

One of the primary ways to study errors is through an accurate model of the experiment itself. For the case relevant here (the spatial distribution of galaxies), this is optimally achieved in a three step process. First, the halo clustering is predicted by following the evolution of dark matter particles in an N-body simulation. Second, the properties of galaxies within these haloes are predicted using empirical or semi-analytical recipes (Baugh, 2006). And lastly, the appropriate flux limit, sample selection, redshift completeness and the geometry of the survey need to be applied to the catalogues. The combination of these steps results in a realistic mock of the observational experiment from which different sources of uncertainties can be studied. Unfortunately, the first step in this process is computationally very demanding, particularly when large scales are probed. There are two main reasons: i) The large number of independent N-body simulations needed to assess the errors. An adequate estimation of the variance would require several dozens of realizations of the density field (e.g. a 10% error on the variance for a Gaussian distribution requires  $\sim 50$  realizations), and an order of magnitude more simulations to robustly compute the full covariance matrix. ii) The huge dynamic range required to resolve haloes which are likely to host the galaxies observed. For instance, in box sizes of a few gigaparsecs, only cluster-mass haloes can typically be identified robustly (Fosalba et al., 2007; Teyssier et al., 2008). Although algorithms and computer hardware are constantly improving, finite computational resources impose a limit on N-body simulations, and carrying out ensembles where future experiments can be recreated is currently a prohibitively expensive computational price.

Several authors have devised a number of different algorithms to predict galaxy clustering efficiently and to overcome the difficulties stated above. Amongst the simplest are realizations of Gaussian or log-normal density fields (e.g. Mesinger and Furlanetto, 2007; Percival et al., 2001). More complicated ideas have been implemented using second order perturbation theory (Monaco et al., 2002; Scoccimarro and Sheth, 2002). In a different approach, the use of dark matter particles to mimic galaxy clustering has been followed in several studies by using a prescription based on the local density (Cole et al., 1998; Cabre and Gaztanaga, 2008). However, none of them has achieved completely the combination of simplicity and accuracy desirable when modelling a given cosmological experiment.

The objective of this chapter is to present a proof of concept of a new scheme to create efficiently mock catalogues of the large scale distribution of galaxies. Our approach uses the density field extracted from low resolution simulations to predict a halo population

whose properties are derived from higher resolution simulations. In §6.2 we provide details of our method along with the theoretical motivation. In §6.3 we apply our algorithm to an ensemble of N-body simulations to investigate the limitations of the procedure. The haloes created by this algorithm can be combined with higher mass counterparts, which are identified directly in the low resolution simulations, to extend the range of halo masses included in the simulations. The resulting hybrid halo catalogue can be fed into a semi-analytic galaxy formation model or combined with a halo occupation distribution (HOD) model from which the final mock galaxy catalogue is generated. To illustrate the feasibility of the idea, our procedure is shown in action in §6.4 where we use a HOD to predict the errors in the clustering of LRGs. Finally, in §6.5, we present a brief summary of our findings.

## 6.2 Method

Here we present the procedure that we employ to generate a halo population for low resolution dark matter simulations. In the first subsection we give the motivation and main ideas behind the method (§6.2.1) and in the second subsection (§6.2.2) we detail the steps to be followed for a practical implementation.

### 6.2.1 Theoretical Motivation

Assuming that the abundance of haloes at a given position is a function of the local underlying nonlinear dark matter density (cf. §4.2), we can write the smoothed density field of haloes of mass  $M$  as:

$$\delta_{\text{h}}(x|R, M) = f_M(\delta_{\text{dm}}(x|R)), \quad (6.1)$$

where  $f_M$  is a smooth and arbitrary function (in principle different for haloes of different mass), the  $\delta(x)$  is the density contrast defined as  $\rho(x)/\langle\rho(x)\rangle - 1$  and the subscripts h and dm refer to the density field of haloes and dark matter respectively.  $R$  is the scale on which both density fields are smoothed.

On large enough scales, the dark matter density approaches the mean value,  $|\delta_{\text{dm}}(x|R)| \ll 1$ , which allows us to express Eq. 6.1 as a Taylor series about  $\delta_{\text{dm}}$  (see e.g Fry and Gaztanaga, 1993):

$$\delta_h(x|R, M) = \sum_{k=0}^{\infty} \frac{b_k(M)}{k!} \delta_{\text{dm}}^k(x|R). \quad (6.2)$$

Note that the coefficients  $b_k$  are usually referred to as the bias parameters. In particular,  $b_1$  is also known as the linear bias. Now it is straightforward to write down an expression for the expected number density of haloes in a given mass in a region in which the dark matter density field has been smoothed :

$$N_h(x|R) = \langle N_h(x|R) \rangle \left[ 1 + b_1 \delta_{\text{dm}}(x|R) + \frac{b_2}{2} \delta_{\text{dm}}^2(x|R) + O(\delta_{\text{dm}}^3(x|R)) \right], \quad (6.3)$$

where the brackets  $\langle \rangle$  represent an average over all different smoothing regions. As we will discuss below, it is possible to use this expression to construct a halo density field with the right abundance, clustering properties expected in a dark matter simulation.

### 6.2.2 Implementation

It can be seen clearly that the expected abundance of haloes at a given location (Eq. 6.3), is a function of 3 quantities: i) the dark matter density field at the same location, ii) the mean number density of haloes of the same mass and iii) the bias parameters. The core of our method is that it is possible to extract the dark matter density field directly from low resolution simulations and both the bias parameters and the mean number of haloes can be computed from a higher resolution N-body simulation. And, as a consequence of putting these ingredients together, a population of dark matter haloes, which spans an arbitrarily wide range of masses, can be easily created for low resolution dark matter simulations.

This artificial population of haloes (which we will refer to as sub-resolution haloes) has, by construction, the right abundance and clustering on scales larger than the smoothing scale. In fact, not only the two-point statistics are reproduced but, *in principle*, the correct volume averaged higher order statistics are also present in the distribution of haloes (as can be seen from Eq. 6.2).

There are, naturally, limitations in the sub-resolution halo catalogues. First, note that our expressions are only valid when  $\delta_{\text{dm}} \ll 1$  which sets, in practice, the minimum possible smoothing scale and ultimately, the smallest scale on which the clustering properties can be reproduced. Second, in a practical implementation, Eqs. 6.2 and 6.3 have to be truncated at a given order which creates two problems. i) As a natural consequence,

statistics of orders higher than the truncation can not be reproduced. ii) In regions where the dark matter density field is low, Eq. 6.3 could predict a negative number of massive haloes! e.g. to first order,  $b(M) > 1$  and  $\delta_{\text{dm}} < -1/b$  implies  $\delta_{\text{hh}} < -1$ . Consequently, we expect our procedure to break down for haloes more massive than  $M_*$ . We will investigate these restrictions in more detail in the following sections when we present our algorithm in action.

## 6.3 Results

We now apply and test the procedure outlined in the previous section. In 6.3.1 we provide details of the implementation along with some general characteristics of the resulting halo catalogues. In section 6.3.2 we present three basic tests and a comparison with haloes identified directly in a high resolution N-body simulation.

### 6.3.1 The sub-resolution halo catalogue

Following the algorithm described in §6.2 we have computed a sub-resolution halo catalogue for the three outputs ( $z = 0, 0.5$  and  $1$ ) of each of the 50 simulations in the L-BASICC ensemble (see section 2.2.1 for details of these simulations). Note that one of these simulations simulates exactly the same initial density field as the BASICC simulation described in Chapter 2.

The application of our algorithm consists of three basic steps. The first one is the construction of the dark matter density field in the simulations. We have achieved this by placing particles on a grid using the nearest grid point mass assignment scheme (Hockney and Eastwood, 1981). We use a grid of  $256^3$  cells (the cell size is  $5.2 h^{-1} \text{Mpc}$ ) which is set so that  $\langle \delta^2 \rangle \sim 1$ . We therefore expect to obtain an inaccurate estimation of the halo clustering on scales smaller than a few grid cells.

The next step is to tabulate the bias and number density of haloes as a function of mass. We extract such relationships from the higher resolution BASICC simulation (see §2.2) in logarithmic mass bins of width  $\Delta \log_{10} M = 0.426$ . Both quantities are computed by smoothing the haloes and dark matter field in  $256^3$  cells and then averaging the values across the grid.

Finally, these three quantities are brought together to compute the expected number density of haloes on every point of the grid. There are several details of the placement of haloes that are worth noting. i) The actual number of haloes in each cell is generated from

a Poisson distribution with the expectation value as the mean. ii) The haloes are placed randomly within each of the smoothing volumes, each of them has a peculiar velocity equal to the mean velocity of the dark matter particles within the same cell. iii) Eq. 6.3 has been truncated to linear order.

As a result we have generated 50 independent sub-resolution halo catalogues at three redshifts. Each contains approximately 17 million haloes with mass between  $5.48 \times 10^{11} h^{-1} M_{\odot}$  and  $1 \times 10^{16} h^{-1} M_{\odot}$  at  $z = 0$ . In the following subsection we will explore the properties of the catalogues.

### 6.3.2 Abundance and clustering

In this subsection we compare the abundance and clustering strength in our sub-resolution halo catalogues with the same quantities derived from haloes directly identified by a FoF algorithm in a high resolution simulation (for details of the FoF catalogues refer to §2.2).

In the upper panels of Fig. 6.1 we show the differential halo mass function from our catalogues (blue filled circles) and that from haloes directly identified in the BASICC simulation. In the lower panels we can see in greater detail the differences between the two populations. As expected, this figure shows excellent agreement in the number density of low mass haloes. The difference is smaller than 10% for objects of mass  $M < 7.51 \times 10^{13} h^{-1} M_{\odot}$  at  $z = 0$ ,  $M < 2.7 \times 10^{13} h^{-1} M_{\odot}$  at  $z = 0.5$  and  $M < 1.14 \times 10^{13} h^{-1} M_{\odot}$  at  $z = 1$ . On the contrary, there is a strong disagreement at the high mass end. This is likely to be caused by the fact that Eq. 6.2 is inconsistent for highly biased haloes in low density regions where  $\delta_h < -1$  (the problem is alleviated in the low mass regime where  $b \lesssim 1$ ). As a consequence of haloes of a fixed mass become more biased with increasing redshift, the behaviour of the halo mass function is reasonable for a reduced range of masses at high redshift.

We extend the comparison by investigating the clustering strength in the sub-resolution catalogues. Each column of Fig. 6.2 displays the linear bias parameter as a function of the peak height,  $\delta_c/\sigma(M, z)$  on the bottom axis and as a function of mass on the top axis. Note that we have computed the bias by smoothing the halo and dark matter density field in cells of size  $167 h^{-1} \text{Mpc}$ , i.e.  $b = \langle \delta_{hh} \rangle / \langle \delta_{mm} \rangle$ . As in the previous plot, the vertical lines indicate the maximum halo mass in which the result from the sub-resolution haloes agrees to within 10% with that of the resolved haloes. Similar to the behaviour seen in Fig. 6.1, at the high mass end, the sub-resolution haloes tend to fail to reproduce the properties seen in the resolved FoF catalogues, which suggests a common origin for the

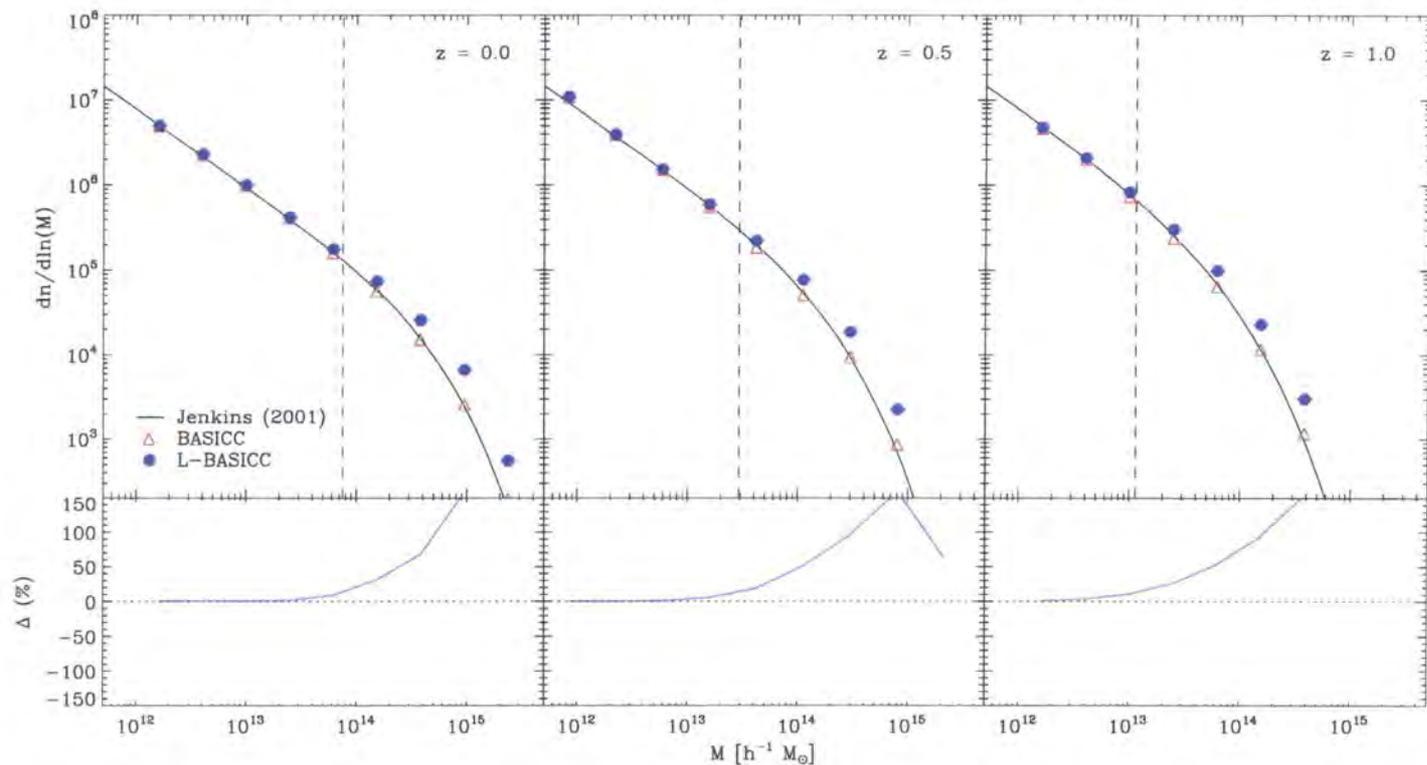


Figure 6.1: Comparison between the mass function of FoF haloes resolved in the BASICC simulation and the mean of sub-resolution halo catalogues built from the L-BASICC ensemble. Top: The mean number of haloes per logarithmic mass bin, as a function of their mass. Red triangles show this quantity from haloes in a high resolution simulation (BASICC), while the blue symbols with error bars show the mean and rms from our algorithm applied to 50 simulations (L-BASICC). Each column shows a different redshift and the corresponding prediction from Jenkins et al (2001) (dotted lines). The vertical dashed lines indicate the halo mass at which the number of haloes resolved directly in the simulations and those created by our algorithm first differ by 10%. The relative difference between these catalogues is shown in the bottom panels.

