The clustering of galaxies in hierarchical galaxy formation models

HAN SIK KIM

The clustering of galaxies in hierarchical galaxy formation models

Hansik Kim

A Thesis presented for the degree of
Doctor of Philosophy

Institute for Computational Cosmology
Department of Physics
University of Durham
England

September 2010

The clustering of galaxies in hierarchical galaxy formation models

Hansik Kim

Submitted for the degree of Doctor of Philosophy
September 2010

Abstract

Galaxy clustering encodes information about the values of cosmological parameters and also about the physical processes behind galaxy formation and evolution. The GALFORM semi-analytical model is the theoretical approach we used to model galaxy formation. We start by studying the luminosity dependence of galaxy clustering which is measured accurately in the local Universe. We have compared the clustering predictions of three publicly available galaxy formation models with clustering measurements from the 2dFGRS and found that two new processes need to be included in order to understand the observed clustering. We then study the distribution of cold gas in dark matter haloes central to the processes of galaxy formation. We present the cold gas mass function and its evolution with redshift. We have found that the clustering predicted by the semi-analytic models agrees well with the HIPASS measurements of Meyer et al. (2007). We have calculated effective volume for redshift surveys planned with Square Kilometre Array (SKA) and compared with that of the optical Euclid mission. Finally, we study the clustering of faint extragalactic sources which are one of the foregrounds in PLANCK maps. We predict the clustering of faint extragalactic sources using a hybrid GRASIL+GALFORM+N-body model. We have compared the hybrid scheme with analytic clustering estimates. On large scales the two approaches agree, but for multipoles with $l > 500$ the results differ significantly, with the hybrid approach being the more accurate.

Declaration

The work in this thesis is based on research carried out between 2007 and 2010 at the Institute for Computational Cosmology, Department of Physics, University of Durham. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text. Portions of this thesis have been published or are to be published in the following papers:

Kim H-.S., Baugh C. M., Cole S., Frenk C. S., 2009, MNRAS, 400, 1527

Kim H-.S., Baugh C. M., Benson A. J., Cole S., Frenk C. S., Lacey C. G., Power C., Schneider M., 2010, submitted to MNRAS, arXiv1003.0008
⋆Section. 4.4.3 mainly worked by M. Schneider as a co-author of this paper.

Kim H-.S., Lacey C. G., Cole S., Baugh C. M, Frenk C. S., Efstathiou G., 2010, to be submitted to MNRAS

Acknowledgements

I am sincerely and heartily grateful to my supervisors, Prof. Carlton Baugh, Prof. Carlos Frenk and Prof. Shaun Cole, for the support and guidance. They showed me throughout my dissertation writing. I am not sure it would have not been possible without their help. Besides I would like to thank to Dr. Cedric Lacey and Dr. Chris Power.

I cannot forget a help and support from Dr. J.H. Lee and Dr. Jay Cho. I would like to thank a few friends who shared my worry and happiness, Jaewoo Kim, Y.C. Lee, Y.J. Yun and J.Y. Lee. Also, I would like to thank the members of Durham Korean Church who gave a lot of useful information for living in UK to me, especially J.H. Lee and K.Y. Park.

I am profoundly grateful to my parent and brothers for their moral and material support. I would also like to thank my father-in-law and mother-in-law from the bottom of my heart for giving me a lot of support. I thank to Korean Government for financial support during my PhD course.

I thank Richard Bower and Filipe Abdalla for insightful comments about my thesis as examiners.

Finally, I really would like to say to my family, Keunhae, Chaeeun and Dangyeol(?),
"I love you and I always thank that you are in my life."

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Accurate measurements of the acoustic peaks imprinted on the cosmic microwave background (CMB) power spectrum by COBE and WMAP (Smoot et al. 2002; Spergel et al. 2006) have shown that the Universe was close to homogeneous on large scales and isotropic at the level of one part in 10^5 when it was 380,000 years old. The CMB observations give tight constraints on the values of fundamental cosmological parameters. However, degeneracies exist between some fundamental parameters which cannot be decided by CMB data alone.

To break the degeneracies, we need to combine these CMB data with other observations, such as type Ia supernovae (Riess et al. 1998, 2004; Perlmutter et al. 1999) and the large scale structure of the Universe as traced by galaxies from the 2dF Galaxy Redshift Survey (2dFGRS, Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS, York et al. 2000). Furthermore, a pattern of oscillatory features called the baryonic acoustic oscillations (BAO) imprinted on the matter power spectrum can be used to constrain the dark energy equation of state (Percival et al. 2001; Cole et al. 2005; Einsenstein et al. 2005).

Currently a simple cold dark matter model with a cosmological constant (ΛCDM) works well to describe the observed Universe. In the ΛCDM framework, the Universe started from the "Big Bang". The Universe expanded exponentially through the inflationary epoch, which produced a scale invariant perturbations which are the seeds of the structure in today's Universe. These seeds were gradually amplified by gravitational instability as the Universe evolved. The Universe has a decelerating

expansion after the inflationary epoch before entering into the phase of accelerating expansion where the cosmological constant Λ dominates the dynamics.

Galaxy surveys have increased in size by order of magnitude since the mid 1990s. The accurate measurement of the galaxy distribution on large scales by the 2dFGRS and SDSS constrains cosmological parameters and also gives us a chance to understand galaxy formation model. The surveys are large enough to measure clustering in sub-samples, split by intrinsic galaxy properties such as colour, morphological type and luminosity. Also many multiwavelength surveys are underway and scheduled to find further information about the cosmological parameters and galaxy evolution and formation; these include SCUBA2 (Scott & White 1999), ASKAP(Johnston et al. 2008), SKA (Albrecht et al. 2006) and Planck (Planck bluebook 2005).

1.1 The observed clustering of galaxies

The clustering of galaxies encodes important information about the values of the cosmological parameters, in so for as it can be related to the spatial distribution of the underlying dark matter, and also about the physical processes behind galaxy formation. In the cold dark matter (CDM) hierarchical structure formation theory, the evolution of galaxies takes place inside dark matter haloes (White & Rees 1978). The formation and evolution of CDM is governed by gravity and can be modelled accurately using N-body simulations. We do not yet have the same level of knowledge of the fate of the baryons, which depends on the physics of gas accretion, star formation and feedback processes. Recent improvements in astronomical instrumentation have led to a wealth of new information becoming available on galaxy clustering, both locally and at earlier epochs. In particular, the unprecedented size and number of galaxies in the 2dF Galaxy Redshift Survey (2dFGRS; Colless et al. 2001) and the SDSS (York et al. 2000) make it possible to quantify how the clustering signal depends on intrinsic galaxy properties, such as luminosity, stellar mass or colour. The variation of clustering strength with an intrinsic galaxy property encodes important information about how galaxies populate haloes. A discrepancy between observations and theoretical predictions points to the need to revise the model to

perhaps incorporate new processes. In this thesis, we consider three probes of galaxy clustering, made at different redshift.

The luminosity dependence of clustering has been measured by 2dFGRS and SDSS (Norberg et al. 2001, 2002; Zehavi et al. 2002, 2005, 2010). These results imply that galaxies with a particular luminosity resides in dark matter haloes of a certain mass. Therefore, such trends in the clustering signed provide a constraint on models of galaxy formation and evolution. In Chapter 3, we show how to interpret the observed luminosity dependence of clustering using the GALFORM sem-analytic model framework.

Cold gas is central to galaxy formation yet little is known about the cold gas content of galaxies particularly in comparison with observations at optical wavelengths. The primary probe of atomic hydrogen, damped Lyman-α and 21cm line emission, is incredibly weak. It is only in recent years that a robust and comprehensive census of atomic hydrogen (HI) in the local universe has been made possible through the HI Parkes All Sky Survey (HIPASS; Barnes et al. 2001; Zwaan et al. 2003, 2005). So far, the distribution of cold gas in galaxies has been observed only in the local Universe. The new generation of 21cm line surveys, such as ASKAP (Johnston et al. 2008), SKA (Albrecht et al. 2006), and Murchison Widefield Array (MWA, Moralez et al. 2004), will produce a remarkable breakthrough in cosmology and galaxy formation and evolution. We present predictions for the distribution of HI in Chapter 4.

The clustering of faint extragalactic sources is one of the important foregrounds expected in the cosmic microwave background measurements from Planck. Planck will constrain the cosmological parameters much more accurately than WMAP or COBE due to the improved angular resolution and broader spectral coverage. To extract robust information from the Planck CMB maps, we need an accurate model of the foreground contamination due to galaxies which are fainter than the Planck detection limit. In Chapter 5, we model the foreground by faint extragalactic sources using state of art GRASIL+GALFORM+N-body model.

We now briefly review our knowledge of galaxy clustering in these three areas.

1.1.1 The clustering of optically selected galaxies

The first attempt to quantify the difference between the clustering of early- and late-type galaxies was made using a shallow angular survey, the Uppsala catalogue, with morphological types assigned from a visual examination of photographic plates (Davis & Geller 1976). Elliptical galaxies were found to have a higher-amplitude angular correlation function than spiral galaxies. In addition, the slope of the correlation function of ellipticals was found to be steeper than that of spiral galaxies at small angular separations. More recently, the comparison of clustering for different types has been extended to three dimensions using redshift surveys. Again, similar conclusions have been reached in these studies, namely that ellipticals display stronger clustering than spirals (Lahav & Saslaw 1992; Santiago & Strauss 1992; Iovino et al. 1993; Hermit et al. 1996; Loveday et al. 1995; Guzzo et al. 1997; Willmer et al. 1998).

Over the past decade, a variety of clustering studies in the local Universe have established an increasingly refined and quantitative characterization of the dependence of galaxy clustering on luminosity, morphology, colour and spectral type (e.g., Brown, Webster & Boyle 2000; Norberg et al. 2001, 2002; Zehavi et al. 2002; Budavari et al. 2003; Madgwick et al. 2003; Zehavi et al. 2005; Li et al. 2006; Swanson et al. 2008; Loh et al. 2010; Zehavi et al. 2010). The general trends of the clustering for different intrinsic properties are : (i) Luminous galaxies generally cluster more strongly than faint galaxies which shows their tendency to reside in higher mass dark matter haloes or denser environments. (ii) Galaxies with bulge-dominated morphologies, red colors, or spectral types indicating old stellar populations also exhibit stronger clustering and a preference for dense environments. Also, at intermediate and high redshifts, significant progress has been made in recent years in measuring galaxy clustering(e.g., Brown et al. 2003; Daddi et al. 2003; Adelberger et al. 2005; Lee et al. 2006; Phleps et al. 2006; Coil et al. 2006, 2008; Meneux et al. 2008, 2009; Abbas et al. 2010).

1.1.2 The clustering of the HI selected galaxies

We do not yet have a clear picture of how much cold gas there is in the Universe at different epochs and how this gas is distributed between dark matter haloes of different mass. The neutral hydrogen content of galaxies has been probed at high redshifts (z $>$ 2) using the absorption of the Lyman-α line by gas clouds along the line of sight to distant quasars (e.g. Lanzetta et al. 1991; Wolfe et al. 1995; Storrie-Lombardi, Irwin & Wolfe 1996; Peroux et al. 2005; Wolfe et al. 2005). A complementary probe of the atomic hydrogen content of galaxies and the physical state of the gas is the 21cm line. A blind survey of 21cm line absorption of gas illuminated by background radio sources has been proposed as an unbiased probe of damped Lyman-α clouds, which would extend to objects with high dust content, unlike Lyman-α absoprtion (Kanekar & Briggs 2004; Kanekar et al. 2009).

To date there is only one clustering measurement using HI observations estimated by Meyer et al. (2007), which was made using the HI Parkes All-Sky Survey (HIPASS) Catalogue (HICAT; Meyer et al. 2004). The clustering strength of HI galaxies is weaker than that of optical galaxies.