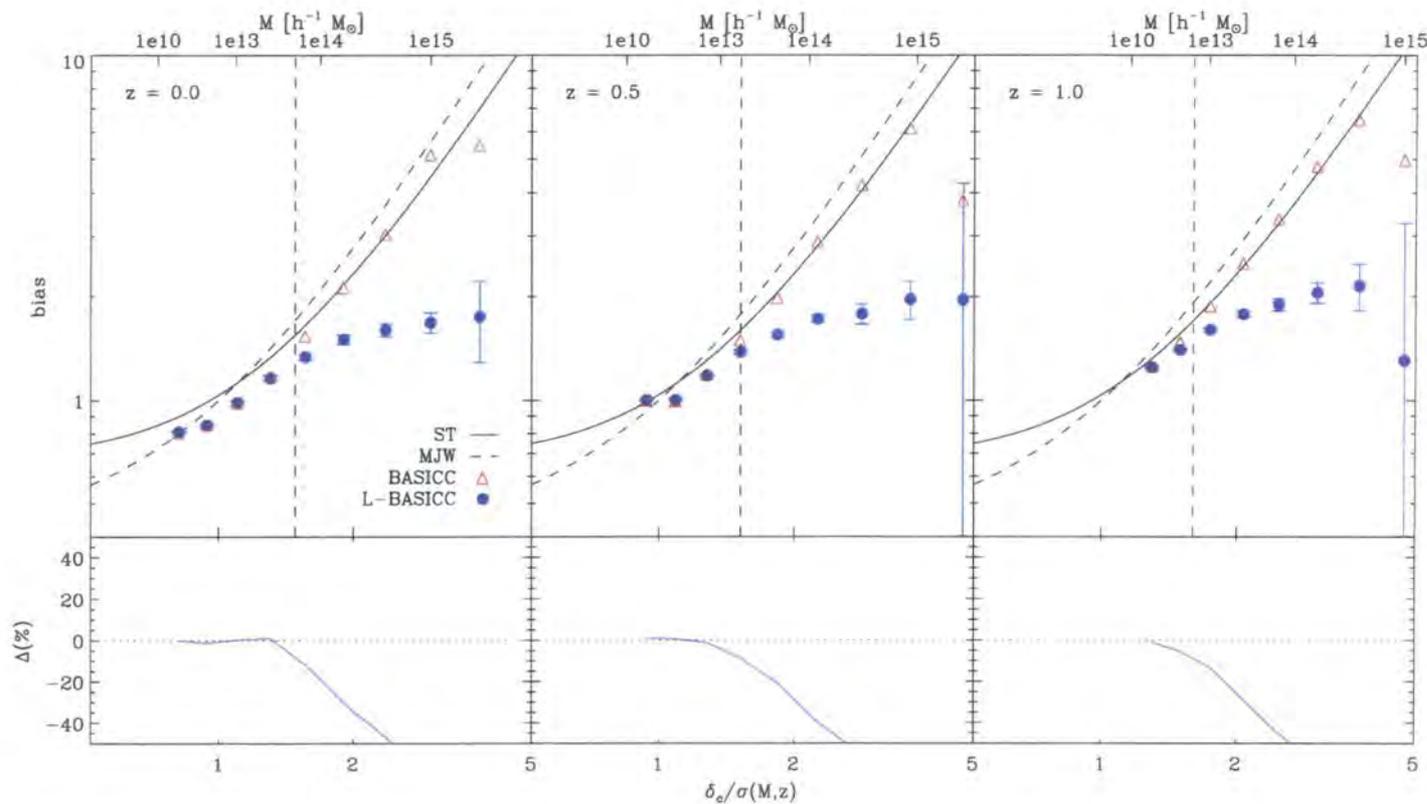


Figure 6.2: Top: The linear bias parameter as a function of halo mass (top axis) or peak height  $\nu = \delta_c/\sigma(M, z)$  (bottom axis) for three different redshifts as indicated on each panel. The blue filled circles with error bars show the mean and rms bias for the sub-resolution haloes generated by our algorithm applied to the 50 low resolution L-BASICC simulations. The bias measured from FoF haloes in the high-resolution BASICC simulation is shown as red triangles. Theoretical predictions by Mo, Jing & White (1998) and Sheth & Tormen (2001) are also shown. The lower panels show the relative difference between the halo bias from the two different catalogues.

misbehaviour seen in the abundance and clustering of sub-resolution haloes. Note, that the 10%-difference mass limit is slightly smaller than that derived from the mass function at;  $z = 0$   $M_{\max} = 5.23 \times 10^{13} h^{-1} M_{\odot}$  while at  $z = 0.5$  and  $z = 1$   $4.3 \times 10^{13} h^{-1} M_{\odot}$  and  $6.73 \times 10^{12} h^{-1} M_{\odot}$  respectively.

Finally, we explore the spherically averaged clustering of the halo catalogues in redshift space. Fig. 6.3 shows the ratio between the linear bias parameter measured in redshift space and that measured in real space for the sub-resolution haloes as well as for the FoF haloes. In linear perturbation theory, this quantity is equivalent to the square root of the Kaiser “boost factor” (Kaiser, 1987):

$$f = \left( 1 + \frac{1}{5}\beta + \frac{2}{5}\beta^2 \right), \quad (6.4)$$

where  $\beta = \Omega_m(z)^{0.6}/b$  and  $b$  is the linear bias parameter. This expression is over plotted in Fig. 6.3 for comparison.

Despite the scatter between the sub-resolution halo catalogues, we find reasonably good agreement between the theoretical expectations and the results from the BASICC haloes. Given the results presented in previous figures, it is not surprising to see the differences for haloes corresponding to high peaks. Nevertheless, our scheme to assign peculiar velocities to haloes performs satisfactorily in the regime where the abundance and clustering in real space are properly imprinted in the sub-resolution catalogues.

## 6.4 Application: Large scale clustering of Luminous Red Galaxies

Recently, the clustering of Luminous red galaxies (LRG) has been of great importance in probing different cosmological scenarios. The low number density but strong clustering of these galaxies means that the spatial distribution of LRG can be mapped over vast regions of the sky at low observational cost. This enables tight constraints to be placed on cosmological parameters, in particular by measuring the BAO feature (Eisenstein et al., 2005; Gaztanaga et al., 2008; Cabre and Gaztanaga, 2008). Unfortunately, there is still an incomplete understanding of the uncertainties associated with their clustering.

Semi-analytical modelling and observational evidence suggests that LRGs populate not only very massive haloes, but also they can be found in haloes with masses as small as  $10^{11} h^{-1} M_{\odot}$  (Almeida et al., 2008; Wake et al., 2008). Therefore, the modelling of LRG clustering, and the BAO feature imprinted on it, requires huge simulations with

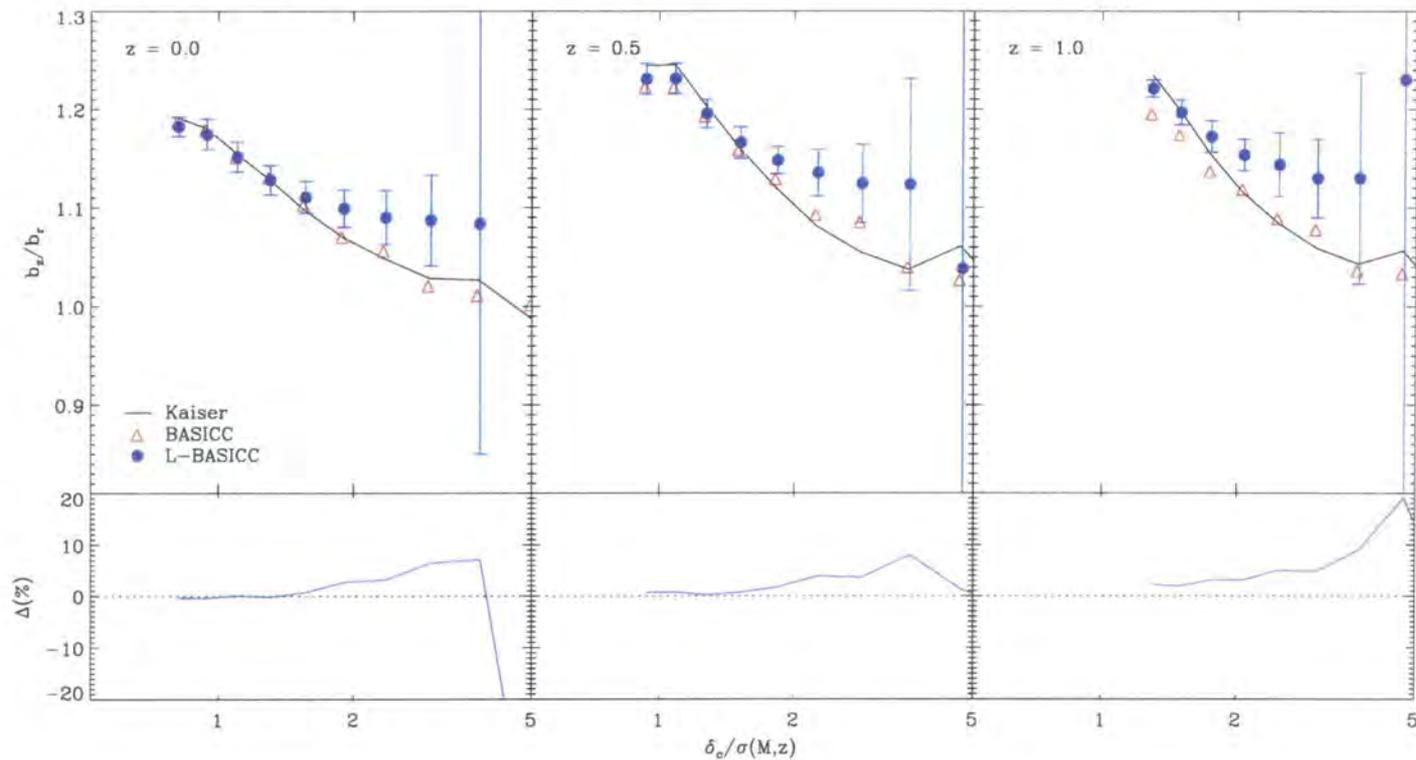


Figure 6.3: The linear bias parameter for haloes measured in redshift space  $b_z$  divided by that measured in real space  $b_r$ , as a function of the peak high (bottom axis) or halo mass (top axis). The mean and rms of this quantity, measured from our sub-resolution haloes in an ensemble of low resolution simulations, is displayed as blue symbols with error bars. As red triangles we display the results measured from FoF haloes in the BASICC simulation. For comparison, we have also included the prediction based on linear theory (solid line, see Eq. 6.4).

considerable dynamic ranges in mass. Although such extremes can be achieved in modern supercomputers, the computational cost is enhanced to inaccessible levels when studying uncertainties or subtle features present in the clustering (such as small scale dependent bias) due to the multiple realizations required.

In this section we approach this problem with the help of the algorithm that we have described in this chapter. Specifically, we have generated 50 LRG catalogues to compute the mean and variance of the two-point correlation function. Details of the creation of the LRG catalogues as well as the clustering measurements are presented in the following subsections.

#### 6.4.1 The haloes and LRG catalogues

The starting point in the creation of our LRGs mock catalogues is to predict the abundance and spatial distribution of the dark matter haloes that are likely to host such galaxies. For this purpose, we have created 50 hybrid halo catalogues, each of them spanning 4 orders of magnitude in mass within a volume of  $2.4 h^{-3} \text{Gpc}^3$  at  $z = 0.5$ .

The halo catalogues are hybrid in the sense that they consist of two types of haloes. High mass ones ( $M > 1.7 \times 10^{13} h^{-1} M_{\odot}$ ) correspond to objects identified directly by a FoF algorithm, with at least 10 particles, in each of the L-BASICC simulations. Then, smaller masses sub-resolution haloes ( $5.48 \times 10^{11} < M/(h^{-1} M_{\odot}) < 1.7 \times 10^{13}$ ) were created using the algorithm described in §6.2. In this way, we are effectively extending the dynamic range of the L-BASICC simulations towards lower masses. Combining the two types of haloes also eliminates the need to reproduce the troublesome high mass end in the sub-resolution catalogues.

Once we have generated the catalogues that contain all the haloes that are expected to host LRGs, we use a Halo Occupation Distribution (HOD) model to determine how many LRGs on average populate each dark matter halo (for a review on the halo model see Cooray and Sheth, 2002). Following Wake et al. (2008) we can express the mean number of central LRGs,  $N_c$  as a function of the host halo mass,  $M_{\text{halo}}$  as:

$$\langle N_c | M_{\text{halo}} \rangle = \exp(-M_{\text{min}}/M_{\text{halo}}), \quad (6.5)$$

and the mean number of satellite LRGs,  $N_s$  as:

$$\langle N_s | M_{\text{halo}} \rangle = (M_{\text{halo}}/M_1)^{\alpha}. \quad (6.6)$$

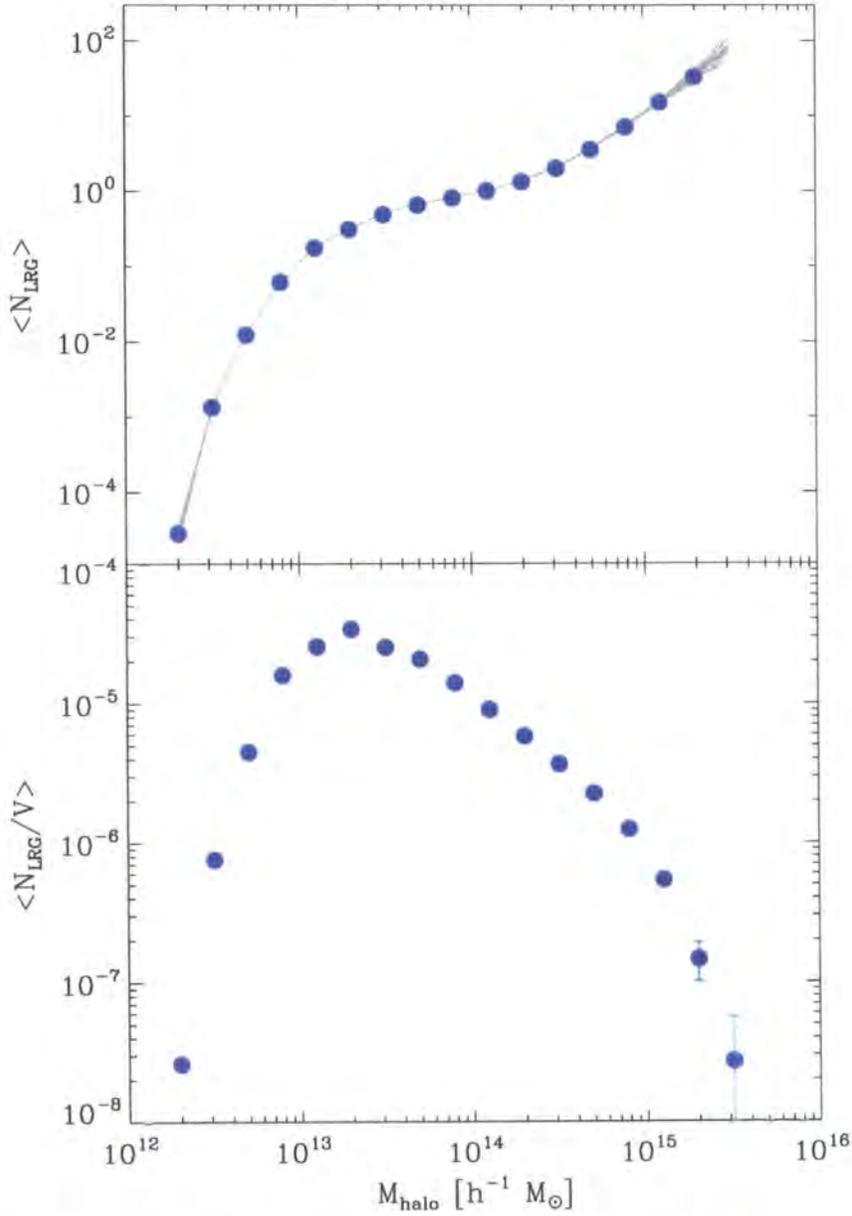


Figure 6.4: Top: the mean number of LRGs per halo as a function of the host halo mass. Bottom: the contribution to the mean volume density from LRGs hosted in haloes of different mass. The grey lines show the quantities in each of our simulations while the blue symbols show the mean and rms.

Consequently, the total number of LRGs has a expected value:

$$\langle N_{\text{LRG}} | M_{\text{halo}} \rangle = \langle N_{\text{c}} | M_{\text{halo}} \rangle [1 + \langle N_{\text{s}} | M_{\text{halo}} \rangle] \quad (6.7)$$

where  $\alpha$ ,  $M_{\text{min}}$  and  $M_1$  are, in principle, free parameters but that can be constrained either by observations or by semi-analytic modelling. Indeed, Wake et al. (2008), using the measured clustering of 2SLAQ LRGs, found that the numbers that best fit  $M_{\text{min}}$ ,  $M_1$  and  $\alpha$  are  $2.19 \times 10^{13} h^{-1} M_{\odot}$ ,  $2.82 \times 10^{13} h^{-1} M_{\odot}$  and 1.86 respectively. In the second step of our modelling we have assumed that in Eq. 6.5 and 6.6 follow a Poisson distribution which combined with the values from Wake et al. allow us to place LRGs within our hybrid catalogues. Note that the alternative approach of applying a full semi-analytic modelling to the hybrid halo catalogues could also have been taken.