1.1.3 The clustering of faint, dusty galaxies

The discovery of the cosmic infrared background came from the analysis of observations by the Infrared Astronomical Satellite (IRAS) and COBE (Schlegel et al. 1998; Chen et al. 1999). The first surveys of extragalactic sources using SCUBA and MAMBO implied that a population of very luminous galaxies at high redshift made a significant contribution to the energy generated by all galaxies over the history of the Universe (Blain et al. 1999). Several hundreds of these galaxies are now known (Smail et al., 1997; Barger et al., 1998; Hughes et al., 1998; Barger et al., 1999; Eales et al., 1999, 2000; Lilly et al., 1999; Bertoldi et al., 2000; Borys et al., 2002; Chapman et al., 2002a; Cowie et al., 2002; Dannerbauer et al., 2002; Fox et al., 2002; Scott et al., 2002; Smail et al., 2002; Webb et al., 2002). The 850 μm sources have a median redshift of z$\sim$2 (Chamman et al. 2005). The results of these mm/submm extragalactic galaxy surveys provide complementary information

to deep surveys for galaxies made in the radio (Richards, 2000), far-IR (Puget et al., 1999), mid-IR (Elbaz et al., 1999) and optical (Steidel et al., 1999) wavebands. Submm observations are a vital component of the search for a coherent picture of the formation and evolution of galaxies.

Deep surveys at submm-wavelengths have to date provided relatively little information about the spatial distribution of the detected galaxies. There is some indication from the UK 8 mJy SCUBA survey (Almaini et al., 2002; Fox et al., 2002; Ivison et al., 2002; Scott et al., 2002) and from the widest-field MAMBO surveys (Carilli et al., 2001) that the clustering strength of SCUBA-selected galaxies is greater than that of faint optically-selected galaxies, yet less than that of K $\sim$ 20 ERO samples (Daddi et al., 2000). Webb et al. (2002) point out that the angular clustering signal expected in submm surveys is likely to be suppressed by smearing in redshift, as the submm galaxies should have a wide range of redshifts, and so the spatial correlation function might in fact be stronger than that of the EROs. Studies of the spatial correlations of sub-mm galaxies have typically been limited to at most 100 sources, leading either to a limit on the clustering amplitude (Blain et al. 2004) or a marginal detection of clustering in projection (Scott et al. 2006).

1.2 Quantifying the large-scale distribution of galaxies

We use the correlation function and halo occupation distribution formalism to understand the galaxy distribution in the Universe.

The two point correlation function of galaxy clustering is a popular statistic with which to test galaxy formation models, because it depends on the way that galaxies populate dark matter haloes and is fairly easy to measure.

A theoretical view of galaxy clustering can be obtained from the Halo Occupation Distribution (HOD) which is a component of the halo model of galaxy clustering. The halo model breaks the large scale structure of the universe up into clumps of dark matter, with the HOD providing the distribution of galaxies within each of the dark matter clumps. The HOD is used to describe three connected properties of

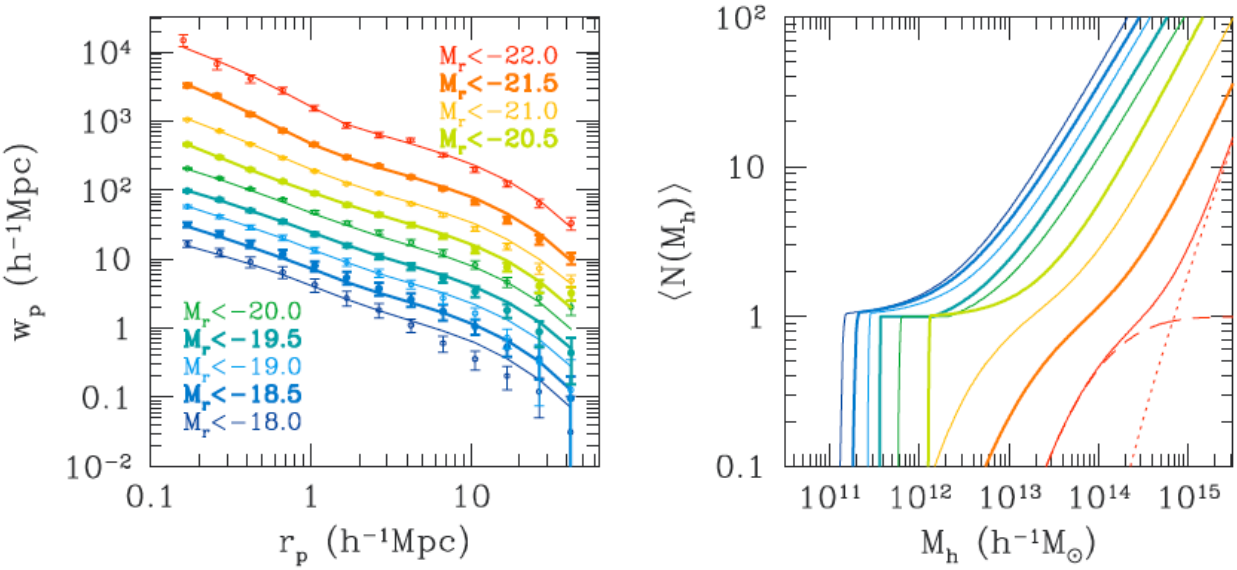

Figure 1.1: Luminosity dependence of galaxy clustering measured from SDSS and the associated HODs. The left panel shows the measured $w_p(r_p)$ and the right panel shows the best-fit HOD models for all luminosity-threshold samples. Taken from Zehavi et al. (2010).

the halo model: the probability distribution relating the mass of a dark matter halo to the number of galaxies that form within that halo; the distribution in space of galactic matter within a dark matter halo; the distribution of velocities of galactic matter relative to dark matter within a dark matter halo.

Fig. 1.1 shows examples of these two methods to quantify and understand the large-scale distribution of galaxies.

1.2.1 Clustering of galaxies : auto-correlation function

We start from a smooth matter density field $\rho(\mathbf{r})$, we can define a overdensity

$$\delta(\mathbf{r}) = \frac{\rho(\mathbf{r}) - \bar{\rho}}{\bar{\rho}}, \tag{1.1}$$

which satisfies $\langle\delta(\mathbf{r})\rangle$=0. Here we introduce the Fourier convention:

$$\hat{\delta}(\mathbf{k}) = \frac{1}{V}\int d^3r\ \exp[i\mathbf{k}\cdot\mathbf{r}]\delta(\mathbf{r}),$$

$$\delta(\mathbf{r}) = \frac{V}{(2\pi)^3}\int d^3k\ \exp[-i\mathbf{k}\cdot\mathbf{r}]\hat{\delta}(\mathbf{k}), \tag{1.2}$$

where V is a finite box of volume with periodic boundary condition. The two-point autocorrelation function is defined as

$$\xi(\mathbf{r}) = \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) \rangle. \tag{1.3}$$

Using Eq. (1.2) and the Hermitian property of Fourier transforms of real fields $\hat{\delta}(-\mathbf{k}) = \hat{\delta}^{\dagger}(\mathbf{k})$, we can derive

$$\xi(\mathbf{r}) = \frac{V}{(2\pi)^3} \int d^3k |\hat{\delta}(\mathbf{k})|^2 \exp[-i\mathbf{k}\cdot\mathbf{r}]. \tag{1.4}$$

We also see that the correlation function is the Fourier transform of the power spectrum defined as

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle = \delta^D(\mathbf{k}-\mathbf{k}')P(k), \tag{1.5}$$

where $\delta^D(\mathbf{k})$ is the Dirac-Delta distribution. Therefore the power spectrum is related to the correlation as

$$P(k) = \frac{1}{V} \int d^3r \xi(r) \exp[i\mathbf{k}\cdot\mathbf{r}] = \frac{4\pi}{V} \int dr \xi(r) r^2 j_0(kr), \tag{1.6}$$

where $j_0(kr)$ is the spherical Bessel function of order 0, $j_0(kr) = sin(kr)/kr$.

An alternative definition of the correlation function is in terms of the excess probability, compared to a random distribution, of finding another galaxy at a distance r from a randomly chosen object,

$$\mathrm{d}P = \bar{n}^2[1+\xi(r)]\delta V_1 \delta V_2, \tag{1.7}$$

where $\bar{n}$ is the mean number density of the galaxy sample and $\delta V_1, \delta V_2$ are volume elements. In the case of a galaxy catalogue from a semi-analytic model combined with an N-body simulation, since the computational volume is periodic, we can measure the correlation function using

$$1 + \xi(r) = D_s D_s / \bar{n}^2 V \mathrm{d}V, \tag{1.8}$$

where $D_s D_s$ is the number pairs of galaxies with separations in the range r to $r+\delta r$ in the catalogues from semi-analytic models, $\bar{n}$ is mean density of galaxies, V is volume of simulation box, and $\mathrm{d}V$ is the differential volume.

In practice, to estimate correlation function in a real survey with boundaries, the expected numbers of pairs is usually estimated by creating a random catalogue

which is much larger in number than the sample under study, and by counting pairs both within the two catalogues and between them. A variety of characterized estimators for $1+\xi$ (Hamilton 1993; Landy & Szalay 1993),

$$\begin{aligned} 1+\xi_1 &= < \mathrm{DD} > / < \mathrm{RR} > \\ 1+\xi_2 &= < \mathrm{DD} > / < \mathrm{DR} > \\ 1+\xi_3 &= < \mathrm{DD} >< \mathrm{RR} > / < \mathrm{DR} >^2 \\ 1+\xi_4 &= 1 + < (\mathrm{D}-\mathrm{R})^2 > / < \mathrm{RR} >, \end{aligned} \tag{1.9}$$

where DD, DR, and RR are pairs of galaxies in the data-data, data-random, and random-random catalogues. The above estimations of the correlation function are for the 3-dimensional case and the analysis of real observational data. Basically, all estimators are equivalent for a large volume. However the quadratic estimator, ξ_4, is more robust for a small volume because the expected pair count might be sensitive to the sample boundary.

Observational data has redshift-space distortions caused by peculiar galaxy velocities along the line of sight which distort the form of $\xi(r)$. To isolate this distortion, Hawkins et al. (2003) measured ξ in two dimensions $\xi(\sigma,\pi)$, both perpendicular(σ) and along the line of sight(π). Integration over the π direction for $\xi(\sigma,\pi)$ leads to a projected clustering statistic, $\Xi(\sigma)$, which is independent of redshift-space distortions,

$$\Xi(\sigma) = 2\int_0^\infty \mathrm{d}\pi \xi(\sigma,\pi) = 2\int_0^\infty \mathrm{dy}\xi\left(\sigma^2+\mathrm{y}^2\right)^{1/2}, \tag{1.10}$$

where y is the separation along the line of sight. Assuming a power law correlation function, $\xi(r) = (r/r_0)^{-\gamma}$, then we can easily obtain r_0 and γ from the projected correlation function, $\Xi(\sigma)$, using the analytic solution of Eq. (1.10):

$$\frac{\Xi(\sigma)}{\sigma} = \left(\frac{r_0}{\sigma}\right)^\gamma \frac{\Gamma(1/2)\Gamma[(\gamma-1)/2]}{\Gamma(\gamma/2)} = A(\gamma)\left(\frac{r_0}{\sigma}\right)^\gamma, \tag{1.11}$$

where Γ is the incomplete gamma function.

1.2.2 Halo Occupation Distribution

In the CDM model, galaxies form inside virialized clumps of dark matter called haloes, the formation and evolution which haloes are governed by gravity. Galaxy

clustering can be predicted by adopting a halo occupation distribution (HOD), which describes the statistical relation between galaxies and dark matter haloes.