Each of our final catalogues contain 398963 galaxies, or equivalently, a number density of  $1.66 \times 10^{-4} h^3 \text{Mpc}^{-3}$ . Even though, on average, there is less than one LRG per sub-resolution halo, jointly they host a total of 114243 galaxies which represents the 28% of whole LRG sample. Fig. 6.4 shows the resulting mean number of LRG in our catalogues per halo (top panel) as well as the total number LRGs (bottom panel), both as a function of the mass of the host halo.

#### 6.4.1.1 Correlation Function of LRG

At this point, we are finally in a position to investigate the clustering of LRGs. We measure the correlation function of our catalogues following the procedure outlined by Barriga and Gaztañaga (2002), Eriksen et al. (2004) and Sanchez et al. (2008). This approach is considerably more efficient than a direct pair count when computing correlation functions on large scales in catalogues containing a large number of particles.

In brief, the method consists of a pixelization of the density field from which the spherically averaged correlation function can be estimated as:

$$\hat{\xi}(|x - x'|) = \frac{1}{N} \sum_{i,j} \delta_i(x) \delta_j(x') \quad (6.8)$$

where  $\delta_i = (n_i - \langle n \rangle) / \langle n \rangle$  is the density fluctuation in the  $i$ -th bin of the grid. We have chosen  $N_{\text{grid}} = 256$ . This method gives an accurate estimation of the correlation function for scales larger than a few grid cells.

Fig. 6.5 shows the result of applying this procedure to compute the correlation function for LRGs in each of our 50 catalogues. The top panel displays the measurements in real

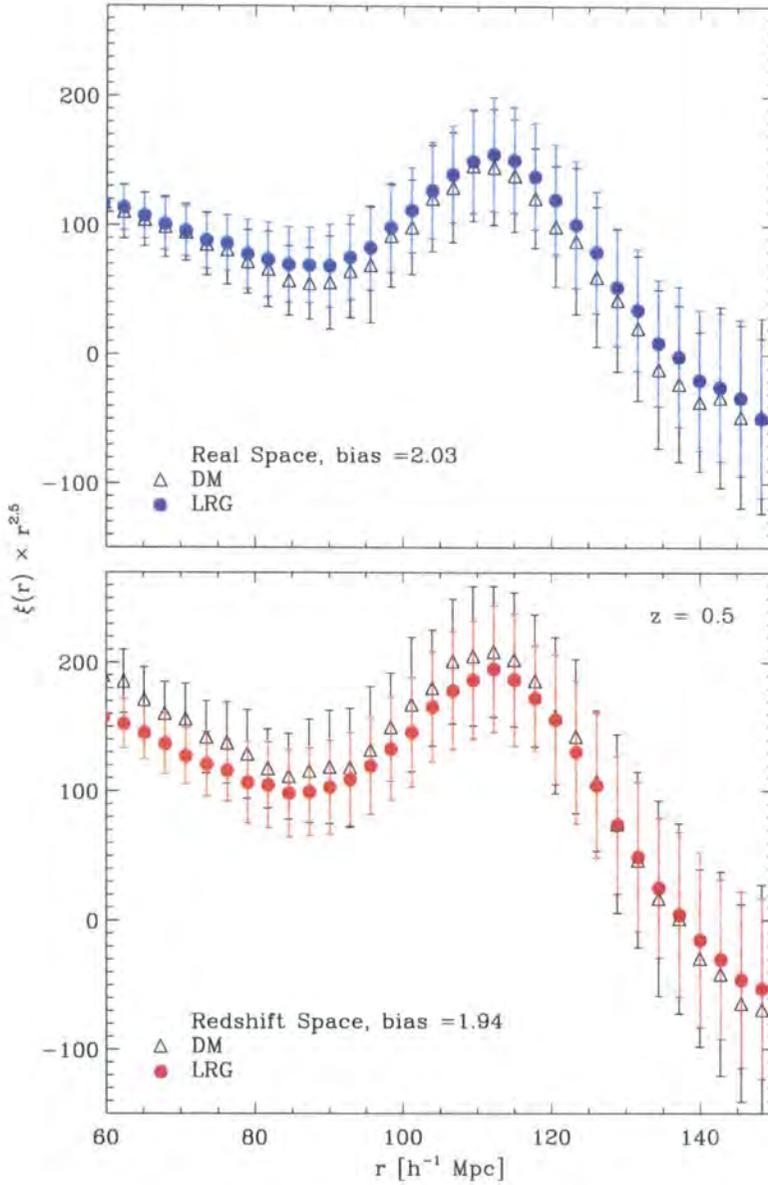


Figure 6.5: The mean and variance of the LRG correlation function from our catalogues at  $z = 0.5$ . As indicated by the legend the top and bottom panels display the results in real and redshift space respectively. In each case filled circles indicate the clustering measured from the LRG catalogues created by combining hybrid halo catalogues (mixture of sub-resolution and FoF haloes) with a HOD inferred by Wake et al. (2008). The triangles show the correlation function measured from dark matter samples diluted to match the number density of the LRG catalogues. To allow a full comparison, we have normalized each measurement (and the respective variance) so that  $\langle \xi(r) \rangle = 1$ . Note that we display  $\xi(r) \times r^{2.5}$  in the y-axis to enhance the acoustic peak.

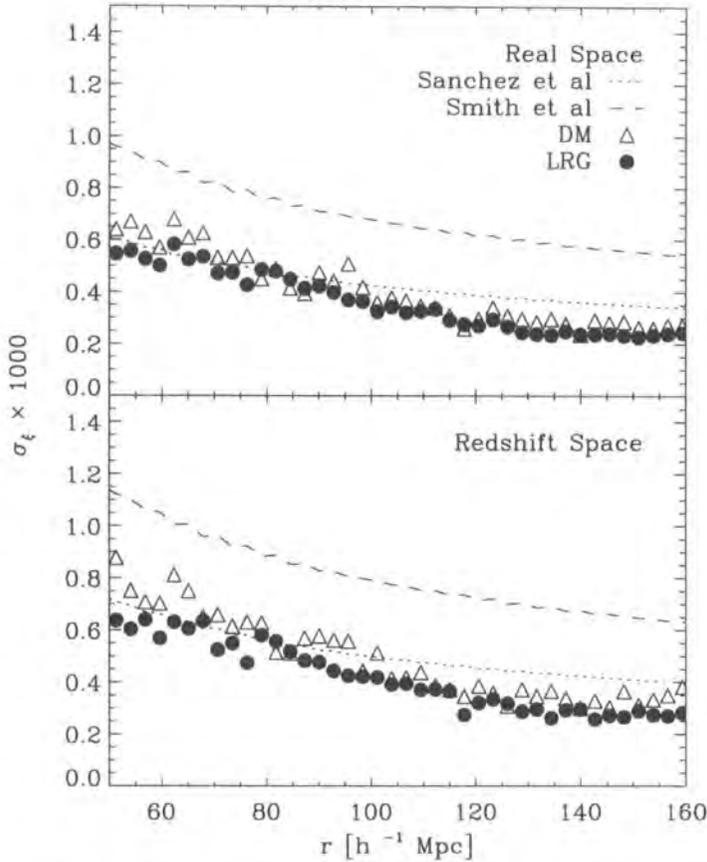


Figure 6.6: The variance in the 2-pt correlation function measured from 50 LRGs catalogues in real space (top) and redshift space (bottom) at  $z = 0.5$ . We also plot two theoretical estimates for the variance as dotted and dashed lines.

space while the bottom panel shows the results in redshift space. In both cases the mean and variance of the measurements are indicated by the filled circles and error bars. In order to contrast our results, we have also measured the correlation function for a set of dark matter particles at  $z = 0.5$  from the L-BASICC simulations. Each of these dark matter samples has been diluted to have the same number density as our LRG catalogues. Note that the y-axis displays  $\xi \times r^{2.5}$  instead of  $\xi$ , in this way the acoustic peak is highlighted. In addition, the results (including the errors) in both real and redshift space have been renormalized as described in the figure caption.

By comparing the correlation function of LRGs with that of dark matter we can see the effects of galaxy bias. As it is clear from Fig. 6.5 that the scaled measurements agree fairly well with each other, the LRG bias is roughly scale independent over the range of

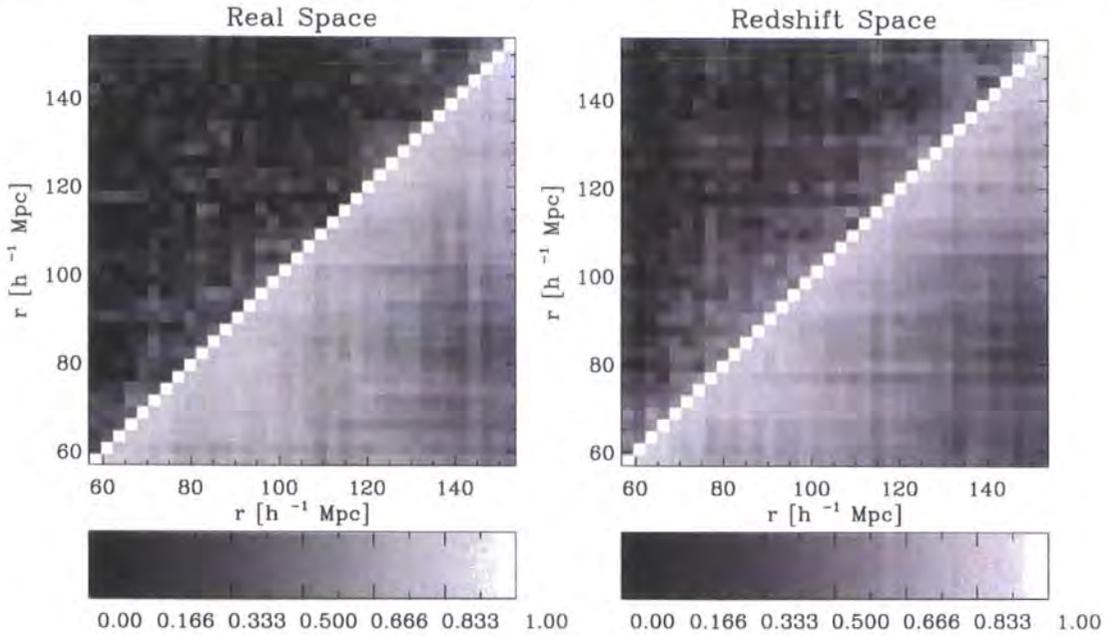


Figure 6.7: The normalized covariance matrix from our ensemble of LRG mock catalogues (lower triangular region) and from an ensemble of DM catalogues with the same number density as the LRG (upper triangular region). The left plot shows displays the results in real space while the right displays the covariance matrix in redshift space.

separation plotted here. Nonetheless, there is a small dependence on scale in real space which seems to be accentuated in redshift space. However, the discrepancy can not be measured with enough significance due to the magnitude of the errors associated with the simulation volume. Further investigation of this issue is needed.

In Fig. 6.6 we compare the variance measured from our ensemble of LRG catalogues (filled circles) with that measured from the diluted dark matter samples (triangles). The dotted and dashed lines show two theoretical predictions for the variance based on power spectrum measurements (see Sanchez et al., 2008, for more details).

We extend this comparison in Fig. 6.7 where we display the normalized covariance matrix (Cohn, 2006; Smith et al., 2008),  $C_{\xi}(r, r') \equiv \langle (\xi(r) - \bar{\xi}(r))(\xi(r') - \bar{\xi}(r')) \rangle / \sigma(r)\sigma(r')$ , in real space (left hand side plot) and in redshift space (right hand side). The above diagonal parts of the matrices show the measured covariance for the diluted dark matter samples. The below diagonal parts show the covariance for the LRG catalogues.

In general we find that the theoretical predictions by Sanchez et al. (2008) and the sample of dark matter provide a fairly good description of variance in the LRG sample.

On the contrary, the predictions of Smith et al. (2008) overestimate the dispersion in our measurements, which, as discussed in Sanchez et al. (2008), appears to be due to the underestimation of the binning in the correlation function measurements (which reduces the variance between adjacent scales).

In addition, the non-diagonal parts of the covariance matrix show a significant disagreement between the LRG and dark matter samples. In particular, our LRG catalogues show stronger off-diagonal correlations than dark matter. This could originate in one of several factors. One possibility is the contribution of the higher order moments of the halo density field. As we showed in Chapter 4, the higher order moments of dark matter and haloes are considerably different. As an example, recall that even if the dark matter density field is Gaussian (i.e. the higher moments are zero), the dark haloes will have non-zero higher order correlations which will contribute into the covariance matrix. Of course, the observed disagreement in the covariance matrix could also have been introduced artificially by our algorithm. Both of these issues require further investigation.

## 6.5 Conclusions

Due to the large areas of the sky that future surveys are expected to map, the derived information regarding galaxy clustering will be of great accuracy and it may even perform the task of distinguishing between different models for gravity or help to unveil the nature of dark energy. It is therefore fundamental to understand the systematic and statistical errors associated with the measurements.

In this section we have illustrated the feasibility of a new scheme that allows a rapid and efficient creation of galaxy mock catalogues. Our method uses the dark matter density field from low resolution N-body simulations combined with the bias parameters and mass functions extracted from higher resolution simulations, to predict statistically the expected density field of dark matter haloes. Since low-resolution simulations are relatively easy to generate, our procedure allows the investigation of uncertainties in both the measurements themselves and in the procedures employed to extract useful information from the data.

During this chapter we have shown that, on large scales, the generated halo population behaves fairly well over a considerable range of masses. At  $z = 0$  in particular, the abundance and clustering strength in both real and redshift space, of haloes less massive than  $7.51 \times 10^{13} h^{-1} M_{\odot}$  agree within 10% with those computed directly from FoF haloes

identified in a high-resolution simulation. For high mass haloes or at higher redshifts, our procedure performs less satisfactorily.

An interesting application of our scheme is to the creation of hybrid catalogues. High mass haloes can be extracted directly from the cosmological simulations whilst low mass ones can be artificially generated. In this way, we can employ our algorithm in the regime where it works best. As an example, we have created 50 such catalogues from the L-BASICC simulations which are combined with a HOD for LRGs. From the resulting galaxy catalogues we were successfully able to predict their mean correlation function along with the full covariance matrix. We found that the variance in a dark matter sample as well as analytical estimates are in agreement with measurements from the LRG catalogues. In spite of this, significant differences in the off-diagonal terms of the covariance matrix were also found.

The algorithm presented here can be improved in a number of ways. Areas which could be improved include the way we compute the dark matter density field, the placement of haloes within the smoothing volume and the inclusion of higher order bias terms. Also extended features can be incorporated such as exclusion effects between haloes, or the implementation of a new function that better encapsulates the relationship between the halo and dark matter density field. We plan to address these ideas in the future. With these improvements, the technique presented in this chapter or similar ideas will hopefully help to improve the understanding and treatment of uncertainties in observations and, therefore, allow the full potential of measurements of the large scale distribution of galaxies to be reached.

# Chapter 7

## *The construction of a halo light-cone*

### 7.1 Introduction

Comparing theoretical models with experiments is a fundamental part of science. In cosmology for instance, confronting models with the observed clustering properties of galaxies has been of enormous importance not only for the understanding of the galaxy formation process but also in distinguishing between different cosmological scenarios (e.g. Gaztanaga et al., 2008). Naturally, in order to correctly interpret the observations, a suitable and accurate theoretical model is desirable.

When studying the clustering and other properties of objects in the Universe, the data is commonly compared with models that are evaluated at a single cosmic epoch. This procedure is justified when dealing with local or narrow band galaxy surveys which focus on a very restricted redshift range. In such cases, the properties and distribution of the observed galaxies evolve little over the depth of the survey. Hence the observations can be safely approximated as a population of galaxies that resides at the median redshift of the observations.

There are, however, several current and future surveys that are not limited to a small redshift range. Indeed, these experiments will observe galaxies over a considerable fraction of the age of the Universe. For instance, the DEEP2 survey has a principal redshift range of  $0.75 < z < 1.4$  (Davis et al., 2007) and the main Pan-STARRS survey will detect galaxies up to redshift 1.2 (Cai et al., 2008). Furthermore, there are a number of observations, besides the clustering, that are sensitive to structure along the line of sight, e.g. the Integrated Sachs-Wolfe effect or gravitational weak lensing.

In order to properly explore and understand such measurements, models are required where the variation of the spatial distribution and properties of galaxies with redshift are consistently taken into account. These models are usually referred to as light-cones

since objects appear in the catalogues when they cross the past light-cone of an observer. Several approaches have been made to construct these theoretical models (Yamamoto and Suto, 1999; Hamana et al., 2001; Blaizot et al., 2005; Kitzbichler and White, 2007; Muñoz and Loeb, 2008). Amongst the most realistic are those derived from dark matter N-body simulations. In the following, I will briefly describe the algorithms used along with their advantages and disadvantages.

The most straightforward way to construct a light-cone is to store the position and velocity of each dark matter particle whenever it crosses the light-cone of a fiducial observer. This condition is checked on the fly between consecutive time-steps of the simulation (Evrard et al., 2002). This approach generates an accurate description of the dark matter clustering. However, in practice there is a limit to the number of light-cones that can be generated, i.e. they cannot be built after the N-body simulation has finished. A further problem is the extra memory and CPU usage that building the light-cone adds to the calculation, which reduces the mass and time resolution of the simulation itself.

One way to overcome the shortcomings of the approach of Evrard et al. is to store the position and velocity of each particle at discrete times (snapshots). Then the construction of a light-cone is carried out by interpolating in post-processing the trajectory of particles between these snapshots. Since this is done a posteriori, it does not add extra hardware requirements to the simulation. In spite of this advantage, the high velocities and eccentric orbits of dark matter particles within a halo cause the interpolated positions to be inaccurate (a particle can spend many orbital times inside a halo between two snapshots, so an interpolation would not capture the main features of its trajectory since there are many fewer snapshots than time steps). As a result, the inner density profile of haloes are smoothed which leads to a upper limit on the minimum number of particles needed to identify a halo or subhalo robustly in comparison with the number of particles needed to resolve these structures in a snapshot (small overdensities are smeared out in the light-cones).

A recent common approach is to identify haloes in the stored snapshots and then populate these haloes with galaxies whose position are interpolated (Blaizot et al., 2005; Kitzbichler and White, 2007). In this case the problems outlined above are diminished. It is possible to create a variety of light-cone geometries, depths, etc and, simultaneously, not distort the small scale clustering since the haloes and its substructures are identified in snapshots. Nonetheless, the galaxies properties have to be interpolated or taken from the closest snapshot which add discreteness effects or possible inaccuracies in the predicted

properties of the galaxy population.

As is clear from the above, each of these methods have their own advantages and disadvantages and none of them completely fulfil the degree of accuracy and flexibility that would be ideal. The purpose of this chapter is to introduce a new method to predict the spatial distribution of galaxies in the light-cone from N-body simulations, in which the problems of the procedures mentioned above are solved. Instead of constructing light-cones based on dark matter particles or galaxies, we follow dark matter haloes between snapshots, to predict their trajectories over time. In this way, we find the position of haloes at the exact moment at which they cross the light-cone as well as their merger history. Our method has the advantage of requiring only the interpolation of the halo positions (which are better described as a smooth curve between snapshots than those of the dark matter particles). In addition, it is straightforward to combine the halo light-cone with semi-analytic models of galaxy formation to predict galaxy properties at the precise light-cone crossing.

The layout of this chapter is as follows. First, in §7.2 we give the analytic formulae for the correlation function and mass function of haloes on the light-cone. These expressions will help in the a posteriori testing of our results. We describe our procedure to build a light-cone in §7.3. In §7.4 we present a comparison between our results and the analytic estimates. Finally, we summarize our findings in §7.5.

## 7.2 Haloes on the light-cone

Several statistics of haloes and dark matter on the observed light-cone can be easily written analytically by a weighted integral of quantities expressed at a single redshift. In this section we will just list the relevant formulae (for a more detailed discussion see Matarrese et al., 1997; Yamamoto and Suto, 1999; Hamana et al., 2001; Muñoz and Loeb, 2008).

One of the most basic statistics on a light-cone that covers a redshift interval  $[z_1, z_2]$ , is the number density of haloes as a function of their mass,  $n_{\text{LC}}(z_1 < z < z_2, M)$ . This can be expressed simply by averaging the weighted number density of haloes as a function of redshift and mass:

$$n_{\text{LC}}(z_1 < z < z_2, M) = \frac{\int_{z_1}^{z_2} dz \frac{dV}{dz} n(z, M)}{\int_{z_1}^{z_2} dz \frac{dV}{dz}}, \quad (7.1)$$

where  $dV/dz$  is the comoving volume element per solid angle and  $n(z, M)$  is the number density of haloes of mass between  $M$  and  $M + dM$  at redshift  $z$ .

In a similar way, we can define an effective linear bias parameter for haloes on the light-cone, in terms of a volume averaged bias factor:

$$b_{\text{LC}}(z_1 < z < z_2, > M) = \frac{\int_{z_1}^{z_2} dz \frac{dV}{dz} b_{\text{eff}}(z, M)}{\int_{z_1}^{z_2} dz \frac{dV}{dz}}, \quad (7.2)$$

and

$$b_{\text{eff}}(z, M) = \frac{\int_M^\infty dm b(m, z) n(m, z)}{\int_M^\infty dm n(m, z)}, \quad (7.3)$$

where  $b(z, m)$  is the bias factor of haloes of mass  $m$  at redshift  $z$ .

Putting these expressions together we can write the correlation function in real space for haloes on the light-cone:

$$\xi_{\text{LC}}^{hh}(z_1 < z < z_2, > M, r) = \frac{\int_M^\infty dm \int_{z_1}^{z_2} dz \xi^{mm}(z, r) \frac{dV}{dz} b_{\text{eff}}^2(z, M)}{\int_{z_1}^{z_2} dz \frac{dV}{dz}} \quad (7.4)$$

and for the dark matter:

$$\xi_{\text{LC}}^{mm}(z_1 < z < z_2, r) = \frac{\int_{z_1}^{z_2} dz \frac{dV}{dz} \xi^{mm}(z, r)}{\int_{z_1}^{z_2} dz \frac{dV}{dz}}, \quad (7.5)$$

where  $\xi^{mm}(z, r)$  is the dark matter correlation function at redshift  $z$ . Note that this expression assumes that evolution of clustering is negligible on the scale  $r$ .

### 7.3 Building a light-cone

The first step in our attempt to populate the observed light-cone with galaxies is to predict the abundance, mass and position of dark matter haloes. In brief we achieve this goal by following the merger history of every halo. In this way, a unique trajectory is built for every halo and the precise moment of the light cone crossing can be found. The following subsection contains a more detailed description of our procedure (§7.3.1). In the subsequent subsection we apply this procedure to generate a light-cone from the Millennium simulation (§7.3.2).

### 7.3.1 Algorithm

We start by replicating the simulation box as many times as needed to cover the redshift range and angular size in which we are interested. Due to the periodic boundary conditions of the simulation, there is an indistinguishable transition between replications. However, note that spurious clustering signals could appear on scales similar or larger than the box size.

In a second step, we find the coordinates of dark matter haloes in each available snapshot using a Friends-of-Friends algorithm (FoF, Davis et al., 1985). We then predict the position and velocity of each halo between snapshots assuming that each dimension can be described by a polynomial in time:

$$x(t) = A + Bt + Ct^2 + Dt^3, \quad (7.6)$$

$$v(t) = B + 2Ct + 3Dt^2, \quad (7.7)$$

where if  $t_1$  and  $t_2$  are the times of consecutive snapshots,

$$A = x(t_1), \quad (7.8)$$

$$B = v(t_1), \quad (7.9)$$

$$C = -2v(t_1) - 3x(t_1) - v(t_2) + 3x(t_2), \quad (7.10)$$

$$D = v(t_1) + v(t_2) + 2[x(t_1) - x(t_2)], \quad (7.11)$$

$v(t_1)$  and  $x(t_1)$  are the velocity and position of the halo, and  $v(t_2)$  and  $x(t_2)$  correspond to the same quantities but for its descendant halo. This is a key aspect of our algorithm. In each snapshot we identify the descendant of a halo as the structure that contains the majority of the 10% most bound particles. This allows us to create a unique path for each halo over its whole life. Also note that Eqs. 7.6-7.11 include the constraint that the time derivative of the position must be equal to the velocity.

In addition, we interpolate the halo mass between snapshots assuming a smooth growth, i.e.

$$M(t) = M(t_1) \exp \left[ \frac{\log(M(t_2)) - \log(M(t_1))}{t_2 - t_1} (t - t_1) \right]. \quad (7.12)$$

Once we know the position of a halo at all times, we look for the time at which its distance to our fiducial observer is the same as the distance that a photon would travel

from the time at which the halo lives to the observer. This is the moment at which the halo crosses the light cone thus we store this position, velocity and mass. An interesting issue regarding our method is the treatment of halo mergers, i.e. when two or more haloes have the same descendant in the following snapshot. In such cases it is possible that two or more haloes cross the light-cone in locations closer than their radii. Clearly, there should be a merger process and we should identify these objects as a single structure. However, here we have not implemented such procedure, and trace the haloes separately until the snapshot where they merge.

Along with storing the coordinates of the halo we also store its merger history. The latter can be fed into a semi-analytic galaxy formation model. This represents a great advantage over previous attempts of constructing light-cones, because it implies that we can follow the galaxy formation process up to the exact point at which the host halo crosses the light-cone. Hence no interpolation of galaxy properties is needed.

### 7.3.2 Light-cones from the Millennium Simulation

In order to test and illustrate our algorithm, we have produced a light-cone from outputs from the Millennium Simulation (cf. §5.2). This light-cone spans from redshift 0 to 1 which requires replicating the box 6 times along each axis. In order to decrease spurious periodicity in the halo distribution we have created our catalogue along the vector  $(x, y, z) = (0.91, 0.38, 0.099)$  (the observer is located at  $(0, 0, 0)$ ). The angular dimension is  $9 \text{ deg} \times 9 \text{ deg}$  which corresponds to approximately to ten times the angular extent of one field of the proposed Pan-STARRS Medium Deep Survey.

As a result, we have obtained 13077347 haloes more massive than  $1.72 \times 10^{10} h^{-1} M_{\odot}$ . The most massive object is located at  $z = 0.334$  and has a mass of  $2.425 \times 10^{15} h^{-1} M_{\odot}$  (which corresponds to 2816415 dark matter particles). In the following section we will investigate the properties of our catalogue in more detail.

## 7.4 Results

To start this section, we show an image in Fig. 7.1 of the halo distribution in the light-cone from  $z = 0$  to  $z = 1$ . The brightness in each pixel is proportional to the logarithm of the density projected along the line of sight. The density field is estimated using a Gaussian kernel of size fixed to 6 Mpc.

This image clearly shows the filamentary structure of the density field on large scales.

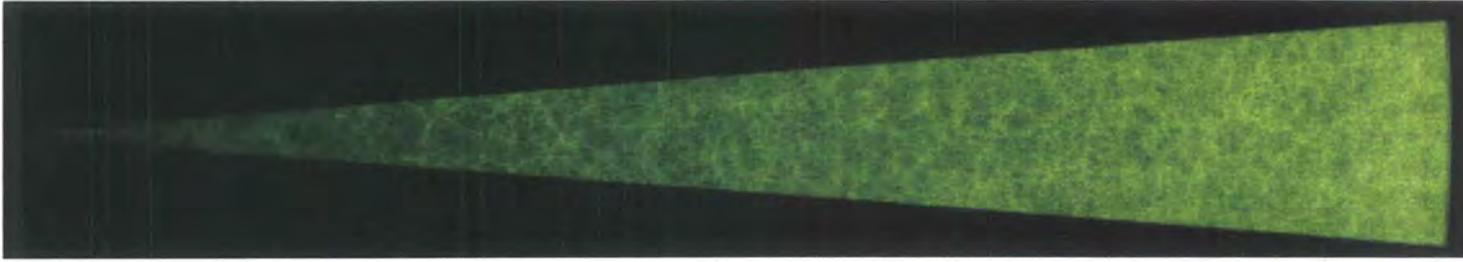


Figure 7.1: The projected halo density in the light-cone in a slice of depth 50 Mpc. The picture covers from redshift 0 to 1 and corresponds to 9 deg. The brightness of each pixel is proportional to the logarithm of the density integrated along the page.

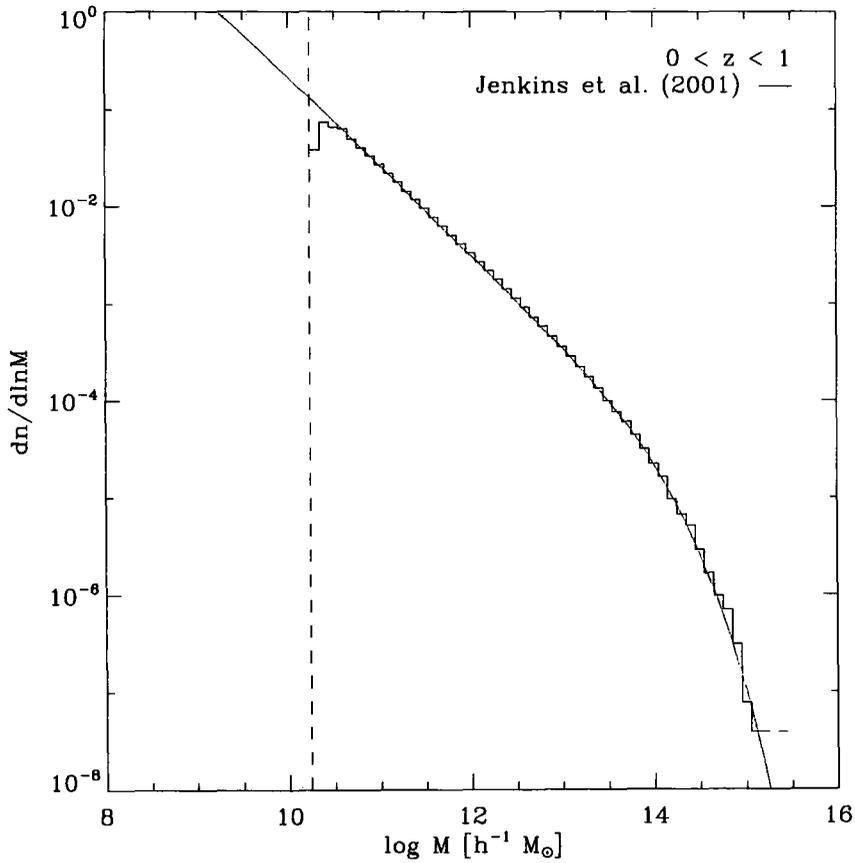


Figure 7.2: The mass function of haloes measured in the light-cone over the redshift range  $0 < z < 1$ . The analytic expression, Eq. 7.1, using the Jenkins et al (2001) mass function is plotted (red line) for comparison. The vertical dashed line indicates the minimum mass of haloes identified in the Millennium Simulation. This limit corresponds to  $M > 1.72 \times 10^{10} h^{-1} M_{\odot}$  (20 particles).

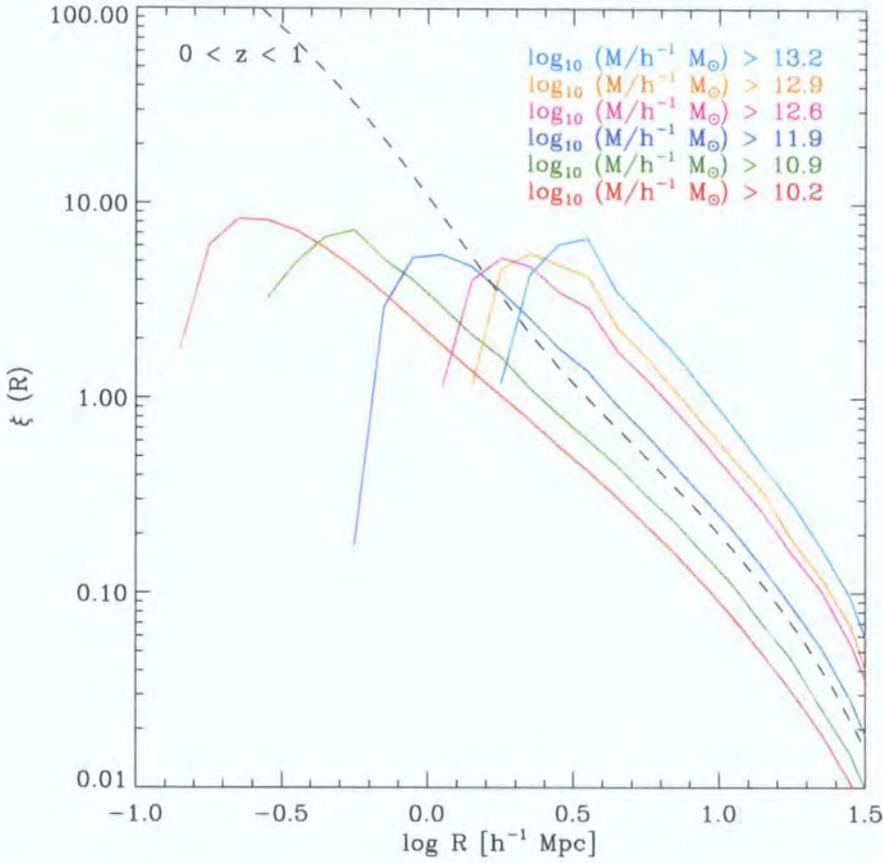


Figure 7.3: The two-point correlation function for haloes in the light-cone. Curves of different colour show results in real space for haloes of different mass. As indicated by the key, the lowest amplitude (red) curve corresponds to haloes with mass in excess of  $1.72 \times 10^{10} h^{-1} M_{\odot}$  while the top (blue) line indicates the correlation function for haloes with  $M > 1.58 \times 10^{13} h^{-1} M_{\odot}$ . These samples contain 13077347 and 17776 haloes respectively. The analytic prediction for the dark matter correlation function in the light-cone (Eq. 7.5) is also over plotted as a dashed (black) line.