In the halo model (Seljak 2000; Peacock & Smith 2000; Sheth et al. 2001; Berlind & Weinberg 2002) the two-point correlation function of the galaxies is a combination of two terms : the 1-halo (1h) term is caused by galaxy pairs inside a single halo, small separation, and the 2-halo (2h) term due to galaxy pairs in individual haloes, large separation (see detail Kravtsov et al. 2004). :

$$\xi_{gg}(r) = \xi_{gg}^{1h}(r) + \xi_{gg}^{2h}(r) + 1. \tag{1.12}$$

The two terms of Eq. (1.12) are given by

$$1 + \xi_{gg}^{1h}(r) = \frac{1}{2}\bar{n}_g^{-2} \int n(M)\langle N(N-1)\rangle_M \lambda(r|M) dM, \tag{1.13}$$

$$\xi_{gg}^{2h}(r) = \xi_{mm}^{lin}(r)\bar{n}_g^{-2} \int n(M_1) b_h(M_1)\langle N\rangle_{M_1} dM_1 \int n(M_2) b_h(M_2)\langle N\rangle_{M_2} \lambda(r|M_1, M_2) dM_2, \tag{1.14}$$

where $\bar{n}_g$ is the mean number density of galaxies in the sample, $n(M)$ is the halo mass function described by Press & Schechter

$$\frac{M}{\bar{\rho}} n(M) dM = f(\nu) d\nu = \frac{1}{\nu}\sqrt{\frac{\nu}{2\pi}} \exp\left[-\frac{\nu}{2}\right], \tag{1.15}$$

where ν is the peak height which is related with the critical density thresholds, δ_c, and the root mean square overdensity within spheres of radius R, $\sigma(M)$; $\nu = [\frac{\delta_c}{\sigma(M)}]$. The Press & Schechter mass function has been improved by Sheth & Tormen 1999 to give a better fit to the N-body simulation. $b_h(M)$ is the large-scale linear bias of halos described as

$$b_h(M) = \left(1 + \frac{\nu - 1}{\delta_c(z)} + \frac{2p}{\delta_c(z)} \frac{1}{1 + (q\nu)^p}\right), \tag{1.16}$$

where parameters have a value q=0.707 and p=0.3 for Sheth & Tormen 1999, q=1 and p=0 for Press & Schechter formula. $\lambda(r|M)$ is the convolution of the radial profile of galaxies within halos with itself introduced by Navarro , Frenk & White (NFW), $\lambda(r|M_1, M_2)$ is the convolution of two different radial profiles, and $\xi_{mm}^{lin}(r)$ is the linear dark matter correlation function (see Sheth et al. 2001; Berlind & Weinberg 2002).

The HOD characterizes the probability distribution $P(N|M)$ of finding a mean number, $< N >_M$, of galaxies in haloes of mass M_H. Combined with the spatial distribution of haloes and assumptions about the position and velocity of the galaxies with the halo, the HOD allows us to make a prediction of galaxy clustering. Although, $< N >_M$ generally refers to all galaxies, we usually separate the contribution from central and satellite galaxies. The formula for central galaxies, $< N >_{cen}$ is simplified to a step function proposed by Kravtsov et al. (2004) and a power-law $< N >_{sat}$ for satellites as:

$$< N >_{\rm cen} = \begin{cases} 1 & \text{if } M_H \geq M_{min} \\ 0 & \text{else} \end{cases}$$

$$< N >_{\rm sat} = \left(\frac{M_H}{M_{min}}\right)^{\alpha}. \tag{1.17}$$

Starting from this simple formula to describe the form of the HOD, we can produce a more accurate description of the distribution of galaxies between dark matter haloes by increasing the number of parameters for satellite and central galaxies or by modifying the form of the function for central galaxies. Zheng et al. (2007) adopt the following parameterization for the HOD of central galaxies :

$$< N_{\rm cen}(M) >= \frac{1}{2}\left[1 + {\rm erf}\left(\frac{{\rm log}M - {\rm log}M_{\rm min}}{\sigma_{{\rm log}M}})\right)\right], \tag{1.18}$$

where erf is the error function, ${\rm erf(x)} = \frac{2}{\sqrt{\pi}}\int_0^{\rm x} {\rm e}^{-{\rm t}^2}{\rm dt}$, $M_{\rm min}$ is the characteristic minimum mass of haloes which can host central galaxies above the luminosity threshold, and $\sigma_{\rm logM}$ is the width of the cutoff profile. Retaining the same cutoff profile used for the central galaxies, these authors assumed the HOD for satellite galaxies for haloes with $M > M_0$ could be written as:

$$< N_{\rm sat}(M) >= \frac{1}{2}\left[1 + {\rm erf}\left(\frac{logM - logM_{min}}{\sigma_{logM}}\right)\right] \times \left(\frac{M - M_0}{{M_1}'}\right)^{\alpha}, \tag{1.19}$$

where M_0 is the mass scale of the drop, ${M_1}'$ characterizes the amplitude, and α is the asymptotic slope at high halo mass. The satellite galaxy HOD is approximately a power law at the high mass end with a slope close to unity. In contrast to this, at

lower mass, it drops more steeply than a power law with a shape characterized by the last term in Eq. (1.19).

1.3 Motivation of thesis

Semi-analytic models predict the luminosity and color dependence of the galaxy correlation function (e.g., Kang et al. 2005; Croton et al. 2006; De Lucia et al. 2007; Bower et al. 2006). HOD modeling has been applied to interpret clustering data from a number of surveys at low and high redshift (e.g.,van den Bosch et al. 2003; Magliocchetti & Porciani 2003; Yan, Madgwick & White 2003; Zheng 2004; Yang et al. 2005; Zehavi et al. 2005; Cooray 2006; Blake, Collister & Lahav 2008; Brown et al. 2008; Quadri et al. 2008; Wake et al. 2008; Kim et al. 2009; Zheng et al. 2009; Ross, Percival & Brunner 2010). In Chapter 3, we show that semi-analytic models cannot reproduce the observed luminosity dependence clustering. This discrepancy between the model predictions and observational measurements indicates that there are missing physical processes in the semi-analytic galaxy formation model.

Now, we have new generation of HI 21cm line survey, such as ASKAP, SKA, and MeerKat. A number of possible models of the abundance and clustering of HI sources have been tried. Empirical models rely on observations of HI in the Universe (Abdalla & Rawlings 2005; Abdalla, Blake & Rawlings 2010). The fully numerical approach uses cosmological gas dynamic simulations (Popping et al. 2009, Duffy et al. 2008). The halo occupation distribution formalism based on the observation results of HI clustering (Wyithe et al. 2009). In Chapter 4, we use the GALFORM semi-analytic model to predict the abundance and clustering of HI sources. The evolution of cold gas mass function was predicted by Power et al. (2010). The main issues to be tackled by future HI 21cm surveys are how the HI galaxies are distributed at high redshift and how well these surveys can constrain the cosmological parameters.

The contribution of clustering by the extragalactic sources is important to calibrate the Planck maps to get robust constraints for cosmological parameters. Predictions for the high redshift sources have made using the GALFORM semi-analytic

model (Baugh et al. 2005; Le Delliou et al. 2005, 2006; Orsi et al. 2008; Lacey et al. 2008, 2010) with the model able to reproduce the luminosity function of submm galaxies, the evolution of the population of Lyman-α emitting galaxies, and their clustering over the redshift range z $\sim$ 3 - 6. The contribution of undetected extra-galactic sources to the angular clustering to be measured by Planck is predicted in Chapter 5.

Chapter 2

The semi-analytic galaxy formation simulation : GALFORM

In the cold dark matter hierarchical structure formation theory, galaxies grow inside dark matter haloes (White & Rees 1978). The formation of structure in the dark matter is governed by gravity and is well understood. However, the fate of the baryonic component is much more complicated. In this thesis we try to constrain the physics of galaxy formation using galaxy clustering using a theoretical approach called semi-analytic modelling. Semi-analytic models are an alternative to hydrodynamic simulation that allow a rapid exploration of the parameter space of galaxy formation physics, generating large, statistically useful samples of galaxies with a wide range of predicted properties. So, this approach offers the best way of modelling the large-scale distribution of galaxies in the Universe.

The number of galaxies in any given halo which should be closely related to its merger history which can be extracted from an N-body simulation or generated using a Monte-Carlo method. Once a merger tree has been created, a suite of analytic prescriptions is used to model the formation and evolution of galaxies in each halo, starting with the highest redshift progenitors and moving forward in time all the way to the final halo at z = 0.

The GALFORM semi-analytic galaxy formation model is a synthesis of many

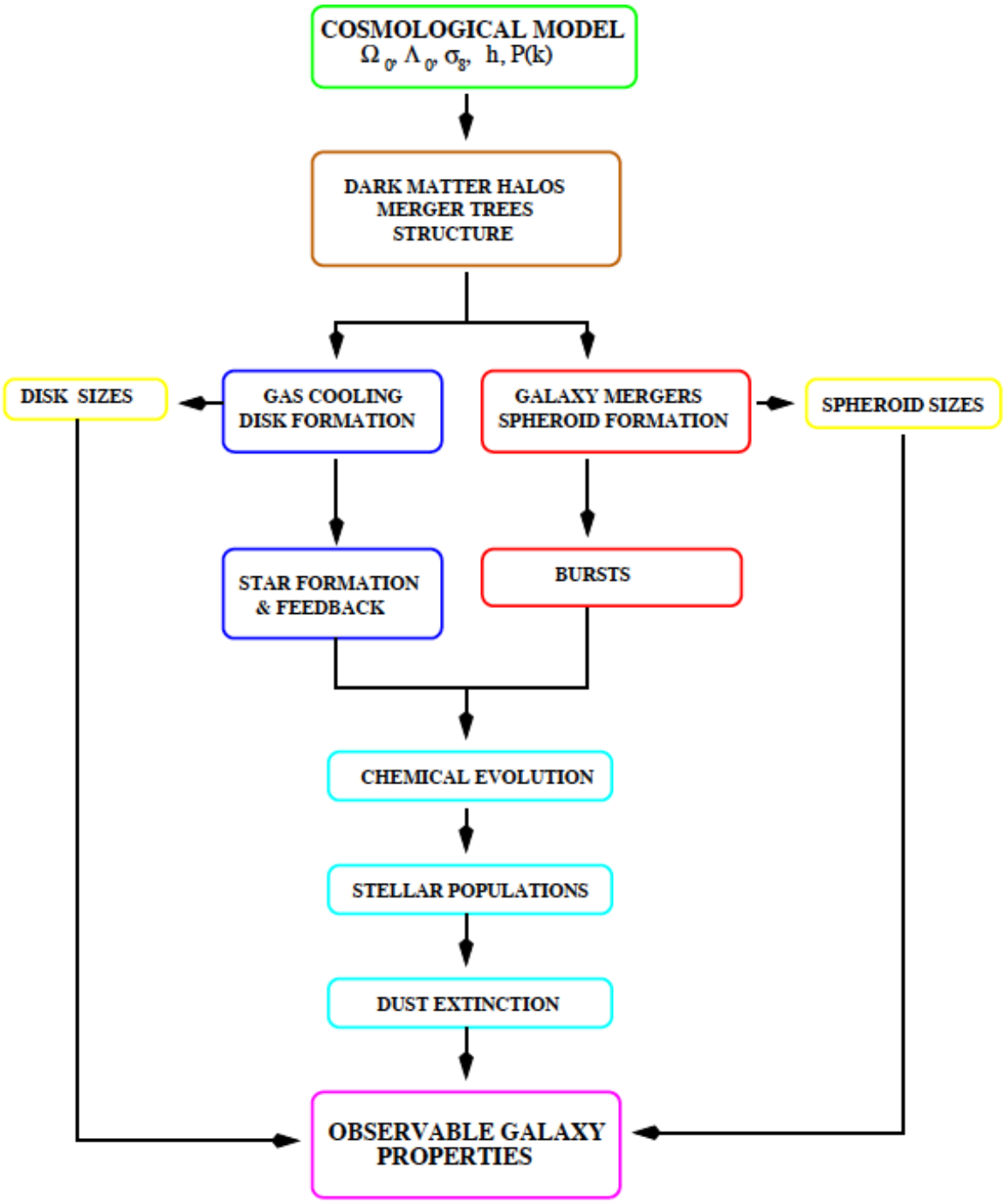

Figure 2.1: A schematic showing the components of the GALFORM semi-analytic galaxy formation model processing. This shows how different physical processes are combined to make predictions for the observable properties of galaxies. Adapted from Baugh (2006).

physical processes, each of which has been developed to understand a particular aspect of the complicated nonlinear processes of galaxy formation. The main properties and processes that we model within the framework include : (1) The gravitationally driven assembly of dark matter haloes. (2) The density and angular momentum profiles of dark matter and hot gas. (3) The radiative cooling of gas and its collapse to form centrifugally supported disks. (4) Star formation in disks. (5) Feedback processes, resulting from the injection of energy from supernovae (SNe) and AGN heating. (6) Chemical enrichment of the interstellar medium (ISM) and hot halo gas which govern the gas cooling rate and the properties of the stellar populations in a galaxy. (7) The dynamical friction of galaxy's orbit within a dark matter halo and its possible merger with the central galaxy. (8) The formation of galactic spheroids. (9) The spectrophotometric evolution of stellar populations. (10) The effect of dust extinction on galaxy luminosities and colours, and its dependence on the inclination of a galaxy. (11) The generation of emission lines from interstellar gas ionized by young hot stars. In this section we briefly describe the main physical processes included in the GALFORM model which are discussed in this thesis. A comprehensive overview of the GALFORM model can be found in Cole et al. (2000); see also the review by Baugh (2006).