In addition, we can see the evolution in the abundance and clustering of haloes with redshift. A high number density of haloes at low redshifts (left hand side). On the other hand, haloes are rarer at higher redshifts (right hand side) due to the fixed mass resolution.

A basic yet important test of our results is the distribution of halo masses. The mass function of haloes in the light-cone is plotted in Fig. 7.2. We show the results for haloes

with  $M > 1.72 \times 10^{10} h^{-1} M_{\odot}$  in the redshift interval  $0 < z < 1$ . We have also included the analytic prediction given by Eq. 7.1 which is computed using the mass function of Jenkins et al. (2001). Overall, the prediction and measurement from the light-cone are in very good agreement over the mass range resolved in the Millennium Simulation. However, at the very low mass end, there is a slight inconsistency. The number density of haloes is 37% lower than the predictions. Such a discrepancy may result from the limitations of our algorithm. First, we can not identify the progenitors of haloes resolved with fewer than  $\sim 20$  particles since they are likely to fall below our resolution limit. Nevertheless, this problem can be alleviated easily by using a lower particle limit to construct the merger tree or simply by using only haloes above a higher mass cut.

The auto-correlation function for haloes of different mass in real space is shown in Fig. 7.3. We compute  $\xi(r)$ , using the Landy and Szalay (1993) estimator which compares pairs of objects at a given separation with objects whose positions are randomly distributed. Explicitly,

$$\xi(r) = \frac{1}{RR} \left[ DD \left( \frac{n_r}{n_d} \right)^2 - 2DR \left( \frac{n_r}{n_d} \right) + RR \right], \quad (7.13)$$

where  $DD$ ,  $DR$  and  $RR$  are the data-data, data-random and random-random pair counts respectively, and  $n_r$  and  $n_d$  are the number of objects in the random and halo catalogues. The random catalogue is constructed following the same redshift distribution as the real halo catalogue, and with the same angular geometry as the light cone. Note that the use of a random catalogue is required due to the evolution of the mean density of haloes with redshift and the none periodic light cone boundary.

Fig. 7.3 shows the clustering measurements for haloes in 6 different ranges of mass. On large scales, where the two-halo term dominates, we can see that the shape of the correlation function for haloes seems to be a scaled version of that for the dark matter. The amplitude increases as the sample is restricted to more massive haloes. On the other hand, each one of the haloes samples shows a dramatic turnover on smaller scales. This feature results from an exclusion effect between haloes, i.e. two haloes cannot be located closer than the sum of their radii. Since the virial radius increases monotonically with the mass of the halo, the cut-off appears on larger scales for more massive haloes.

One key aspect in modelling galaxy clustering is the calculation of the bias of the host dark matter haloes. In Fig. 7.4 we test our results by comparing the bias derived from the light-cone against the analytical model presented in §7.2. Note that the model

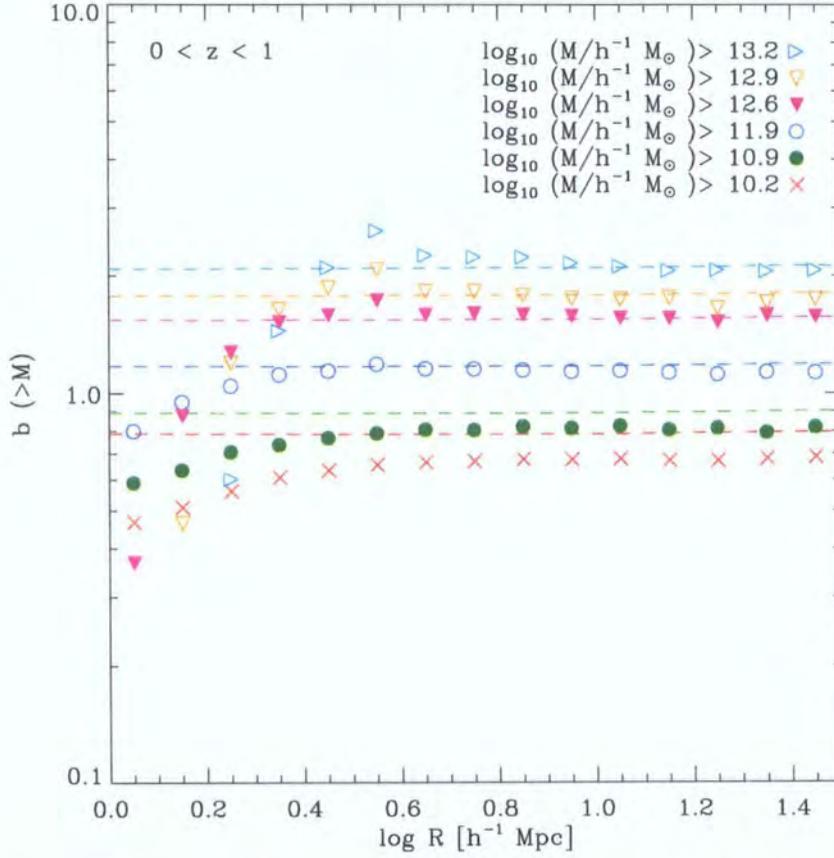


Figure 7.4: The linear bias parameter as a function of scale for haloes in the light-cone. The bias is defined as  $b^2 = \xi_{LC}^{hh} / \xi_{LC}^{mm}$ , where  $\xi_{LC}^{hh}$  and  $\xi_{LC}^{mm}$  are, respectively, the two-point correlation function for haloes and for dark matter in the light-cone. The symbols indicate our measurements while the dashed lines indicate the analytical prediction given by the square root of the ratio between Eq. 7.4 and Eq. 7.5.

is constructed using the expressions provided by Jenkins et al. (2001) and Scoccimarro et al. (2001) respectively for the number counts and bias of haloes.

The measurement for high mass haloes shows fairly good agreement with the model. However, there is a noticeable disagreement for the smallest mass haloes. The origin of this discrepancy could lie in a number of factors. First, it has been shown that analytic biasing prescriptions fail to describe, to an accuracy better than 10%, the bias measured in numerical simulations over the whole halo mass range employed in the construction of the light-cone (see Fig. 4.7 in Chapter 4). In general, while these prescriptions work quite well for highly biased objects, they tend to over-predict the bias for haloes with smaller masses. Such discrepancies could explain part of the disagreement with our measurements. Another possibility is that our model for the dark matter correlation function is an overestimation of the true one or that the definition of bias in the light-cone is flawed and requires improvement. Nevertheless, our haloes provide a reasonable description of the clustering expected in the light-cone, enclosing key features of its clustering and distribution.

## 7.5 Summary

In this chapter we have presented a new method to generate light-cones from cosmological N-body simulations. In contrast to previous approaches, our method consists of two steps. First, we compute the position of haloes at the moment of light-cone crossing. This is done by following their merger history which is interpolated between snapshots. Second, the halo light-cone can easily be combined with semi-analytic models of galaxy formation with which it is possible to generate a galaxy catalogue over a large redshift range. We plan to apply the second step to our halo catalogues in the future.

The advantages of our procedure are: i) Since the net velocity of haloes is much smaller than that of dark matter particles, their trajectory between snapshots and, therefore, their time of light-cone crossing can be constrained accurately. Since we are dealing with dark matter structures, not particles, the small scale clustering of galaxies cannot be artificially distorted. In fact, our method is just a spatial redistribution of haloes. ii) The observable properties of the galaxies can be computed up to exactly the time in which they are observed.

We have applied our procedure to build a light-cone from haloes identified in the Millennium Simulation. This light-cone contains haloes more massive than  $1.7 \times 10^{10} h^{-1} M_{\odot}$

on a solid angle of  $9 \text{ deg} \times 9 \text{ deg}$  from redshift 0 to 1. The tests carried out in Section 7.4 show reasonable agreement between our results and analytic expressions for the mass distribution of haloes.

Although our model already reproduces key features of the halo catalogue, it can be improved in several ways. The most important is probably an improved treatment of halo mergers. As two haloes can have the same descendant in the following snapshot, they can cross the light-cone in positions located closer than their joint radius. In fact, they could even end up one inside another. In the current implementation of our algorithm, we keep both haloes as a separate structure. However, it is possible to add a pruning process which would merge these objects and increase the mass (and subhaloes content) of the halo produced accordingly. Likewise, it is possible to follow the merger history of subhaloes between snapshots and include or exclude them from a host halo according to their light-cone crossing time.

The creation of light-cones poses an interesting and much needed extra layer of sophistication in galaxy formation modelling. As new surveys continue to expand the region of the Universe that is mapped, these more accurate tools will become essential for the correct and complete exploitation of the new datasets.



# Chapter 8

## *Conclusions*

The subjects covered in this thesis can be grouped into three areas. The main results are discussed below along with possible directions for future work.

### 8.1 Galaxy surveys and BAO

Over the next decade, several ambitious galaxy surveys will carry out one of the most promising approaches to reveal the properties of the dark energy: the measurement of baryonic acoustic oscillations (BAO). Due to the large volumes mapped, the systematic errors may become larger than the random uncertainties, degrading the measurements and therefore the constraints on the dark energy. For this reason, in order to be able to exploit the forthcoming data to the full, precise theoretical modelling of galaxy clustering and associated errors is essential. In the first part of this thesis (chapters 2 and 3), I devoted particular attention to this issue by giving an assessment of the effects which alter the appearance of the power spectrum on large scales. I also developed a new technique to study uncertainties in theoretical predictions.

In particular, in Chapter 2, I presented one of the most realistic predictions to date of the expected BAO signature in future galaxy surveys. A key piece of this analysis was the **BASICC** (Baryonic Acoustic Oscillation Simulation at the ICC) N-body simulation, which follows the evolution of more than 3 billion particles that represent the matter in a  $2.4 \text{ Gpc}^3/h^3$  region of the Universe. The analysis presented in Chapter 2 improved previous studies in three ways. First, the volume of the simulations used was comparable to that of future surveys. This allowed me to model the growth of structures on very large scales. Second, we were able to resolve most of the haloes that are likely to host the galaxies that will be detected in future surveys. Third, by combining the dark matter clustering with **GALFORM**, the semi-analytical model of galaxy formation developed at Durham, we were able to predict positions, velocities, colours and magnitudes, along with several other properties, for more than 1 million galaxies at  $z \sim 1$  in a physically motivated way.

Using these simulated universes and applying analysis techniques based on those in-

tended for use in the real data, I then explored the constraints on the dark energy equation of state from BAO measurements. One of the most important results presented is that 1% deviations from the underlying acoustic scale are possible within a single simulation due to sampling fluctuations. However, the expected value is, at the sub-percent level, an unbiased estimator of the acoustic scale. Finally, I made a forecast of the accuracy with which different proposed galaxy surveys will measure the dark energy equation of state.

Likewise, we emphasized the importance of accurate modelling of nonlinear effects to predict realistic clustering of galaxies, even on scales as large as those of the acoustic oscillations. In particular, we showed that i) the nonlinear growth of the dark matter power spectrum is evident even on scales of  $k < 0.03 h\text{Mpc}^{-1}$ . ii) The form of the redshift space distortion to the clustering depends on the type of object selected to trace the density field, being very different for galaxies, haloes, and dark matter. iii) The galaxy bias exhibits a scale dependence, for  $k > 0.15 h\text{Mpc}^{-1}$ , which is sensitive to the selection criteria applied.

In Chapter 3 I extended the study presented in Chapter 2 by a short analysis of the impact of photometric redshift errors on the BAO signature. The distortions induced in both the overall shape of the power spectrum and in the appearance of the BAO were shown, along with a simple analytic model to predict the damping in the power spectrum induced by photometric redshift errors. These results will be useful in the task of finding out how to optimally extract the information enclosed in photometric surveys.

## 8.2 Dark matter haloes and their substructures

In the second part of this thesis (chapters 4 and 5) I focused on more theoretical aspects of the clustering of dark matter haloes and in the kinematic of their substructures, both of which could have important implications for models of galaxy clustering.

The classic work of Press & Schechter (1974) predicts that the spatial distribution of halos is a function only of their mass. However, in recent years numerical simulations have reached sufficient precision to challenge this picture. The clustering strength of haloes, as measured by the 2-point correlation function and a linear bias parameter, is found to depend on other halo properties in addition to mass such as the age, spin and number of substructures. However, due to the challenging nature of the measurements, such dependencies were completely unexplored in higher order clustering. In chapter 6, I eschewed the traditional 2-point statistics applied in previous studies to extend the

analysis to higher orders. This achievement was made possible thanks to the huge volume covered by the BASICC run and, most importantly, to the development of a novel way to extract the higher order bias parameters.

The new method consists of cross-correlating density fluctuations in the halo distribution with fluctuations in the dark matter. Under the assumption that the halo density field is a smooth, local function of the underlying nonlinear density field, we show that i) this new estimator is equally sensitive to the high order parameters as the traditional auto-correlation functions and, ii) it has a much better noise performance due to the larger number density of dark matter particles in modern cosmological simulations, compared with the number density of haloes. Using this fast, simple and yet robust, estimator we were able to measure in narrow mass bins and at different redshifts the bias parameters up to order 4 for haloes corresponding to  $5\sigma$  peaks. By comparing samples of haloes with low and high values of the concentration parameter, we showed that the higher order bias parameters also depend on the concentration as well as mass.

The determination of the higher order bias parameters of dark matter haloes presented in this chapter will be useful for the modelling of galaxy clustering such as the halo occupation distribution models. At the same time, future surveys will cover volumes from which the accurate measurement of higher order moments will be possible, allowing a comparison with theoretical models. On a different front, an interesting application is to extend the use of higher order cross-correlations to other cosmological measurements, for instance, galaxy-cluster cross-correlations or the correlation of the microwave background anisotropies with the structure along the line of sight. The different permutations of cross moments would perhaps break degeneracies and provide useful information.

The connection between dark matter structures and galaxies is a pillar in linking observations and theory. Therefore the study of dark matter structures in general and of substructure within dark matter haloes in particular, is of great interest. In Chapter 5 I investigated the properties and fate of substructures within dark matter haloes. In the first half of the chapter we revisited several properties of the population of subhaloes in a simulation with better statistics than those used previously. I found that the subhalo mass function departs significantly from a power law for massive subhaloes. I also found that both the radial and angular distributions of substructures are functions of subhalo mass. In particular, high mass subhaloes tend to be less radially concentrated and to have angular distributions which are more perpendicular to the spin of the host halo than their less massive counterparts. In the second half of the chapter, we investigated the

whether mergers between substructures can occur. Such events could have important observational consequences. In fact, an interaction between substructures could trigger an episode of star formation or a morphological transformation. We found that mergers between subhaloes do occur and that they tend to involve substructures that were already dynamically associated before accretion into the main halo. In general, the rate of subhalo-subhalo mergers is lower than the rate at which subhaloes merge into the central halo, particularly for large substructures. However, for low mass subhaloes, these merger rates are equivalent and the subhalo-subhalo merger rate would be more likely, as there are more subhaloes to aim at. In the simulations, due to the finite dynamic range, we only see this scenario in practice in cluster-mass haloes, which are the best resolved. Since this type of interaction is usually neglected, our results have implications for the modelling of satellites in hierarchical galaxy formation models.

### 8.3 Improving current mock catalogues

In the last two chapters of this thesis I have presented new ideas under development to improve the modelling of the large scale structure in future observations.

As discussed in Chapter 7, another important layer of realism required in the theoretical predictions is a consistent incorporation of evolution of the clustering along the line of sight. Even though this effect was neglected in Chapters 2 and 3, we developed and presented a method to construct a halo light-cone in Chapter 7. In particular, our method will be of great importance in the modelling of current and future surveys such as DEEP2, Dark Energy Survey and Pan-STARRS (all of which will include galaxies over a vast redshift range) as well as to study cosmological probes sensitive to the evolution of the clustering.

In spite of the sophistication of our models, they are still limited in several aspects. For instance, the effects induced by the selection function and survey geometry were neglected. These ingredients can be incorporated easily to our current mock catalogues. A more pressing limitation of the models is their failure to reproduce the observed dependence of clustering strength with luminosity, which restricts the accuracy of the forecast errors. The solution to this problem may require a revision of the physical ingredients of the models.

A fundamental aspect in any measurement is a realistic treatment of errors. The information to be extracted from future datasets would be devoid of meaning if it is

not accompanied by a proper assessment of the uncertainties on the measurement. The method presented in Chapter 6 provided a means to address this issue by allowing the generation, at low computational cost, of halo catalogues from low resolution simulations. In this way, we are therefore able to produce a large number of realistic galaxy catalogues from different realizations of a primordial density field and hence assess the error on a given measurement. I expect the ideas presented in this chapter to help the understanding of uncertainties in measurements of the large scale structure of galaxies, and to improve methods to extract cosmological information from clustering measurements and even in the optimization of future experiments.

An immediate application of the ideas presented in these chapters would be their implementation into larger simulations to predict the location and properties of all the observable galaxies in the whole sky up to intermediate redshift. Resolving Milky-Way type haloes in a simulation that covers the whole sky up to redshift 1 is out of the reach for current supercomputers. However, the dynamic range can be extended using the method outlined in Chapter 6. Then, the construction of a halo light-cone would follow the method provided in Chapter 7. The galaxy population would be predicted using the semi-analytical model as described in Chapter 2.

The studies presented throughout this thesis have attempted improve the current understanding of several aspects of the clustering of the dark matter, haloes and galaxies. We consider that our findings will result in a more accurate modelling of the observed large scale structure of the Universe as well as an improved treatment of systematic and statistical errors. We believe that all this will yield to a more complete exploitation of future surveys as well as leading to the improvement of our knowledge of the Universe that surrounds us.



# Bibliography

- A. Albrecht, G. Bernstein, R. Cahn, W. L. Freedman, J. Hewitt, W. Hu, J. Huth, M. Kamionkowski, E. W. Kolb, L. Knox, J. C. Mather, S. Staggs, and N. B. Suntzeff. Report of the Dark Energy Task Force. *arXiv:astro-ph/0609591*, September 2006.
- C. Alcock and B. Paczynski. An evolution free test for non-zero cosmological constant. *Nature*, 281:358–+, October 1979.
- C. Almeida, C. M. Baugh, and C. G. Lacey. The structural and photometric properties of early-type galaxies in hierarchical models. *MNRAS*, 376:1711–1726, April 2007. doi: 10.1111/j.1365-2966.2007.11530.x.
- C. Almeida, C. M. Baugh, D. A. Wake, C. G. Lacey, A. J. Benson, R. G. Bower, and K. Pimbblet. Luminous red galaxies in hierarchical cosmologies. *MNRAS*, 386:2145–2160, June 2008. doi: 10.1111/j.1365-2966.2008.13179.x.
- L. Amendola, C. Quercellini, and E. Giallongo. Constraints on perfect fluid and scalar field dark energy models from future redshift surveys. *MNRAS*, 357:429–439, February 2005. doi: 10.1111/j.1365-2966.2004.08558.x.
- R. Angulo, C. M. Baugh, C. S. Frenk, R. G. Bower, A. Jenkins, and S. L. Morris. Constraints on the dark energy equation of state from the imprint of baryons on the power spectrum of clusters. *MNRAS*, 362:L25–L29, September 2005. doi: 10.1111/j.1745-3933.2005.00067.x.
- R. E. Angulo, C. M. Baugh, C. S. Frenk, and C. G. Lacey. The detectability of baryonic acoustic oscillations in future galaxy surveys. *MNRAS*, 383:755–776, January 2008. doi: 10.1111/j.1365-2966.2007.12587.x.
- G. W. Angus, B. Famaey, and H. S. Zhao. Can MOND take a bullet? Analytical comparisons of three versions of MOND beyond spherical symmetry. *MNRAS*, 371: 138–146, September 2006. doi: 10.1111/j.1365-2966.2006.10668.x.

- M. A. Aragón-Calvo, R. van de Weygaert, B. J. T. Jones, and J. M. van der Hulst. Spin Alignment of Dark Matter Halos in Filaments and Walls. *ApJ*, 655:L5–L8, January 2007. doi: 10.1086/511633.
- J. Ariel Keselman and A. Nusser. Halo assembly bias in the quasi-linear regime. *MNRAS*, 382:1853–1858, December 2007. doi: 10.1111/j.1365-2966.2007.12495.x.
- J. Barriga and E. Gaztañaga. The three-point function in large-scale structure - I. The weakly non-linear regime in N-body simulations. *MNRAS*, 333:443–453, June 2002. doi: 10.1046/j.1365-8711.2002.05431.x.
- C. M. Baugh. A primer on hierarchical galaxy formation: the semi-analytical approach. *Reports of Progress in Physics*, 69:3101–3156, December 2006. doi: 10.1088/0034-4885/69/12/R02.
- C. M. Baugh and G. Efstathiou. A Comparison of the Evolution of Density Fields in Perturbation Theory and Numerical Simulations - Part One - Nonlinear Evolution of the Power Spectrum. *MNRAS*, 270:183–+, September 1994.
- C. M. Baugh, E. Gaztanaga, and G. Efstathiou. A comparison of the evolution of density fields in perturbation theory and numerical simulations - II. Counts-in-cells analysis. *MNRAS*, 274:1049–1070, June 1995.
- C. M. Baugh, C. G. Lacey, C. S. Frenk, G. L. Granato, L. Silva, A. Bressan, A. J. Benson, and S. Cole. Can the faint submillimetre galaxies be explained in the  $\Lambda$  cold dark matter model? *MNRAS*, 356:1191–1200, January 2005. doi: 10.1111/j.1365-2966.2004.08553.x.
- C. M. Baugh et al. The 2dF Galaxy Redshift Survey: hierarchical galaxy clustering. *MNRAS*, 351:L44–L48, June 2004. doi: 10.1111/j.1365-2966.2004.07962.x.
- J. D. Bekenstein. Relativistic gravitation theory for the modified Newtonian dynamics paradigm. *Phys. Rev. D*, 70(8):083509–+, October 2004. doi: 10.1103/PhysRevD.70.083509.
- A. J. Benson, S. Cole, C. S. Frenk, C. M. Baugh, and C. G. Lacey. The nature of galaxy bias and clustering. *MNRAS*, 311:793–808, February 2000.
- A. J. Benson, C. S. Frenk, C. G. Lacey, C. M. Baugh, and S. Cole. The effects of photoionization on galaxy formation - II. Satellite galaxies in the Local Group. *MNRAS*, 333:177–190, June 2002. doi: 10.1046/j.1365-8711.2002.05388.x.

- A. J. Benson, R. G. Bower, C. S. Frenk, C. G. Lacey, C. M. Baugh, and S. Cole. What Shapes the Luminosity Function of Galaxies? *ApJ*, 599:38–49, December 2003. doi: 10.1086/379160.
- A. A. Berlind et al. The Halo Occupation Distribution and the Physics of Galaxy Formation. *ApJ*, 593:1–25, August 2003. doi: 10.1086/376517.
- F. Bernardeau. Skewness and kurtosis in large-scale cosmic fields. *ApJ*, 433:1–18, September 1994. doi: 10.1086/174620.
- F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro. Large-scale structure of the Universe and cosmological perturbation theory. *Phys. Rep.*, 367:1–3, September 2002.
- G. Bertone, D. Hooper, and J. Silk. Particle dark matter: evidence, candidates and constraints. *Phys. Rep.*, 405:279–390, January 2005. doi: 10.1016/j.physrep.2004.08.031.
- E. Bertschinger. Simulations of Structure Formation in the Universe. *ARA&A*, 36:599–654, 1998. doi: 10.1146/annurev.astro.36.1.599.
- P. Bett, V. Eke, C. S. Frenk, A. Jenkins, J. Helly, and J. Navarro. The spin and shape of dark matter haloes in the Millennium simulation of a  $\Lambda$  cold dark matter universe. *MNRAS*, 376:215–232, March 2007. doi: 10.1111/j.1365-2966.2007.11432.x.
- J. Blaizot, Y. Wadadekar, B. Guiderdoni, S. T. Colombi, E. Bertin, F. R. Bouchet, J. E. G. Devriendt, and S. Hatton. MoMaF: the Mock Map Facility. *MNRAS*, 360:159–175, June 2005. doi: 10.1111/j.1365-2966.2005.09019.x.
- C. Blake and S. Bridle. Cosmology with photometric redshift surveys. *MNRAS*, 363:1329–1348, November 2005. doi: 10.1111/j.1365-2966.2005.09526.x.
- C. Blake and K. Glazebrook. Probing Dark Energy Using Baryonic Oscillations in the Galaxy Power Spectrum as a Cosmological Ruler. *ApJ*, 594:665–673, September 2003. doi: 10.1086/376983.
- C. Blake, D. Parkinson, B. Bassett, K. Glazebrook, M. Kunz, and R. C. Nichol. Universal fitting formulae for baryon oscillation surveys. *MNRAS*, 365:255–264, January 2006. doi: 10.1111/j.1365-2966.2005.09706.x.

- C. Blake, A. Collister, S. Bridle, and O. Lahav. Cosmological baryonic and matter densities from 600000 SDSS luminous red galaxies with photometric redshifts. *MNRAS*, 374:1527–1548, February 2007. doi: 10.1111/j.1365-2966.2006.11263.x.
- R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey. Breaking the hierarchy of galaxy formation. *MNRAS*, 370:645–655, August 2006. doi: 10.1111/j.1365-2966.2006.10519.x.
- A. Cabre and E. Gaztanaga. Clustering of luminous red galaxies I: large scale redshift space distortions. *arXiv:astro-ph/0807.2460*, 807, July 2008.
- Y.-C. Cai, R. E. Angulo, C. M. Baugh, S. Cole, C. S. Frenk, and A. Jenkins. Mock galaxy redshift catalogues from simulations: implications for Pan-STARRS1. *arXiv:astro-ph/0810.2300*, October 2008.
- C. Carbone, L. Verde, and S. Matarrese. Non-Gaussian Halo Bias and Future Galaxy Surveys. *ApJ*, 684:L1–L4, September 2008. doi: 10.1086/592020.
- S. M. Carroll, W. H. Press, and E. L. Turner. The cosmological constant. *ARA&A*, 30:499–542, 1992. doi: 10.1146/annurev.aa.30.090192.002435.
- D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky. A Direct Empirical Proof of the Existence of Dark Matter. *ApJ*, 648:L109–L113, September 2006. doi: 10.1086/508162.
- J. D. Cohn. Power spectrum and correlation function errors: Poisson vs. Gaussian shot noise. *New Astronomy*, 11:226–239, January 2006. doi: 10.1016/j.newast.2005.08.002.
- J. M. Colberg and others. Clustering of galaxy clusters in cold dark matter universes. *MNRAS*, 319:209–214, November 2000.
- S. Cole and N. Kaiser. Biased clustering in the cold dark matter cosmogony. *MNRAS*, 237:1127–1146, April 1989.
- S. Cole and C. Lacey. The structure of dark matter haloes in hierarchical clustering models. *MNRAS*, 281:716–+, July 1996.
- S. Cole, S. Hatton, D. H. Weinberg, and C. S. Frenk. Mock 2dF and SDSS galaxy redshift surveys. *MNRAS*, 300:945–966, November 1998.

- S. Cole, C. G. Lacey, C. M. Baugh, and C. S. Frenk. Hierarchical galaxy formation. *MNRAS*, 319:168–204, November 2000.
- S. Cole et al. The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *MNRAS*, 362:505–534, September 2005. doi: 10.1111/j.1365-2966.2005.09318.x.
- M. Colless, G. Dalton, S. Maddox, W. Sutherland, P. Norberg, S. Cole, J. Bland-Hawthorn, T. Bridges, R. Cannon, C. Collins, W. Couch, N. Cross, K. Deeley, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, D. Madgwick, J. A. Peacock, B. A. Peterson, I. Price, M. Seaborne, and K. Taylor. The 2dF Galaxy Redshift Survey: spectra and redshifts. *MNRAS*, 328:1039–1063, December 2001. doi: 10.1046/j.1365-8711.2001.04902.x.
- A. Cooray and R. Sheth. Halo models of large scale structure. *Phys. Rep.*, 372:1–129, December 2002.
- M. Crocce and R. Scoccimarro. Renormalized cosmological perturbation theory. *Phys. Rev. D*, 73(6):063519–+, March 2006a. doi: 10.1103/PhysRevD.73.063519.
- M. Crocce and R. Scoccimarro. Memory of initial conditions in gravitational clustering. *Phys. Rev. D*, 73(6):063520–+, March 2006b. doi: 10.1103/PhysRevD.73.063520.
- M. Crocce, S. Pueblas, and R. Scoccimarro. Transients from initial conditions in cosmological simulations. *MNRAS*, 373:369–381, November 2006. doi: 10.1111/j.1365-2966.2006.11040.x.
- D. J. Croton, L. Gao, and S. D. M. White. Halo assembly bias and its effects on galaxy clustering. *MNRAS*, 374:1303–1309, February 2007. doi: 10.1111/j.1365-2966.2006.11230.x.
- D. J. Croton et al. The 2dF Galaxy Redshift Survey: higher-order galaxy correlation functions. *MNRAS*, 352:1232–1244, August 2004. doi: 10.1111/j.1365-2966.2004.08017.x.
- N. Dalal, M. White, J. R. Bond, and A. Shirokov. Halo Assembly Bias in Hierarchical Structure Formation. *arXiv:astro-ph/0803.3453*, 803, March 2008.
- M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. *ApJ*, 292:371–394, May 1985. doi: 10.1086/163168.

- M. Davis et al. The All-Wavelength Extended Groth Strip International Survey (AEGIS) Data Sets. *ApJ*, 660:L1–L6, May 2007. doi: 10.1086/517931.
- G. De Lucia et al. Substructures in cold dark matter haloes. *MNRAS*, 348:333–344, February 2004. doi: 10.1111/j.1365-2966.2004.07372.x.
- A. Dekel and O. Lahav. Stochastic Nonlinear Galaxy Biasing. *ApJ*, 520:24–34, July 1999. doi: 10.1086/307428.
- J. Diemand, B. Moore, and J. Stadel. Velocity and spatial biases in cold dark matter subhalo distributions. *MNRAS*, 352:535–546, August 2004. doi: 10.1111/j.1365-2966.2004.07940.x.
- J. Diemand, M. Kuhlen, and P. Madau. Formation and Evolution of Galaxy Dark Matter Halos and Their Substructure. *ApJ*, 667:859–877, October 2007. doi: 10.1086/520573.
- J. Diemand, M. Kuhlen, P. Madau, M. Zemp, B. Moore, D. Potter, and J. Stadel. Clumps and streams in the local dark matter distribution. *Nature*, 454:735–738, August 2008. doi: 10.1038/nature07153.
- D. Dolney, B. Jain, and M. Takada. Baryon oscillations and dark-energy constraints from imaging surveys. *MNRAS*, 366:884–898, March 2006. doi: 10.1111/j.1365-2966.2005.09606.x.
- G. Efstathiou. Comets to Cosmology; Proceedings of the Third IRAS Conference. In A Lawrence, editor, *Comets to Cosmology; Proceedings of the Third IRAS Conference*, page 312, 1988.
- G. Efstathiou, M. Davis, S. D. M. White, and C. S. Frenk. Numerical techniques for large cosmological N-body simulations. *ApJS*, 57:241–260, February 1985. doi: 10.1086/191003.
- D. J. Eisenstein and W. Hu. Baryonic Features in the Matter Transfer Function. *ApJ*, 496:605–+, March 1998. doi: 10.1086/305424.
- D. J. Eisenstein, H.-J. Seo, E. Sirko, and D. N. Spergel. Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak. *ApJ*, 664:675–679, August 2007a. doi: 10.1086/518712.