Figure. 2.1 shows the basic scheme of the GALFORM semi-analytic model. GALFORM combines different physical processes to make predictions for the observational properties of galaxies, starting from initial conditions specified by the background cosmology.

2.1 Formation of Dark matter haloes

2.1.1 Cosmology

We compute the formation and evolution of galaxies using GALFORM within the framework of the ΛCDM model of structure formation. The growth of dark matter haloes is governed by the background cosmology. The cosmologies in the models used in this thesis are not identical. The Bower et al. (2006), De Lucia et al. (2007), Font et al. (2008), and MHIBow06 models all have the Millennium cosmology based

on the WMAP1 results (Spergel et al. 2003). The background cosmology of the GpcBow06 model is based on the Sanchez et al. (2009) results. The Baugh et al. (2005) model uses a spatially flat ΛCDM model with 'concordance' parameters. The different cosmological parameters for the background cosmologies are summarized in Table. 2.1.

Table 2.1: The cosmological parameters used in the galaxy formation models used in this thesis. The parameters govern the evolution of structure in the dark matter. The columns are as follows: (1) The name of the model. (2) The present-day matter density Ω_0. (3) The cosmological costant Λ_0. (4) The Hubble constant H_0. (5) The primordial scalar spectral index n_s. (6) The baryon density Ω_b. (7) The fluctuation amplitude σ_8. (8) The source of the dark matter halo merger trees.

	Ω_0	Λ_0	$H_0[\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}]$	n_s	Ω_b	σ_8	merger tree source
Bower et al. (2006) Font et al. (2006) MHIBow06	0.25	0.75	73	1	0.045	0.9	Millennium Simulation
GpcBow06	0.26	0.74	71.5	0.96	0.044	0.8	GigaParsec simulation (GPICC)
Baugh et al. (2005)	0.25	0.75	73	1	0.045	0.9	Monte Carlo (Parkinson et al. 2008)

2.1.2 Dark matter halo merger trees

The haloes could be associated with peaks in the Gaussian random density field of dark matter in the early Universe (Press & Schechter 1974). Press & Shechter have derived the distribution of dark matter halo masses using the relatively simple statistics of Gaussian random field. The number of haloes per unit volume in the mass range M to $M + \delta M$ is $\delta M(\mathrm{d}n/\mathrm{d}M)$ where :

$$\frac{\mathrm{d}n}{\mathrm{d}M}(M,t) = \left(\frac{2}{\pi}\right)^{1/2} \frac{\rho_o}{M^2}\frac{\delta_c(t)}{\sigma(M)}\left|\frac{\mathrm{dln}\sigma}{\mathrm{dln}M}\right| \exp\left[-\frac{\delta_c^2(t)}{2\sigma^2(M)}\right], \tag{2.1}$$

where ρ_0 is the mean number density of the Universe, $\sigma(M)$ is the fractional variance in the density field that contains a mass M, and $\delta_c(t)$ is the critical overdensity for spherical top-hat collapse at time t.

The assembly and merger histories of dark matter haloes are computed using a Monte Carlo method based on the extended Press-Schechter theory (Bower et al. 1990; Bond et al. 1991; Cole et al. 2000) or using halo merger histories directly extracted from N-body simulations (Helly et al. 2003). The dark matter halo merger histories from the revised Monte Carlo approach of Parkinson et al. (2008) agree well with those from N-body simulations.

In the case of the Monte Carlo trees, the halo merger history is calculated using an algorithm which randomly generates a formation path for the haloes. Equation. 2.2 describes the progenitor distribution (equation (2.15) of Lacey & Cole 1993) and is derived from the extension of the Press & Shechter (1974) theory proposed by Bond et al. (1991) and Bower (1991) :

$$f_{12}(M_1, M_2)\mathrm{d}M_1 = \frac{1}{\sqrt{2\pi}}\frac{(\delta_{c1} - \delta_{c2})}{(\sigma_1^2 - \sigma_2^2)^{1.5}} exp\left(-\frac{(\delta_{c1} - \delta_{c2}^2)}{2(\sigma_1^2 - \sigma_2^2)}\right)\frac{\mathrm{d}\sigma_1^2}{\mathrm{d}M_1}\mathrm{d}M_1, \tag{2.2}$$

where the quantities σ_1 and σ_2 are the linear theory rms density fluctuations in spheres of mass M_1 and M_2. The δ_{c1} and δ_{c2} are the critical thresholds on the linear overdensity for collapse at time t_1 and t_2. The quantities that must be specified in order to define the merger tree are the density fluctuation power spectrum, which gives the function $\sigma(M)$, and the cosmological parameters, Ω_0 and Λ_0, which enter through the dependence of $\delta_c(t)$ on the cosmological model. With a zero time-lag this can be interpreted as a merging rate. Repeated application of this merging rate

use to build merger tree The GpcBow06 model in Chapter 4 and the Baugh et al. (2005) model in Chapter 5 use this method to create the merger histories of dark matter haloes.

The Bower et al (2006), the De Lucia et al. (2007), the Font et al. (2008) models used in Chapters 3 and 4, and the MHIBow06 model in Chapter 4 use dark matter halo merger trees extracted directly from the Millennium simulation (Springel et al. 2005).

The main advantage of N-body merger trees is that this method gives direct information about clustering because the N-body simulation gives the location of each dark matter halo along with information which can be used to give the spatial position of galaxies within the halo. The merit of the Monte Carlo method is that this technique allows us to grow the halo merger trees to lower mass dark matter haloes than can be resolved by the Millennium simulation.

2.1.3 Halo properties

We model the internal structure of dark matter haloes in order to calculate the properties of the galaxies which form within them.

Spin distribution

Dark matter haloes obtain angular momentum from the gravitational tidal torques which operated during their formation. The magnitude of the angular momentum is quantified by the dimensionless spin parameter :

$$\lambda_H = \frac{J_H |E_H|^2}{G M_H^{5/2}}, \tag{2.3}$$

where M_H, J_H, and E_H are the total mass, angular momentum, and energy of the halo, respectively. The distribution of λ_H has been measured in various N-body simulation studies (Efstathiou et al. 1988; Cole & Lacey 1996; Lemson & Kauffmann 1999).

Halo density profile

The standard choice of halo density profile in the GALFORM is the NFW model (Navarro et al. 1997) :

$$\rho_{\rm r} = \frac{\Delta_{\rm vir}\rho_{\rm crit}}{f(a_{NFW})}\frac{1}{r/r_{\rm vir}(r/r_{\rm vir}+a_{NFW})^2}(r \leq r_{\rm vir}), \tag{2.4}$$

where $f(a_{NFW}) = \ln(1+1/a_{NFW}) - 1/(1+a_{NFW})$. $\Delta_{\rm vir}$ is the virial over density which is defined by the spherical collapse model. $\rho_{\rm crit}$ is defined by $\rho_{\rm crit} = 3H^2/(8\pi G)$.

Halo rotation velocity

We model the rotational structure of the halo to compute the angular momentum of the halo gas that cools and is used in forming a galaxy. Approximately, the mean rotational velocity, $V_{\rm rot}$, can be related to the halo spin parameter, λ_H, in Eq. 2.3.

$$V_{rot} = A(a_{NFW})\lambda_H V_H, \tag{2.5}$$

where the effective circular velocity of the halo at the virial radius is defined by $V_H \equiv (GM/r_{\rm vir})^{1/2}$. The dimensionless coefficient $A(a_{NFW})$ is a weak function of a_{NFW}, varying from $A \approx 3.9$ for a_{NFW}=0.01 to $A \approx 4.5$ for a_{NFW}=0.3.

2.2 Formation of disks and spheroids

In this section we show how disks and spheroids in galaxies form and how we model the star formation, feedback and chemical evolution in the models used in this thesis.

2.2.1 Disk formation

Disks form by the radiative cooling of gas which is initially in the hot halo. Tidal torques distribute angular momentum to all material in the halo, including the gas, so that when the gas cools and loses its pressure support, it will naturally settle into a disk.

Hot gas distribution

Diffuse gas is assumed to be shock-heated during halo collapse and merging events. We refer to the halo gas as *hot gas* to distinguish it from the *cold gas* in galaxies. To calculate the amount of hot gas in the halo which falls into the disk by cooling, we model the initial temperature and density profiles of the hot gas in the halo. The mean temperature of the gas is related to the virial temperature, defined by :

$$T_{\rm vir} = \frac{1}{2}\frac{\mu m_H}{k} V_H^2, \tag{2.6}$$

where m_H is the mass of the hydrogen atom and μ is the mean molecular mass.

High-resolution hydrodynamic simulations of the formation of galaxy clusters (Navarro et al. 2005; Eke, Navarro & Frenk 1998; Frenk et al. 1999) guide the modelling of the hot gas density profile. We assume that any diffuse gas in the progenitors of a forming halo is shock-heated during the formation process and then settles into a spherical distribution with density profile,

$$\rho_{\rm gas}(r) \propto 1/(r^2 + r_{\rm core}^2), \tag{2.7}$$

where $r_{\rm core} = r_{NFW}/3$.

Cooling process

In the GALFORM model we assume that galactic disks form by the cooling of the diffuse gas in the halo.

Gas can cool via a number of channels. The cooling channels are: (i) inverse Compton scattering of CMB photons by electrons in the hot gas for the very early universe (Rees & Ostriker 1977). (ii) The excitation of rotational or vibrational energy levels in molecular hydrogen through collisions for the haloes with virial temperature below $T \sim 10^4$K. (iii) Emission of photons following transitions between energy levels for the haloes with virial temperature between 10^4K to 10^6K. (iv) Bremsstrahlung radiation as electrons are accelerated in an ionized plasma for the massive clusters ($T \sim 10^7$K).

The cooling time of the gas in the GALFORM is given by

$$\tau_{\rm cool}(r) = \frac{3}{2}\frac{\bar{\mu} m_p k_B T_{\rm gas}}{\rho_{\rm gas}(r)\Lambda(T_{\rm gas}, Z_{\rm gas})}, \tag{2.8}$$

where $T_{\rm gas}$ is the temperature of the gas and $Z_{\rm gas}$ is its metallicity. $\Lambda(T_{\rm gas}, Z_{\rm gas})$ is the cooling function tabulated by Sutherland & Dopita (1993). $\bar{\mu} m_p$ is the mean particle mass. The cooling radius, $r_{\rm cool}(t)$, is defined by the radius at which we have $\tau_{\rm cool} = t$ (age of the halo).

We take into account the free-fall time which determines when the newly cooled gas can be accreted by the disk.

$$t_{\rm ff}(r) = \int_0^r \left[\int_r^y -\frac{2GM(x)}{x^2} {\rm d}x \right]^{-1/2} {\rm d}y. \tag{2.9}$$

The free-fall radius, $r_{\rm ff}$ is defined by the radius at which $t_{\rm ff} = t$. To obtain the total gas mass that cools and is accreted onto the disk during a timestep, Δt, we calculate the mass of the hot gas inside the spherical shell defined by the radius, $r_{\rm min}(t) = \min[r_{\rm cool}(t), r_{\rm ff}(t)]$ and $r_{\rm min}(t + \Delta t) = \min[r_{\rm cool}(t + \Delta t), r_{\rm ff}(t + \Delta t)]$, and equate it to $\dot{M}_{\rm cool} \Delta t$. $\dot{M}_{\rm cool}$ is the cooling rate which is an important quantity for the star formation rate, metal enrichment and feedback processes.