- D. J. Eisenstein, H.-J. Seo, and M. White. On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter. *ApJ*, 664:660–674, August 2007b. doi: 10.1086/518755.
- D. J. Eisenstein et al. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, 633:560–574, November 2005. doi: 10.1086/466512.
- H. K. Eriksen, P. B. Lilje, A. J. Banday, and K. M. Górski. Estimating N-Point Correlation Functions from Pixelized Sky Maps. *ApJS*, 151:1–11, March 2004. doi: 10.1086/381740.
- N. Espino-Briones, M. Plionis, and C. Ragone-Figueroa. Environmental Effects of Dark Matter Halos: The Clustering-Substructure Relation of Group-Size Halos. *ApJ*, 666:L5–L8, September 2007. doi: 10.1086/521622.
- A. E. Evrard, T. J. MacFarland, H. M. P. Couchman, J. M. Colberg, N. Yoshida, S. D. M. White, A. Jenkins, C. S. Frenk, F. R. Pearce, J. A. Peacock, and P. A. Thomas. Galaxy Clusters in Hubble Volume Simulations: Cosmological Constraints from Sky Survey Populations. *ApJ*, 573:7–36, July 2002. doi: 10.1086/340551.
- H. A. Feldman, N. Kaiser, and J. A. Peacock. Power-spectrum analysis of three-dimensional redshift surveys. *ApJ*, 426:23–37, May 1994. doi: 10.1086/174036.
- P. Fosalba, E. Gaztanaga, F. Castander, and M. Manera. The onion universe: all sky lightcone simulations in spherical shells. *arXiv:astro-ph/0711.1540*, 711, November 2007.
- C. S. Frenk, S. D. M. White, G. Efstathiou, and M. Davis. Cold dark matter, the structure of galactic haloes and the origin of the Hubble sequence. *Nature*, 317:595–597, October 1985.
- C. S. Frenk, S. D. M. White, M. Davis, and G. Efstathiou. The formation of dark halos in a universe dominated by cold dark matter. *ApJ*, 327:507–525, April 1988. doi: 10.1086/166213.
- W. J. Frith, P. J. Outram, and T. Shanks. High-order 2MASS galaxy correlation functions: probing the Gaussianity of the primordial density field. *MNRAS*, 373:759–768, December 2006. doi: 10.1111/j.1365-2966.2006.11066.x.

- J. N. Fry and E. Gaztanaga. Biasing and hierarchical statistics in large-scale structure. *ApJ*, 413:447–452, August 1993. doi: 10.1086/173015.
- L. Gao and S. D. M. White. Assembly bias in the clustering of dark matter haloes. *MNRAS*, 377:L5–L9, April 2007. doi: 10.1111/j.1745-3933.2007.00292.x.
- L. Gao, G. De Lucia, S. D. M. White, and A. Jenkins. Galaxies and subhaloes in  $\Lambda$ CDM galaxy clusters. *MNRAS*, 352:L1–L5, August 2004. doi: 10.1111/j.1365-2966.2004.08098.x.
- L. Gao, V. Springel, and S. D. M. White. The age dependence of halo clustering. *MNRAS*, 363:L66–L70, October 2005. doi: 10.1111/j.1745-3933.2005.00084.x.
- E. Gaztanaga and C. M. Baugh. Hierarchical correlations in models of galaxy clustering. *MNRAS*, 273:L1–L6, March 1995.
- E. Gaztanaga, R. Miquel, and E. Sanchez. First Cosmological Constraints on Dark Energy from the Radial Baryon Acoustic Scale. *arXiv:astro-ph/0808.1921*, 808, August 2008.
- S. Ghigna, B. Moore, F. Governato, G. Lake, T. Quinn, and J. Stadel. Dark matter haloes within clusters. *MNRAS*, 300:146–162, October 1998.
- S. Ghigna, B. Moore, F. Governato, G. Lake, T. Quinn, and J. Stadel. Density Profiles and Substructure of Dark Matter Halos: Converging Results at Ultra-High Numerical Resolution. *ApJ*, 544:616–628, December 2000. doi: 10.1086/317221.
- C. Giocoli, G. Tormen, and F. C. van den Bosch. The population of dark matter subhaloes: mass functions and average mass-loss rates. *MNRAS*, 386:2135–2144, June 2008. doi: 10.1111/j.1365-2966.2008.13182.x.
- K. Glazebrook and C. Blake. Measuring the Cosmic Evolution of Dark Energy with Baryonic Oscillations in the Galaxy Power Spectrum. *ApJ*, 631:1–20, September 2005. doi: 10.1086/432497.
- K. Glazebrook et al. Dark Energy and Cosmic Sound:  $w(z)$  Surveys with the Gemini/Subaru Wide-Field Multi-Object Spectrograph. *preprints (astro-ph/0507457)*, July 2005.
- K. Glazebrook et al. The WiggleZ project:  $\Lambda$ CDM and Dark Energy. *preprints (astro-ph/0701876)*, January 2007.

- F. Governato, A. Babul, T. Quinn, P. Tozzi, C. M. Baugh, N. Katz, and G. Lake. Properties of galaxy clusters: mass and correlation functions. *MNRAS*, 307:949–966, August 1999.
- A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, January 1981.
- L. Guzzo et al. A test of the nature of cosmic acceleration using galaxy redshift distortions. *Nature*, 451:541–544, January 2008. doi: 10.1038/nature06555.
- Z. Haiman, J. J. Mohr, and G. P. Holder. Constraints on Cosmological Parameters from Future Galaxy Cluster Surveys. *ApJ*, 553:545–561, June 2001. doi: 10.1086/320939.
- T. Hamana, N. Yoshida, Y. Suto, and A. E. Evrard. Clustering of Dark Matter Halos on the Light Cone: Scale, Time, and Mass Dependence of the Halo Biasing in the Hubble Volume Simulations. *ApJ*, 561:L143–L146, November 2001.
- A. J. S. Hamilton, P. Kumar, E. Lu, and A. Matthews. Reconstructing the primordial spectrum of fluctuations of the universe from the observed nonlinear clustering of galaxies. *ApJ*, 374:L1–L4, June 1991. doi: 10.1086/186057.
- G. Harker, S. Cole, J. Helly, C. Frenk, and A. Jenkins. A marked correlation function analysis of halo formation times in the Millennium Simulation. *MNRAS*, 367:1039–1049, April 2006. doi: 10.1111/j.1365-2966.2006.10022.x.
- S. Hatton, J. E. G. Devriendt, S. Ninin, F. R. Bouchet, B. Guiderdoni, and D. Vibert. GALICS- I. A hybrid N-body/semi-analytic model of hierarchical galaxy formation. *MNRAS*, 343:75–106, July 2003. doi: 10.1046/j.1365-8711.2003.05589.x.
- S. J. Hatton. PhD Thesis. *PhD Thesis, University of Durham*, 1999.
- D. J. Heath. The growth of density perturbations in zero pressure Friedmann-Lemaitre universes. *MNRAS*, 179:351–358, May 1977.
- R. W. Hockney and J. W. Eastwood. *Computer Simulation Using Particles*. Computer Simulation Using Particles, New York: McGraw-Hill, 1981, 1981.
- F. Hoyle, I. Szapudi, and C. M. Baugh. Higher order clustering in the Durham/UKST and Stromlo-APM Galaxy Redshift Surveys. *MNRAS*, 317:L51–L56, September 2000.

- W. Hu and Z. Haiman. Redshifting rings of power. *Phys. Rev. D*, 68(6):063004–+, September 2003.
- W. Hu and N. Sugiyama. Small-Scale Cosmological Perturbations: an Analytic Approach. *ApJ*, 471:542–+, November 1996. doi: 10.1086/177989.
- E. Huff, A. E. Schulz, M. White, D. J. Schlegel, and M. S. Warren. Simulations of baryon oscillations. *Astroparticle Physics*, 26:351–366, January 2007. doi: 10.1016/j.astropartphys.2006.07.007.
- D. Huterer, M. Takada, G. Bernstein, and B. Jain. Systematic errors in future weak-lensing surveys: requirements and prospects for self-calibration. *MNRAS*, 366:101–114, February 2006. doi: 10.1111/j.1365-2966.2005.09782.x.
- G. Hütsi. Power spectrum of the SDSS luminous red galaxies: constraints on cosmological parameters. *A&A*, 459:375–389, November 2006a. doi: 10.1051/0004-6361:20065377.
- G. Hütsi. Acoustic oscillations in the SDSS DR4 luminous red galaxy sample power spectrum. *A&A*, 449:891–902, April 2006b. doi: 10.1051/0004-6361:20053939.
- B. Jain and E. Bertschinger. Second-order power spectrum and nonlinear evolution at high redshift. *ApJ*, 431:495–505, August 1994. doi: 10.1086/174502.
- A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida. The mass function of dark matter haloes. *MNRAS*, 321: 372–384, February 2001.
- C. Y. Jiang, Y. P. Jing, A. Faltenbacher, W. P. Lin, and C. Li. A Fitting Formula for the Merger Timescale of Galaxies in Hierarchical Clustering. *ApJ*, 675:1095–1105, March 2008. doi: 10.1086/526412.
- Y. P. Jing. Accurate Fitting Formula for the Two-Point Correlation Function of Dark Matter Halos. *ApJ*, 503:L9+, August 1998. doi: 10.1086/311530.
- Y. P. Jing. Correcting for the Alias Effect When Measuring the Power Spectrum Using a Fast Fourier Transform. *ApJ*, 620:559–563, February 2005. doi: 10.1086/427087.
- Y. P. Jing, Y. Suto, and H. J. Mo. The Dependence of Dark Halo Clustering on Formation Epoch and Concentration Parameter. *ApJ*, 657:664–668, March 2007. doi: 10.1086/511130.

- R. Juszkiewicz, F. R. Bouchet, and S. Colombi. Skewness induced by gravity. *ApJ*, 412: L9–L12, July 1993. doi: 10.1086/186927.
- N. Kaiser. On the spatial correlations of Abell clusters. *ApJ*, 284:L9–L12, September 1984. doi: 10.1086/184341.
- N. Kaiser. Clustering in real space and in redshift space. *MNRAS*, 227:1–21, July 1987.
- X. Kang, F. C. van den Bosch, X. Yang, S. Mao, H. J. Mo, C. Li, and Y. P. Jing. The alignment between satellites and central galaxies: theory versus observations. *MNRAS*, 378:1531–1542, July 2007. doi: 10.1111/j.1365-2966.2007.11902.x.
- G. Kauffmann, A. Nusser, and M. Steinmetz. Galaxy formation and large-scale bias. *MNRAS*, 286:795–811, April 1997.
- M. G. Kitzbichler and S. D. M. White. The high-redshift galaxy population in hierarchical galaxy formation models. *MNRAS*, 376:2–12, March 2007. doi: 10.1111/j.1365-2966.2007.11458.x.
- A. Klypin, S. Gottlöber, A. V. Kravtsov, and A. M. Khokhlov. Galaxies in N-Body Simulations: Overcoming the Overmerging Problem. *ApJ*, 516:530–551, May 1999a. doi: 10.1086/307122.
- A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada. Where Are the Missing Galactic Satellites? *ApJ*, 522:82–92, September 1999b. doi: 10.1086/307643.
- A. Knebe, S. P. D. Gill, B. K. Gibson, G. F. Lewis, R. A. Ibata, and M. A. Dopita. Anisotropy in the Distribution of Satellite Galaxy Orbits. *ApJ*, 603:7–11, March 2004. doi: 10.1086/381306.
- A. Knebe, H. Yahagi, H. Kase, G. Lewis, and B. K. Gibson. The radial alignment of dark matter subhalos: from simulations to observations. *arXiv:astro-ph/0805.1823*, 805, May 2008a.
- A. Knebe et al. On the relation between the radial alignment of dark matter subhaloes and host mass in cosmological simulations. *MNRAS*, 386:L52–L56, May 2008b. doi: 10.1111/j.1745-3933.2008.00459.x.
- C. S. Kochanek and N. Dalal. Tests for Substructure in Gravitational Lenses. *ApJ*, 610: 69–79, July 2004. doi: 10.1086/421436.

- R. S. Koehler, P. Schuecker, and K. Gebhardt. Probing dark energy with baryonic acoustic oscillations at high redshifts. *A&A*, 462:7–20, January 2007. doi: 10.1051/0004-6361:20065649.
- E. Komatsu, J. Dunkley, M. R.olta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, S. S. Meyer, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *arXiv:astro-ph/0803.0547*, 803, March 2008.
- C. Lacey and S. Cole. Merger rates in hierarchical models of galaxy formation. *MNRAS*, 262:627–649, June 1993.
- S. D. Landy and A. S. Szalay. Bias and variance of angular correlation functions. *ApJ*, 412:64–71, July 1993. doi: 10.1086/172900.
- M. Le Delliou, C. Lacey, C. M. Baugh, B. Guiderdoni, R. Bacon, H. Courtois, T. Sousbie, and S. L. Morris. The abundance of Ly $\alpha$  emitters in hierarchical models. *MNRAS*, 357:L11–L15, February 2005. doi: 10.1111/j.1745-3933.2005.00007.x.
- M. Le Delliou, C. G. Lacey, C. M. Baugh, and S. L. Morris. The properties of Ly $\alpha$  emitting galaxies in hierarchical galaxy formation models. *MNRAS*, 365:712–726, January 2006. doi: 10.1111/j.1365-2966.2005.09797.x.
- G. Lemson and G. Kauffmann. Environmental influences on dark matter haloes and consequences for the galaxies within them. *MNRAS*, 302:111–117, January 1999.
- A. Lewis, A. Challinor, and A. Lasenby. Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. *ApJ*, 538:473–476, August 2000. doi: 10.1086/309179.
- N. I. Libeskind, S. Cole, C. S. Frenk, T. Okamoto, and A. Jenkins. Satellite systems around galaxies in hydrodynamic simulations. *MNRAS*, 374:16–28, January 2007. doi: 10.1111/j.1365-2966.2006.11205.x.
- E. V. Linder. Cosmic growth history and expansion history. *Phys. Rev. D*, 72(4):043529–+, August 2005. doi: 10.1103/PhysRevD.72.043529.
- E. V. Linder and R. N. Cahn. Parameterized Beyond-Einstein Growth. *preprints (astro-ph/0701317)*, January 2007.

- J. Makino and P. Hut. Merger Rate of Equal-Mass Spherical Galaxies. *ApJ*, 481:83–+, May 1997. doi: 10.1086/304013.
- S. Matarrese and L. Verde. The Effect of Primordial Non-Gaussianity on Halo Bias. *ApJ*, 677:L77–L80, April 2008. doi: 10.1086/587840.
- S. Matarrese, P. Coles, F. Lucchin, and L. Moscardini. Redshift evolution of clustering. *MNRAS*, 286:115–132, March 1997.
- A. Meiksin, M. White, and J. A. Peacock. Baryonic signatures in large-scale structure. *MNRAS*, 304:851–864, April 1999.
- A. Mesinger and S. Furlanetto. Efficient Simulations of Early Structure Formation and Reionization. *ApJ*, 669:663–675, November 2007. doi: 10.1086/521806.
- P. Meszaros. The behaviour of point masses in an expanding cosmological substratum. *A&A*, 37:225–228, December 1974.
- H. J. Mo and S. D. M. White. An analytic model for the spatial clustering of dark matter haloes. *MNRAS*, 282:347–361, September 1996.
- H. J. Mo, Y. P. Jing, and S. D. M. White. High-order correlations of peaks and haloes: a step towards understanding galaxy biasing. *MNRAS*, 284:189–201, January 1997.
- P. Monaco, T. Theuns, and G. Taffoni. The pinocchio algorithm: pinpointing orbit-crossing collapsed hierarchical objects in a linear density field. *MNRAS*, 331:587–608, April 2002. doi: 10.1046/j.1365-8711.2002.05162.x.
- B. Moore, N. Katz, and G. Lake. On the Destruction and Overmerging of Dark Halos in Dissipationless N-Body Simulations. *ApJ*, 457:455–+, February 1996. doi: 10.1086/176745.
- B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel, and P. Tozzi. Dark Matter Substructure within Galactic Halos. *ApJ*, 524:L19–L22, October 1999. doi: 10.1086/312287.
- J. A. Muñoz and A. Loeb. Light-cone distortion of the clustering and abundance of massive galaxies at high redshifts. *MNRAS*, 386:2323–2329, June 2008. doi: 10.1111/j.1365-2966.2008.13210.x.