In the Baugh et al. (2005) model in Chapter 5, the process of cooling of hot gas is "locked" during the lifetime of the halo. This means that the cooling of additional gas accreted through infalling haloes or gas that is returned to the halo by feedback processes is not allowed to cool until the next halo forms in the merger tree.

The Bower et al. (2006) model used in Chapters 3 and 4 has an improved calculation of cooling, by allowing the cooling of gas reheated by stellar feedback. The new calculation explicitly transfers reheated gas back to the reservoir of hot gas on a timescale comparable to the halo's dynamical timescale. Bower et al define a time scale for the reincorporation of reheated gas $\tau_{\rm reheat} = \tau_{\rm dyn}/\alpha_{\rm reheat}$ and increment the mass available for cooling over a timestep by $\Delta m = m_{\rm reheat} \Delta t / \tau_{\rm reheat}$, where $\tau_{\rm dyn}$ is the dynamical timescale of disk and $\alpha_{\rm reheat}$ is a parameter. Also, the cooling calculation in the MHIBow06 and GpcBow06 models used in this thesis is based on the prescription using by Bower et al. (2006) model.

The cooling gas is only accreted by the central galaxy in the halo and not by any satellite galaxies in the same halo. Recently, Font et al. (2008) considered an alternative model for cooling in satellite galaxies, by adopting rampressure stripping of the hot gas based on the hydrodynamic simulations of McCarthy et al. (2008).

The Font et al. (2008) model predicts the bimodal distribution of galaxy colours seen in SDSS observations and also reproduces the observed dependence of satellite colours on environment, from small groups to high mass clusters. Font et al. adopt a partially empirical approach to give the cooling rate in the satellite halo as :

$$\dot{M}_{\rm hot} = (1 - \epsilon_{\rm strip} f_{\rm strip}) \frac{M_{\rm reheat}}{\tau_{\rm reheat}} - \dot{M}_{\rm cool} \tag{2.10}$$

The effect of ram-pressure stripping is described by the term $\epsilon_{\rm strip} f_{\rm strip}$, where $f_{\rm strip}$ is the stripping factor and $\epsilon_{\rm strip}$ is a new parameter (representing the time averaged stripping rate after the initial pericentre of the satellite orbit) which is adjusted to fit the observations.

Angular momentum

We assume that when the halo gas cools and collapses into a disk, it conserves its angular momentum. With this assumption, GALFORM is able to match the observed distribution of disk galaxy scale lengths.

Validity of the cooling recipe in the semi-analytic models

There has been considerable discussion regarding the validity of the cooling recipe used in semi-analytical models (Bsirnboim & Dekel 2003; Keres et al. 2005). Keres et al. suggest that the cooling recipe outlined above requires significant revision in light of their simulation results, which confirm the expectations of Binney (1977). Bsirnboim & Dekel and Keres et al. characterize their results in terms of two cooling regimes. One is a cold mode, which is found to dominate in low mass haloes ($3{\times}10^{11} M_{\odot}$) and high redshift (z>3), in which gas is funnelled down filaments onto galaxies, while the other is a hot mode, which is the "traditional" hot accretion mode, in which gas cools from a quasi-static halo. Croton et al. (2006) and Benson & Bower (2010) gives a detail of this cooling recipe in semi-analytic models.

2.2.2 Star formation in disks

Star formation not only converts cold gas into luminous stars, but it also affects the physical state of the surrounding gas by feedback processes such as supernovae and

stellar winds. The processes of gas cooling from the reservoir of the hot gas in halo and accretion onto the disk, star formation from the cold gas, and the reheating and ejection of gas are assumed to occur simultaneously.

Chemical enrichment

We model the chemical enrichment of galaxies. The injected metals enrich both the cold star-forming gas and the surrounding diffuse hot halo gas. Enrichment of the halo gas decreases the cooling time while stellar enrichment affects the colour and luminosity of the stellar populations.

The equations used in GALFORM to describe the transformation of mass and metals between stars and the hot and cold gas phases are :

$$\dot{M}_{\star} = (1 - R)\psi \tag{2.11}$$

$$\dot{M}_{\rm hot} = -\dot{M}_{\rm cool} + \beta\psi \tag{2.12}$$

$$\dot{M}_{\rm cold} = \dot{M}_{\rm cool} - (1 - R - \beta)\psi \tag{2.13}$$

$$\dot{M}^{Z}_{\star} = (1 - R)Z_{\rm cold}\psi \tag{2.14}$$

$$\dot{M}^{Z}_{\rm hot} = -\dot{M}_{\rm cool}Z_{\rm hot} + (pe + \beta Z_{\rm cold})\psi \tag{2.15}$$

$$\dot{M}^{Z}_{\rm cold} = \dot{M}_{\rm cool}Z_{\rm hot} + (p(1 - e) - (1 + \beta - R)Z_{\rm cold})\psi, \tag{2.16}$$

where $M^{Z}_{\rm cold}$ is the mass of metal in the cold gas, $Z_{\rm cold} = M^{Z}_{\rm cold}/M_{\rm cold}$ is the metallicity of the cold gas, and $Z_{\rm hot} = M^{Z}_{\rm hot}/M_{\rm hot}$. The values of R, the fraction of mass recycled by stars, and p, the yield, in these equations are determined by the choice of the IMF. ψ is the instantaneous star formation rate, e is the fraction of newly produced metals ejected directly from the stellar disk to the hot gas phase (typically e=0 in our models assuming that all of the metals produced by supernovae feedback firstly settle into the cold gas phase.). β is the efficiency of stellar feedback (see later for definition).

Figure. 2.2 shows the concept which we use in the GALFORM to describe the transfer of mass and metals between stars and the hot and cold gas phases during a single timestep.

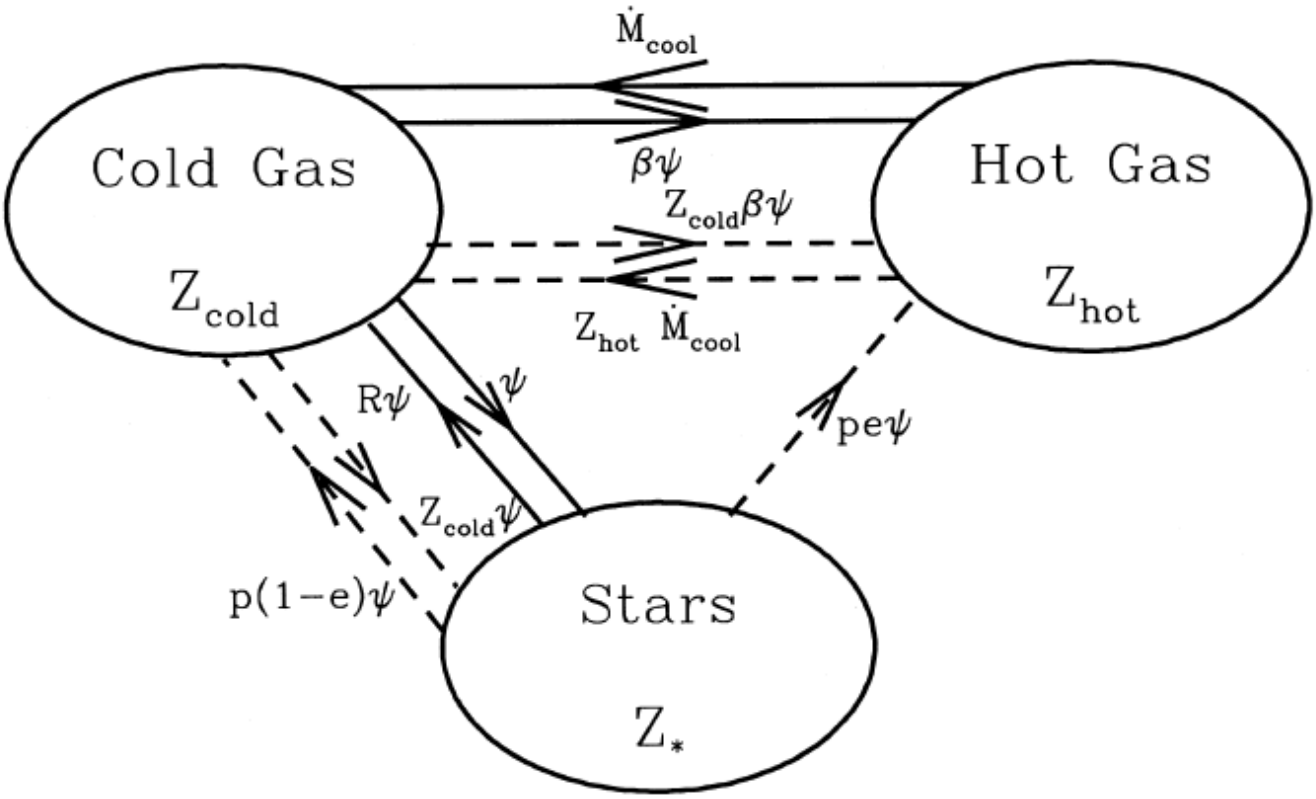

Figure 2.2: A schematic diagram showing the transfer of mass and metals between stars and the hot and cold gas phases during a single timestep, as described by Eqs. 2.11∼ 2.16. Adapted from Cole et al. (2000).

Star formation

Once the hot gas cools onto a rotationally supported gas disk, the process of star formation starts. In the GALFORM model, we assume that the instantaneous star formation rate, ψ, is :

$$\psi = \frac{M_{\rm cold}}{\tau_\star}, \tag{2.17}$$

where $\tau_\star$ is the star formation timescale which is related to the circular velocity of the galaxy disk as shown below. The Bower et al. (2006) and Font et al. (2008) models in Chapters 3 and 4 follow the same prescription for the star formation time scale as in Benson et al. (2003). The star formation timescale depends on the dynamical time of the disk, $\tau_{\rm disk}$:

$$\tau_\star = \varepsilon_\star^{-1} \tau_{\rm disk} \left(\frac{V_{\rm disk}}{200 {\rm kms}^{-1}} \right)^{\alpha_\star}, \tag{2.18}$$

where $\varepsilon_\star$ and $\alpha_\star$ are dimensionless parameters and $\tau_{\rm disk} = r_{\rm disk}/V_{\rm disk}$ with half-mass radius of the disk, $r_{\rm disk}$, and the circular velocity of the disk, $V_{\rm disk}$. In the Baugh et al. (2005), MHIBow05, and GpcBow06 models, we use a slightly different prescription for the star formation timescale :

$$\tau_\star = \tau_\star^0 \left(\frac{V_{\rm disk}}{200 {\rm kms}^{-1}} \right)^{\alpha_\star}, \tag{2.19}$$

where $\tau_\star^0$ is an adjustable parameter. This change was motivated by the need to have more gas available at high redshifts to fuel starbursts in Baugh et al. (2005).

SN feedback

The reheating of cold gas from winds from hot stars and SNe is modelled by :

$$\dot{M}_{\rm eject} = \beta \psi, \tag{2.20}$$

where β is the efficiency of the feedback process defined by :

$$\beta = (V_{\rm disk}/V_{\rm hot})^{-\alpha_{\rm hot}}, \tag{2.21}$$

where $V_{\rm hot}$ and $\alpha_{\rm hot}$ are dimensionless adjustable parameters. The SNe feedback is effective in low mass galaxies and suppresses the formation of low luminosity galaxies, producing a galaxy luminosity function with a reasonably shallow faint

end slope, as observed (Norberg et al. 2002; Blanton et al. 2001). The parameter values adopted to explain the luminosity function at the faint end are different in the different models. The Baugh et al. (2005) model uses $V_{\rm hot} = 300\,{\rm kms}^{-1}$ and $\alpha_{\rm hot} = 2$. The Bower et al. (2006), Font et al. (2008) and the MHIBow06 models use $V_{\rm hot} = 485\,{\rm kms}^{-1}$ and $\alpha_{\rm hot} = 3.2$. The GpcBow06 model uses $V_{\rm hot} = 390\,{\rm kms}^{-1}$ and $\alpha_{\rm hot} = 3.2$.

Feedback processes in high mass dark matter haloes

With current baryon densities, hierarchical galaxy formation models have needed to invoke additional processes to regulate the formation of bright galaxies. The Baugh et al. (2005) model invokes a superwind feedback process to prevent the formation of too many luminous galaxies, as introduced by Benson et al. (2003), in which cold gas is ejected from the hot halo of a galaxy in proportion to the star formation rate. The *superwind* process is supported by evidence of outflows in the spectra of Lyman break galaxies and in local starburst galaxies (Adelberger et al. 2003; Wilman et al. 2005). In contrast to the SNe feedback in Sec. 2.21, the gas ejected by this mechanism is not allowed to cool again, even in more massive haloes. The mass ejection in the *superwind* is modelled by :

$$\dot{M}_{eject} = \beta_{sw}\psi, \tag{2.22}$$

where β_{sw} is an efficiency factor given by :

$$\beta_{sw} = f_{sw}{\rm min}[1, (V_c/V_{sw})^{-2}], \tag{2.23}$$

where f_{sw} and V_{sw} are adjustable parameters (f_{sw}=2 and V_{sw}=200kms^{-1} in Baugh et al. 2005).

The other models used in this thesis employ AGN feedback to prevent the over-production of bright galaxies, based on the Bower et al. (2006) model. This model invokes the suppression of cooling flows in massive haloes, as a result of the energy released following accretion of matter onto a central supermassive black hole. A halo is assumed to be in quasi-hydrostatic equilibrium if the time required for gas to cool at the cooling radius, $t_{\rm cool}(r_{\rm cool})$, exceeds a multiple of the free-fall time at

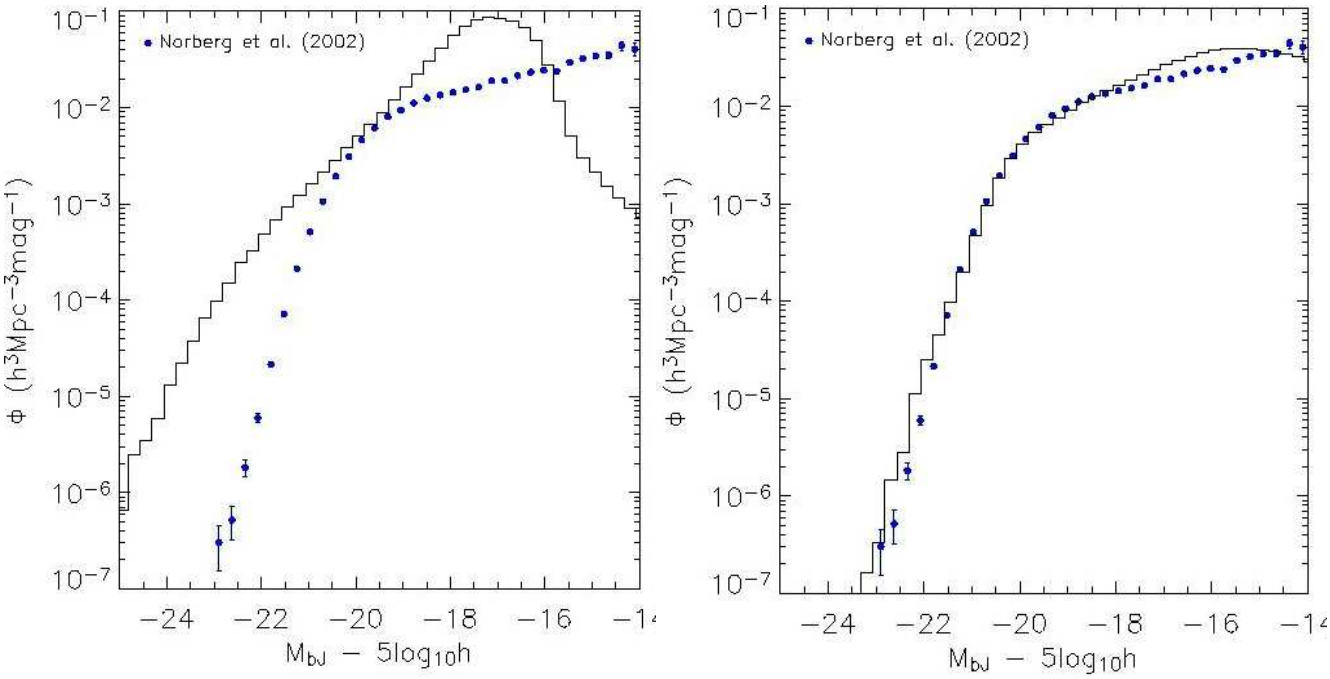

Figure 2.3: Feedback effects for the the b_J luminosity function. Left panel shows the luminosity function from the semi-analytic model before including any kinds of feedback processes. Right panel shows the luminosity function from the semi-analytic model after including a supernova feedback and AGN feedback processes. The points show the observational luminosity function from the 2dFGRS survey. Adapted from Croton et al. 2006

this radius, $t_{\rm ff}(r_{\rm cool})$:

$$t_{\rm cool}(r_{\rm cool}) > \frac{1}{\alpha_{\rm cool}} t_{\rm ff}(r_{\rm cool}), \tag{2.24}$$

where $\alpha_{\rm cool}$ is an adjustable parameter whose value controls the sharpness and position of the break in the optical galaxy luminosity function. The cooling flow in the halo is then shut down completely if the luminosity released by accretion of matter onto the supermassive black hole (SMBH) exceeds the cooling luminosity. The energy released by accretion depends on the mass of the SMBH.

Fig. 2.3 shows the effect of feedback processes to the galaxy luminosity function. To understand the observational luminosity function from the 2dFGRS survey, we included the supernovae feedback process to match the faint end of luminosity function and the AGN process for matching the bright end of luminosity function.

2.3 The formation of spheroids

We assume that the most massive galaxy automatically becomes the central galaxy in the new halo when dark matter haloes merge. All the other galaxies become satellite galaxies orbiting within the dark matter halo. The orbits of these satellite galaxies gradually decay as energy and angular momentum are lost via dynamical friction to the halo material. The satellite galaxies eventually merge into the central galaxy.

2.3.1 Dynamical friction and galaxy merger

When a new halo forms, the GALFORM model assumes that each satellite galaxy enters the halo on a random orbit. The most massive pre-existing galaxy is assumed to become the central galaxy in the new halo. Note that these assumptions are discussed further in Chapter 3 in which the goal is to reproduce the luminosity dependence of clustering in the 2dFGRS data. The time for a satellite's orbit to decay because of the effect of dynamical friction depends on the initial energy and angular momentum of the orbit. The time for an orbit to decay in an isothermal halo is based on the standard Chandrasekhar formula for dynamical friction given by Lacey & Cole (1993) as :

$$\tau_{\rm mrg} = f_{\rm df}\Theta_{\rm orbit}\tau_{dyn}\frac{0.3722}{\ln(\Lambda_{\rm Coulomb})}\frac{M_H}{M_{\rm sat}}, \tag{2.25}$$

where M_H is the mass of the main dark matter halo, which includes the central and satellite galaxies, and $M_{\rm sat}$ is the mass of a dark matter sub-halo which includes the satellite galaxy. We take the Coulomb logarithm to be $\ln(\Lambda_{\rm Coulomb}) = \ln(M_H/M_{\rm sat})$. The dynamical time in the new halo is defined by $\tau_{\rm dyn} \equiv \pi r_{\rm vir}/V_H$. $f_{\rm df}$ is a dimensionless parameter. $\Theta_{\rm orbit}$ is an orbital parameter defined as :

$$\Theta_{orbit} = [J/J_c(E)]^{0.78}[r_c(E)/r_{vir}]^2, \tag{2.26}$$

where E is the initial energy and J is the initial angular momentum of the satellite's orbit. $r_c(E)$ and $J_c(E)$ are the radius and angular momentum of a circular orbit with the same energy as that of the satellite galaxy.

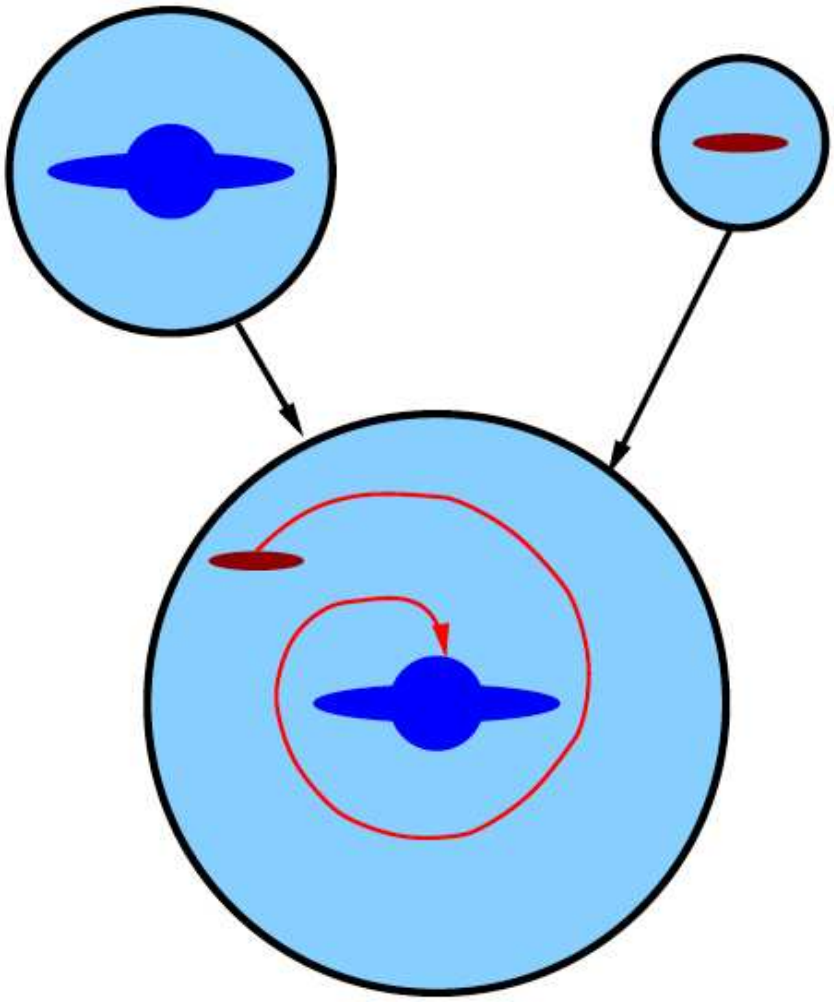

Figure 2.4: A schematic of a galaxy merger between two dark matter haloes. The progenitors of the final halo each contain a galaxy. After the two haloes merge, the more massive galaxy is placed at the centre of the newly formed halo (the size of circle indicates the mass of dark matter). The orbit of the satellite galaxy decays due to dynamical friction. The satellite galaxy may eventually merge with the central galaxy. Adapted from Baugh (2006).

The basic scheme of a galaxy merger in GALFORM is shown in Fig. 2.4. The central galaxy in the newly formed halo is the more massive progenitor galaxy. The less massive galaxy in the progenitors becomes a satellite galaxy in the newly formed halo and may eventually merge with the central galaxy after losing its energy by dynamical friction.

2.4 The GRASIL model

Even though GALFORM itself makes an independent calculation of the spectral energy distribution (SED) of starlight, including a dust extinction model which is described in Cole et al. (2000). This calculation gives similar results to those obtained with GRASIL at optical wavelengths. We accurately compute SED of model galaxies using the spectrophotometric code GRASIL (Silva et al. 1998; Granato et al. 2000). GRASIL computes the emission from the stellar population, the absorption and emission of radiation by dust, and also radio emission by thermal and synchrotron processes powered by massive stars (Bressan et al. 2002). GALFORM calculates the basic ingredients for the GRAISL code to calculate SED for each galaxy: (i) the star formation and metal enrichment history which including the contribution from starbursts. (ii) the scale length of the disk and bulge. components. (iii) the cold gas mass. The dust is modelled in GALFORM as a two phase medium, with a diffuse component and dense molec- ular clouds. The mass split between these components is a model parameter. For example, in the Bau05 model, 25 per cent of the dust is assumed to be in the form of dense clouds.