- D. Nagai and A. V. Kravtsov. The Radial Distribution of Galaxies in  $\Lambda$  Cold Dark Matter Clusters. *ApJ*, 618:557–568, January 2005. doi: 10.1086/426016.
- M. Nagashima, C. G. Lacey, C. M. Baugh, C. S. Frenk, and S. Cole. The metal enrichment of the intracluster medium in hierarchical galaxy formation models. *MNRAS*, 358:1247–1266, April 2005a. doi: 10.1111/j.1365-2966.2005.08766.x.
- M. Nagashima, C. G. Lacey, T. Okamoto, C. M. Baugh, C. S. Frenk, and S. Cole. The metal enrichment of elliptical galaxies in hierarchical galaxy formation models. *MNRAS*, 363:L31–L35, October 2005b. doi: 10.1111/j.1745-3933.2005.00078.x.
- J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. *ApJ*, 490:493–+, December 1997. doi: 10.1086/304888.
- A. F. Neto et al. The statistics of  $\Lambda$  CDM halo concentrations. *MNRAS*, 381:1450–1462, November 2007. doi: 10.1111/j.1365-2966.2007.12381.x.
- R. C. Nichol et al. The effect of large-scale structure on the SDSS galaxy three-point correlation function. *MNRAS*, 368:1507–1514, June 2006. doi: 10.1111/j.1365-2966.2006.10239.x.
- N. D. Padilla and C. M. Baugh. Cluster correlations in redshift space. *MNRAS*, 329:431–444, January 2002. doi: 10.1046/j.1365-8711.2002.04992.x.
- N. Padmanabhan et al. The Clustering of Luminous Red Galaxies in the Sloan Digital Sky Survey Imaging Data. *arXiv:astro-ph/0605302*, May 2007.
- T. Padmanabhan. *Structure formation in the Universe*. Cambridge Univ. Press, Cambridge, UK, June 1993. ISBN 0521424860.
- D. Parkinson, C. Blake, M. Kunz, B. A. Bassett, R. C. Nichol, and K. Glazebrook. Optimizing baryon acoustic oscillation surveys - I. Testing the concordance  $\Lambda$ CDM cosmology. *MNRAS*, 377:185–197, May 2007. doi: 10.1111/j.1365-2966.2007.11586.x.
- J. Peacock and P. Schneider. The ESO-ESA Working Group on Fundamental Cosmology. *The Messenger*, 125:48–+, September 2006.
- J. A. Peacock. *Cosmological Physics*. *Cosmological Physics*, by John A. Peacock, pp. 704. ISBN 052141072X. Cambridge, UK: Cambridge University Press, January 1999., January 1999.

- J. A. Peacock and S. J. Dodds. Reconstructing the Linear Power Spectrum of Cosmological Mass Fluctuations. *MNRAS*, 267:1020–+, April 1994.
- J. A. Peacock and S. J. Dodds. Non-linear evolution of cosmological power spectra. *MNRAS*, 280:L19–L26, June 1996.
- J. A. Peacock and R. E. Smith. Halo occupation numbers and galaxy bias. *MNRAS*, 318: 1144–1156, November 2000.
- P. J. E. Peebles. *The large-scale structure of the universe*. Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p., 1980.
- P. J. E. Peebles. *Principles of physical cosmology*. Princeton Series in Physics. Princeton University Press, Princeton, NJ, USA, 1993.
- P. J. E. Peebles and J. T. Yu. Primeval Adiabatic Perturbation in an Expanding Universe. *ApJ*, 162:815–+, December 1970.
- W. J. Percival and M. White. Testing cosmological structure formation using redshift-space distortions. *arXiv:astro-ph/0808.0003*, 808, July 2008.
- W. J. Percival, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, and et al Cannon. The 2dF Galaxy Redshift Survey: the power spectrum and the matter content of the Universe. *MNRAS*, 327:1297–1306, November 2001. doi: 10.1046/j.1365-8711.2001.04827.x.
- W. J. Percival, D. Scott, J. A. Peacock, and J. S. Dunlop. The clustering of halo mergers. *MNRAS*, 338:L31–L35, February 2003. doi: 10.1046/j.1365-8711.2003.06317.x.
- W. J. Percival, R. C. Nichol, D. J. Eisenstein, D. H. Weinberg, M. Fukugita, A. C. Pope, D. P. Schneider, A. S. Szalay, M. S. Vogeley, I. Zehavi, N. A. Bahcall, J. Brinkmann, A. J. Connolly, J. Loveday, and A. Meiksin. Measuring the matter density using baryon oscillations in the SDSS. *arXiv:astro-ph/0608635*, August 2007.
- S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J.

- Couch, and The Supernova Cosmology Project. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *ApJ*, 517:565–586, June 1999. doi: 10.1086/307221.
- C. Power, J. F. Navarro, A. Jenkins, C. S. Frenk, S. D. M. White, V. Springel, J. Stadel, and T. Quinn. The inner structure of  $\Lambda$ CDM haloes - I. A numerical convergence study. *MNRAS*, 338:14–34, January 2003. doi: 10.1046/j.1365-8711.2003.05925.x.
- W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438, February 1974.
- W. H. Press and E. T. Vishniac. Propagation of adiabatic cosmological perturbations through the ERA of matter-radiation decoupling. *ApJ*, 236:323–334, March 1980. doi: 10.1086/157749.
- A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116:1009–1038, September 1998. doi: 10.1086/300499.
- A. G. Riess, L.-G. Strolger, J. Tonry, S. Casertano, H. C. Ferguson, B. Mobasher, P. Challis, A. V. Filippenko, S. Jha, W. Li, R. Chornock, R. P. Kirshner, B. Leibundgut, M. Dickinson, M. Livio, M. Giavalisco, C. C. Steidel, T. Benítez, and Z. Tsvetanov. Type Ia Supernova Discoveries at  $z \lesssim 1$  from the Hubble Space Telescope: Evidence for Past Deceleration and Constraints on Dark Energy Evolution. *ApJ*, 607:665–687, June 2004. doi: 10.1086/383612.
- A. G. Sánchez, C. M. Baugh, W. J. Percival, J. A. Peacock, N. D. Padilla, S. Cole, C. S. Frenk, and P. Norberg. Cosmological parameters from cosmic microwave background measurements and the final 2dF Galaxy Redshift Survey power spectrum. *MNRAS*, 366:189–207, February 2006. doi: 10.1111/j.1365-2966.2005.09833.x.
- A. G. Sanchez, C. M. Baugh, and R. Angulo. What is the best way to measure baryonic acoustic oscillations? *arXiv:astro-ph/0804.0233*, 804, April 2008.
- A. E. Schulz and M. White. Scale-dependent bias and the halo model. *Astroparticle Physics*, 25:172–177, March 2006. doi: 10.1016/j.astropartphys.2005.11.007.

- R. Scoccimarro. Redshift-space distortions, pairwise velocities, and nonlinearities. *Phys. Rev. D*, 70(8):083007–+, October 2004. doi: 10.1103/PhysRevD.70.083007.
- R. Scoccimarro and R. K. Sheth. PTHALOS: a fast method for generating mock galaxy distributions. *MNRAS*, 329:629–640, January 2002. doi: 10.1046/j.1365-8711.2002.04999.x.
- R. Scoccimarro, R. K. Sheth, L. Hui, and B. Jain. How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering. *ApJ*, 546:20–34, January 2001. doi: 10.1086/318261.
- U. Seljak. Analytic model for galaxy and dark matter clustering. *MNRAS*, 318:203–213, October 2000.
- U. Seljak. Measuring primordial non-gaussianity without cosmic variance. *arXiv:astro-ph/0807.1770*, 807, July 2008.
- U. Seljak and M. S. Warren. Large-scale bias and stochasticity of haloes and dark matter. *MNRAS*, 355:129–136, November 2004. doi: 10.1111/j.1365-2966.2004.08297.x.
- H.-J. Seo and D. J. Eisenstein. Probing Dark Energy with Baryonic Acoustic Oscillations from Future Large Galaxy Redshift Surveys. *ApJ*, 598:720–740, December 2003. doi: 10.1086/379122.
- H.-J. Seo and D. J. Eisenstein. Baryonic Acoustic Oscillations in Simulated Galaxy Redshift Surveys. *ApJ*, 633:575–588, November 2005. doi: 10.1086/491599.
- H.-J. Seo and D. J. Eisenstein. Improved forecasts for the baryon acoustic oscillations and cosmological distance scale. *arXiv:astro-ph/0807.1770*, January 2007.
- L. D. Shaw, J. Weller, J. P. Ostriker, and P. Bode. Statistics of Physical Properties of Dark Matter Clusters. *ApJ*, 646:815–833, August 2006. doi: 10.1086/505016.
- L. D. Shaw, J. Weller, J. P. Ostriker, and P. Bode. The Bound Mass of Substructures in Dark Matter Halos. *ApJ*, 659:1082–1095, April 2007. doi: 10.1086/511849.
- R. K. Sheth and G. Tormen. Large-scale bias and the peak background split. *MNRAS*, 308:119–126, September 1999.
- R. K. Sheth and G. Tormen. On the environmental dependence of halo formation. *MNRAS*, 350:1385–1390, June 2004. doi: 10.1111/j.1365-2966.2004.07733.x.

- R. K. Sheth, H. J. Mo, and G. Tormen. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *MNRAS*, 323:1–12, May 2001. doi: 10.1046/j.1365-8711.2001.04006.x.
- R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman. Stable clustering, the halo model and non-linear cosmological power spectra. *MNRAS*, 341:1311–1332, June 2003. doi: 10.1046/j.1365-8711.2003.06503.x.
- R. E. Smith, R. Scoccimarro, and R. K. Sheth. Scale dependence of halo and galaxy bias: Effects in real space. *Phys. Rev. D*, 75(6):063512–+, March 2007. doi: 10.1103/PhysRevD.75.063512.
- R. E. Smith, R. Scoccimarro, and R. K. Sheth. Motion of the acoustic peak in the correlation function. *Phys. Rev. D*, 77(4):043525–+, February 2008. doi: 10.1103/PhysRevD.77.043525.
- G. F. Smoot et al. Structure in the COBE differential microwave radiometer first-year maps. *ApJ*, 396:L1–L5, September 1992. doi: 10.1086/186504.
- R. S. Somerville and J. R. Primack. Semi-analytic modelling of galaxy formation: the local Universe. *MNRAS*, 310:1087–1110, December 1999.
- D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters. *ApJS*, 148:175–194, September 2003. doi: 10.1086/377226.
- D. N. Spergel, R. Bean, O. Dore', M. R. Nolta, C. L. Bennett, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H. V. Peiris, L. Verde, C. Barnes, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Wilkinson Microwave Anisotropy Probe (WMAP) Three Year Results: Implications for Cosmology. *arXiv:astro-ph/0807.1770*, March 2007.
- V. Springel. The cosmological simulation code GADGET-2. *MNRAS*, 364:1105–1134, December 2005. doi: 10.1111/j.1365-2966.2005.09655.x.

- V. Springel, S. D. M. White, G. Tormen, and G. Kauffmann. Populating a cluster of galaxies - I. Results at  $z=0$ . *MNRAS*, 328:726–750, December 2001. doi: 10.1046/j.1365-8711.2001.04912.x.
- V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, June 2005. doi: 10.1038/nature03597.
- V. Springel, C. S. Frenk, and S. D. M. White. The large-scale structure of the Universe. *Nature*, 440:1137–1144, April 2006. doi: 10.1038/nature04805.
- V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk, and S. D. M. White. The Aquarius Project: the subhalos of galactic halos. *arXiv:astro-ph/0807.1770*, September 2008.
- A. A. Starobinskij. Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations. *Physics Letters B*, 117:175–178, 1982.
- F. Stoehr, S. D. M. White, G. Tormen, and V. Springel. The satellite population of the Milky Way in a  $\Lambda$ CDM universe. *MNRAS*, 335:L84–L88, October 2002. doi: 10.1046/j.1365-8711.2002.05891.x.
- R. A. Sunyaev and Y. B. Zel'dovich. Small-Scale Fluctuations of Relic Radiation. *Ap&SS*, 7:3–+, 1970.
- M. Tegmark et al. Cosmological constraints from the SDSS luminous red galaxies. *Phys. Rev. D*, 74(12):123507–+, December 2006. doi: 10.1103/PhysRevD.74.123507.
- R. Teyssier, S. Pires, S. Prunet, D. Aubert, C. Pichon, A. Amara, K. Benabed, S. Colombi, A. Refregier, and J.-L. Starck. Full-Sky Weak Lensing Simulation with 70 Billion Particles. *arXiv:astro-ph/0807.3651*, 807, July 2008.
- G. Tormen. The rise and fall of satellites in galaxy clusters. *MNRAS*, 290:411–421, September 1997.
- G. Tormen, A. Diaferio, and D. Syer. Survival of substructure within dark matter haloes. *MNRAS*, 299:728–742, September 1998.

- F. C. van den Bosch, G. Tormen, and C. Giocoli. The mass function and average mass-loss rate of dark matter subhaloes. *MNRAS*, 359:1029–1040, May 2005. doi: 10.1111/j.1365-2966.2005.08964.x.
- D. A. Wake, R. K. Sheth, R. C. Nichol, C. M. Baugh, J. Bland-Hawthorn, M. Colless, W. J. Couch, S. M. Croom, R. de Propris, M. J. Drinkwater, A. C. Edge, J. Loveday, T. Y. Lam, K. A. Pimbblet, I. G. Roseboom, N. P. Ross, D. P. Schneider, T. Shanks, and R. G. Sharp. The 2dF-SDSS LRG and QSO Survey: evolution of the clustering of luminous red galaxies since  $z = 0.6$ . *MNRAS*, 387:1045–1062, July 2008. doi: 10.1111/j.1365-2966.2008.13333.x.
- R. H. Wechsler, A. R. Zentner, J. S. Bullock, A. V. Kravtsov, and B. Allgood. The Dependence of Halo Clustering on Halo Formation History, Concentration, and Occupation. *ApJ*, 652:71–84, November 2006. doi: 10.1086/507120.
- A. R. Wetzel, J. D. Cohn, M. White, D. E. Holz, and M. S. Warren. The Clustering of Massive Halos. *ApJ*, 656:139–147, February 2007. doi: 10.1086/510444.
- S. D. M. White. Formation and Evolution of Galaxies: Les Houches Lectures. *arXiv:astro-ph/9410043*, October 1994.
- S. D. M. White and M. J. Rees. Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering. *MNRAS*, 183:341–358, May 1978.
- W. M. Wood-Vasey et al. Observational Constraints on the Nature of Dark Energy: First Cosmological Results from the ESSENCE Supernova Survey. *ApJ*, 666:694–715, September 2007. doi: 10.1086/518642.
- K. Yamamoto and Y. Suto. Two-Point Correlation Function of High-Redshift Objects: an Explicit Formulation on a Light-Cone Hypersurface. *ApJ*, 517:1–12, May 1999. doi: 10.1086/307184.
- D. G. York et al. The Sloan Digital Sky Survey: Technical Summary. *AJ*, 120:1579–1587, September 2000. doi: 10.1086/301513.
- Y. B. Zel'dovich. Gravitational instability: an approximate theory for large density perturbations. *A&A*, 5:84–89, March 1970.

- A. R. Zentner. The Excursion Set Theory of Halo Mass Functions, Halo Clustering, and Halo Growth. *International Journal of Modern Physics D*, 16:763–815, 2007. doi: 10.1142/S0218271807010511.
- A. R. Zentner, A. A. Berlind, J. S. Bullock, A. V. Kravtsov, and R. H. Wechsler. The Physics of Galaxy Clustering. I. A Model for Subhalo Populations. *ApJ*, 624:505–525, May 2005. doi: 10.1086/428898.
- D. H. Zhao, H. J. Mo, Y. P. Jing, and G. Börner. The growth and structure of dark matter haloes. *MNRAS*, 339:12–24, February 2003. doi: 10.1046/j.1365-8711.2003.06135.x.
- Z. Zheng, A. A. Berlind, D. H. Weinberg, A. J. Benson, C. M. Baugh, S. Cole, R. Davé, C. S. Frenk, N. Katz, and C. G. Lacey. Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies. *ApJ*, 633:791–809, November 2005. doi: 10.1086/466510.