Based on the local radiation field, GRASIL calculates the radiative transfer of starlight through the dust and self- consistently solves for the temperature distribution of the dust grains at each point in the galaxy. The dust emission calculate using the temperature distribution of the grains. The composition and size of the dust grains are chosen to match the properties of the local ISM: a mixture of graphite and silicate grains, as well as polycyclic aromatic hydrocarbon (PAH) molecules. The effects of temperature fluctuations in very small grains and PAH molecules are taken into account. Emission from PAHs is calculated using the cross-sections of Li

& Draine (2001). Radio emission from ionized HII regions and synchrotron radiation is included as in Bressan et al. (2002). The output from GRASIL is the complete SED of a galaxy in GALFORM model from the far-UV to the radio (wavelengths 0.1 μm $\leq \lambda \leq$ 1m).

2.5 Summary of models

In this section we show the values of key parameters which are varied in the different models to reproduce observational results. We present the values of these parameters in Table. 2.2. For a full description of the DeLucia07 model used in Chapters 3 and 4, see Croton et al. (2006) and De Lucia et al. (2007).

Table 2.2: The values of selected parameters which are used models in this thesis. The columns are as follows: (1) The name of the model. (2) The equation used to calculate the star formation timescale, $\tau_\star$. (3) The value of $\epsilon_\star$ (Eq. 2.18) or $\tau_0^\star$ (Eq. 2.19) used in the star formation timescale. (4) The feedback processes in high mass halos, specified by $\alpha_{\rm cool}$ (Eq. 2.24) or β_{sw} (Eq. 2.23) . (5) The supernova feedback parameter, $V_{\rm hot}$ (Eq. 2.21). (6) The supernova feedback parameter, α_{hot} (Eq. 2.21). (7) The source of halo merger histories. (8) Comments giving model source or key differences from published models.

	$\tau_\star$	$\epsilon_\star$ or $\tau_\star^0$	$\alpha_{\rm cool}$ or $f_{sw}(V_{sw})$	$V_{\rm hot}[{\rm kms}^{-1}]$	α_{hot}	Merger tree	Comments
Bow06 or Bower06	Eq. 2.18	0.0029	0.58	485	3.2	N-body	Bower et al. (2006)
Font08	Eq. 2.18	0.0029	0.70	485	3.2	N-body	Font et al. (2008) Modified cooling recipe in satellites and yield from Bow06
MHIBow06	Eq. 2.19	8 Gyr	0.62	485	3.2	N-body	Modified star formation recipe from Bow06
GpcBow06	Eq. 2.19	4 Gyr	0.72	390	3.2	Monte Carlo	Different background cosmology and modified star formation recipe from Bow06
Baugh05	Eq. 2.19	8 Gyr	2(200km/s)	300	2	Monte Carlo	Different background cosmology superwind prescription in high mass halo

Chapter 3

Modelling galaxy clustering: Is new physics needed in galaxy formation models?

3.1 Introduction

The clustering of galaxies encodes information about the values of the cosmological parameters and also about the physical processes behind the formation and evolution of galaxies. In the cold dark matter (CDM) hierarchical structure formation theory, galaxies grow inside dark matter haloes (White & Frenk 1991; Cole 1991). The formation of structure in the dark matter is governed by gravity and can be modelled accurately using N-body simulations (e.g. Springel, Frenk & White 2006). However, the fate of baryonic material is much more complicated as it involves a range of often complex and nonlinear physical processes. The efficiency of galaxy formation is expected to depend on the mass of the host dark matter halo (e.g. Eke et al. 2004; Baugh 2006). Modelling the dependence of galaxy clustering on intrinsic properties such as luminosity offers a route to establish how such properties depend upon the mass of the host halo and hence to improve our understanding of galaxy formation.

Over the past decade, models of galaxy clustering have evolved which allow us to interpret observational data and learn more about how galaxies are distributed between dark matter haloes. This development has been led by semi-analytical models,

which can populate large volumes with galaxies in a short time using physically motivated prescriptions (Kauffman, Nusser & Steinmetz 1997; Kauffmann et al. 1999; Benson et al. 2000; Benson et al. 2001; Hatton et al. 2003). Such studies also inspired empirical approaches which involve fitting halo occupation distributions (HODs; Peacock & Smith 2000; Seljak 2000; Cooray & Sheth 2002) and conditional luminosity functions (van den Bosch, Mo & Yang 2003), parametric functions which describe the number of galaxies per halo and the luminosity of galaxies within a halo, respectively.

Recent advances in astronomical instrumentation have produced a wealth of information on galaxy clustering. The enormous volume and number of galaxies in the two-degree field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS York et al. 2000) have made possible accurate measurements of clustering for samples of galaxies defined by various intrinsic properties (Norberg et al. 2001, 2002; Zehavi et al. 2002, 2005; Madgwick et al. 2003; Li et al. 2006). The variation of clustering strength with luminosity tells us how galaxies populate haloes and hence about the physics of galaxy formation. Any discrepancy between the observational measurements of clustering and theoretical predictions points to the need to improve the models, either by refining existing ingredients or adding new ones.

The dependence of galaxy clustering on luminosity has been measured accurately in the local universe (Norberg et al. 2001, 2002, 2009, in preparation; Zehavi et al. 2002, 2005; Li et al. 2006). Over the period spanned by these studies, galaxy formation models have evolved significantly, particularly in the treatment of bright galaxies (see, for example, Benson et al. 2003). The majority of current models invoke some form of heating of the hot gas atmosphere to prevent gas cooling in massive haloes, in order to reproduce the bright end of the galaxy luminosity function. This has implications for the correlation between galaxy luminosity and host dark matter halo mass, which has, in turn, an impact on the clustering of galaxies.

Li et al. (2007) compared the semi-analytical galaxy formation models of Kang et al. (2005) and Croton et al. (2006), two early models with AGN feedback, against measurements of clustering from the SDSS. Qualitatively, the models displayed sim-

ilar behaviour to the real data, but did not match the clustering measurements in detail. For example, Li et al. show that as the luminosity varies the predictions of the Croton et al. model change in clustering amplitude by a similar amount to the observations. The brightest galaxies are the most strongly clustered in the model. However, the clustering strength displays a minimum around L_* before increasing again for fainter galaxies. The luminosity dependence in the SDSS data, on the other hand, is monotonic. Li et al. speculated that the models predict too many galaxies in massive haloes. They demonstrated that the clustering predictions could be improved, but not fully reconciled with the data, by removing satellite galaxies by hand.

In this chapter, we extend this comparison to the 2dFGRS clustering measurements and test the latest galaxy formation models. By using the blue selected 2dFGRS, we widen the range of physics tested to include the processes which influence recent star formation. We compare models produced by different groups which allows us to probe different implementations of the physics. We reach similar conclusions to those of Li et al. and investigate physical ways to achieve the required reduction in the number of satellites.

The structure of this chapter is as follows. We briefly introduce the three semi-analytic models we discuss in Section 3.2. In Section 3.3, we compare the two point correlation function results for the 2dFGRS with the theoretical predictions. In Section 3.4, we explore the mechanisms that drive clustering, particularly the galaxy luminosity – host halo mass relation and give a step-by-step illustration of how the number of galaxies as a function of halo mass (the Halo Occupation Distribution) is connected to the clustering amplitude. We empirically determine the HOD which reproduces the observed luminosity dependence of clustering in Section 3.5. We implement simple models for two new physical processes in Section 3.6, to see if we can modify the existing models to match the observed clustering. Finally, in Section 3.7, we give a summary and conclusions.

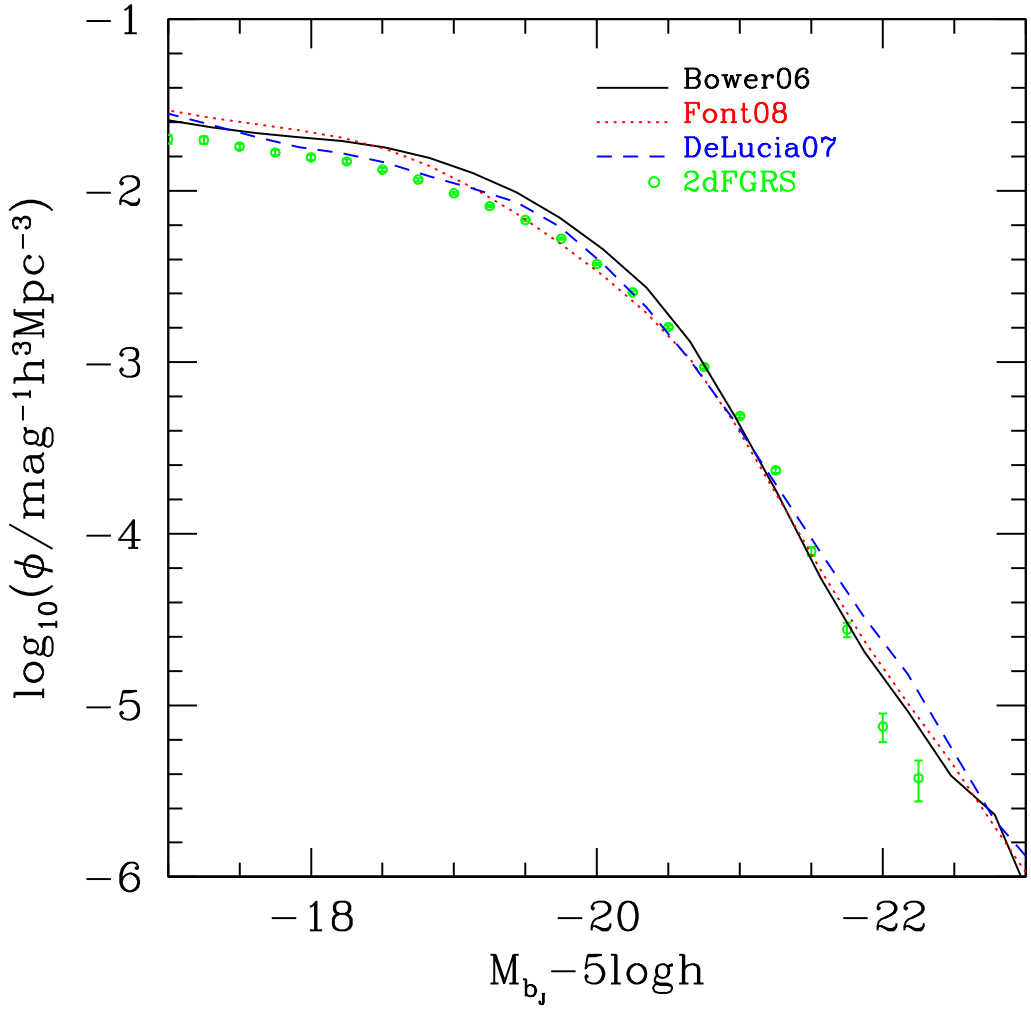

Figure 3.1: The b_J-band luminosity function of the Bower et al. (2006; black, solid line), De Lucia & Blaizot (2007; blue, dashed line) and Font et al. (2008; red, dotted line) models. The green symbols show the estimate of the luminosity function made from the 2dFGRS (from Norberg et al. 2002).

3.2 Galaxy formation models

To make predictions for the clustering of galaxies, we need a theoretical tool which can populate large cosmological volumes with galaxies. Furthermore, it is essential that we have well developed predictions for the properties of the model galaxies, in order that we can extract samples which match different observational selection criteria. Gas dynamic simulations currently struggle to meet both of these requirements. Such calculations demand high resolution which limits the accessible computational volume. Also, the level of sophistication of the model predictions in gas simulations is not always sufficient to make direct contact with observational quantities. Semi-analytical models, on the other hand, meet both of the above requirements, whilst inevitably addressing the physics in a more idealised manner than the gas simulations, and are therefore well suited to clustering studies (for an overview of this approach see Baugh 2006).

In the first half of this chapter we consider predictions for galaxy clustering from three semi-analytical models, those of Bower et al. (2006), de Lucia & Blaizot (2007) and Font et al. (2008). These models are publicly available from the Millennium Galaxy Archive[1]. In the second part, we consider modifications to the Bower et al. model. We shall also refer to the Bower et al. and Font et al. models as the Durham models (and as Bower06 and Font08 respectively in figure labels) and to the de Lucia & Blaizot model as the Munich model (and as DeLucia07 in plots).

The three models listed above are set in the context of structure formation in a cold dark matter universe as modelled by the Millennium Simulation[2] of Springel et al. (2005). The starting point is the merger histories of dark matter haloes, which are extracted from the simulation (note both groups have independent algorithms for constructing merger histories; see Springel et al. 2005 and Harker et al. 2006

[1] http://galaxy-catalogue.dur.ac.uk:8080/Millennium/

[2] The cosmological parameters of the Millennium Simulation are a matter density parameter of $\Omega_{\rm M} = 0.25$, cosmological constant of $\Omega_\Lambda = 0.75$, baryon density $\Omega_{\rm b} = 0.045$, a fluctuation amplitude of $\sigma_8 = 0.9$, spectral index of perturbations of $n_{\rm s} = 1$ and present-day Hubble constant of $H_0 = 73 {\rm kms}^{-1}{\rm Mpc}^{-1}$. The halo mass resolution limit does not affect the galaxy luminosity function at the luminosities studied in this chapter.

for further details). The models follow a common range of processes which involve the baryonic component of the universe: gas cooling, star formation, reheating of cold gas by supernovae, chemical evolution of gas reservoirs, heating of the hot gas halo by AGN and galaxy mergers. The implementation of these processes differs in detail between the models and we refer the reader to the original references for a full description. Moreover, when setting the model parameters, different emphasis was placed on the reproduction of particular observational datasets. Here we simply remark on some key features of the models.

Bower et al. (2006) use the model of Malbon et al. (2007) to describe the growth of supermassive black holes through galaxy mergers, and the accretion of cold and hot gas. The latter process is the key to matching the sharpness of the break in the local optically selected galaxy luminosity function. The energy released by the accretion of hot gas onto the black hole is assumed to match the luminosity which would have been released by gas cooling, thereby suppressing the formation of bright galaxies (see Croton et al. 2006). The Font et al. (2008) model is a development of the Bower et al. model. Firstly, in the Font et al. model the stellar yield in all modes of star formation is twice that adopted in the Bower et al. model. This shifts the locus of the red and blue sequences in the colour magnitude relation into better agreement with local data from the Sloan survey (see Gonzalez et al. 2008 for a comparison of the predicted colour distributions with SDSS observations). Secondly, in the Font et al. model the stripping of the hot gas from newly accreted satellite galaxies is not assumed to be 100% efficient. This is different from the assumption commonly made in semi-analytical models and is motivated by the results of recent gas dynamics simulations carried out by McCarthy et al. (2008). This means that in the Font et al. model galaxies can continue to accrete cold gas even after they have been subsumed into a more massive halo. This results in an improved match to the observed colour distribution of satellite galaxies (Gonzalez et al. 2008). Both the Bower et al. and Font et al. models give very good matches to the stellar mass function over the full redshift range for which observational estimates are available.

The De Lucia & Blaizot (2007) model is a development of the semi-analytical models of Springel et al. (2001), De Lucia et al. (2004) and Croton et al. (2006).

Luminosity and colour dependent clustering were discussed in Springel et al. (2005) and Croton et al. (2006); the De Lucia & Blaizot model gives similar clustering predictions to those from these earlier models.

The parameters of the models are set to give a reasonable reproduction of the present day galaxy luminosity function, as shown by Fig. 3.1. In this chapter we give ourselves the freedom to adjust the luminosities of the model galaxies, whilst maintaining the ranking of galaxy luminosity, to force an exact match to the 2dFGRS luminosity function measured by Norberg et al. (2002a). This small adjustment allows us to rule out abundance differences as a possible source of variations between the clustering predictions of different models. Note that we also have checked the correlation function using the galaxy luminosity bins before forcing to match the 2dFGRS luminosity function. We have nearly identical results between two correlation functions between them. We apply the same methodology to the modified versions of the Bower et al. model discussed in the second part of the chapter.

3.3 Predictions for luminosity dependent clustering

In this section we compare the predictions of the three galaxy formation models (Bower et al. 2006; De Lucia & Blaizot 2007; Font et al. 2008) with measurements of clustering made from the final two-degree field galaxy redshift survey (Norberg et al. 2009b). The observational data are presented in the form of the projected correlation function, $\Xi(\sigma)/\sigma$. This statistic is estimated from the two point correlation function binned in pair separation parallel and perpendicular to the line of sight, $\xi(\sigma, \pi)$:

$$\frac{\Xi(\sigma)}{\sigma} = \frac{2}{\sigma} \int_0^\infty \xi(\sigma, \pi) \mathrm{d}\pi. \tag{3.1}$$

When redshift is used to infer the radial distance to a galaxy, gravitationally induced peculiar motions on top of the Hubble flow cause a distortion to the inferred clustering signal. In principle, the projected correlation function is unaffected by the contribution from peculiar velocities. In practice, the integration in Eq. 3.1 has to be truncated at a finite value of π as the clustering signal on larger scales becomes

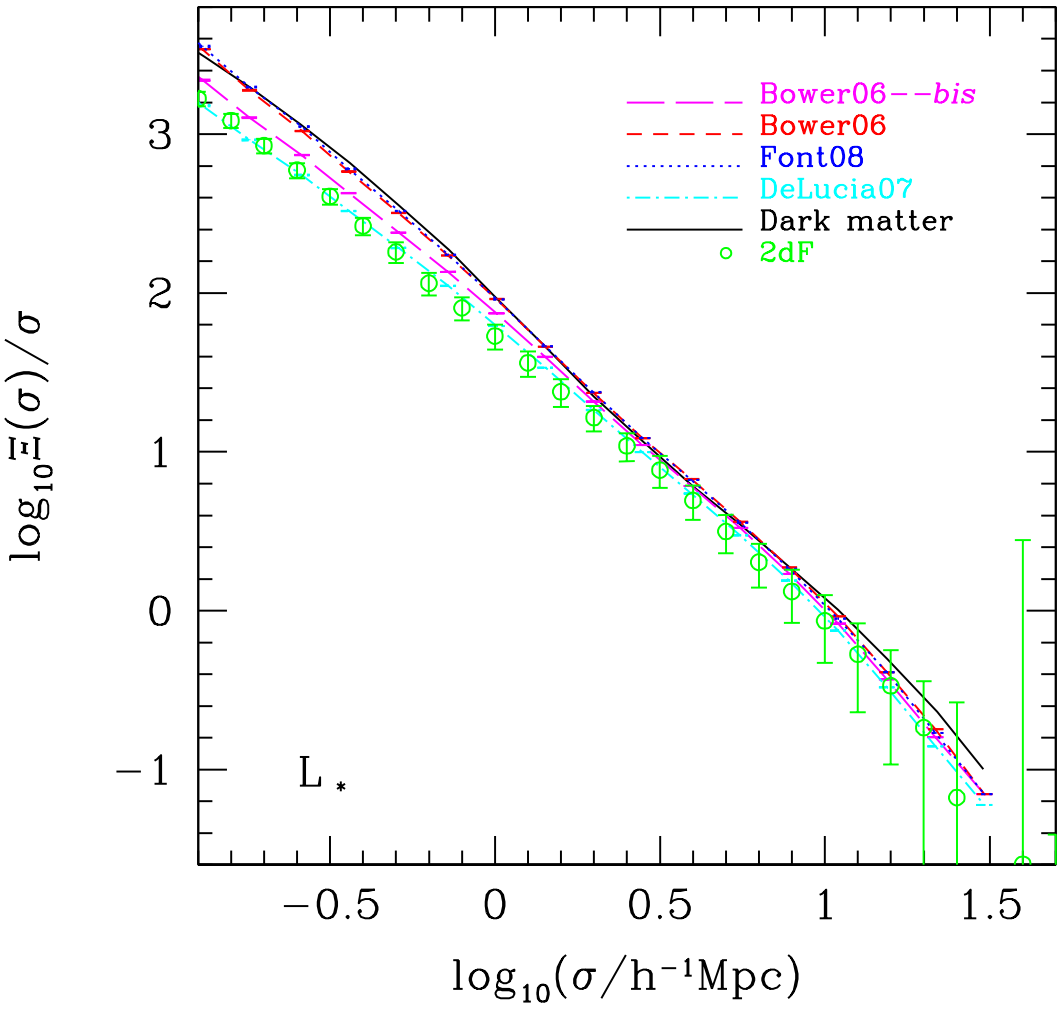

Figure 3.2: The projected correlation function of L_* galaxies measured in the 2dF-GRS by Norberg et al. (2009; open symbols). The model predictions are shown by different coloured lines, as indicated by the key. The projected correlation function of the dark matter in the Millennium simulation is shown by the black solid line.

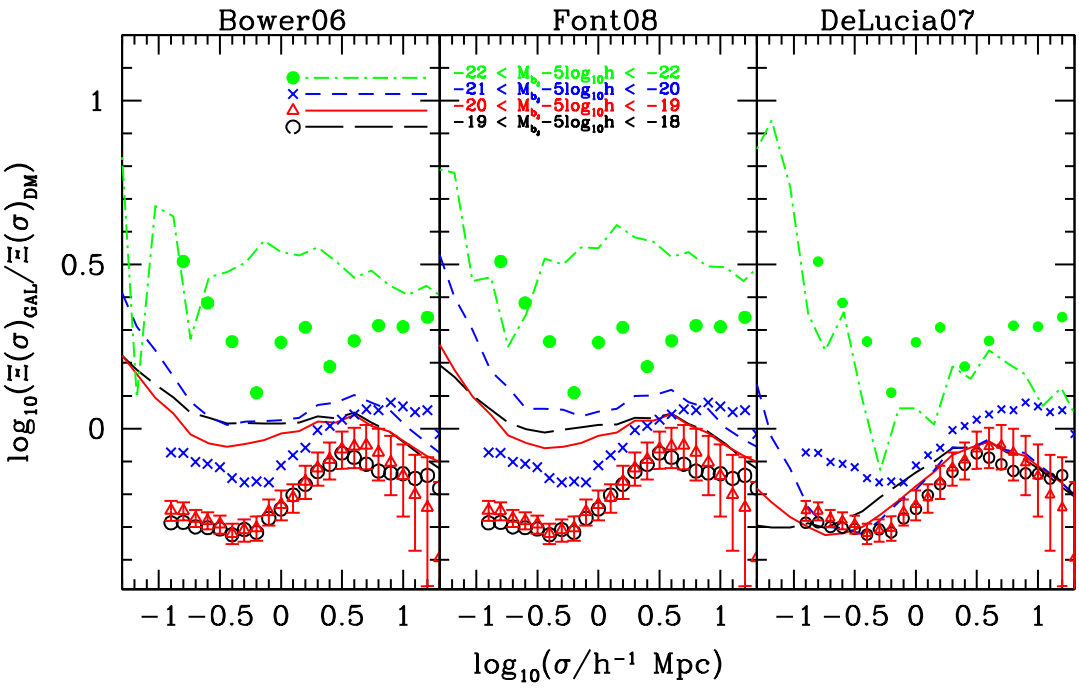

Figure 3.3: The projected galaxy correlation functions divided by the projected correlation function of the dark matter in the Millennium Simulation. The symbols show the ratios for the 2dFGRS clustering measurements. Different colours and line types show the different luminosity bins as indicated by the key. The model predictions are shown by the lines. Each panel shows the predictions for a different model, as indicated by the label.

noisy. Norberg et al. (2009a) show that this truncation has a negligible effect on the form of the projected correlation function on scales below $10h^{-1}$Mpc.

The observational measurements we use in this chapter are from the final 2dFGRS. Previous results for the luminosity dependence of galaxy clustering were presented by Norberg et al. (2001,2002). These papers analysed an intermediate version of the 2dFGRS which consisted of around 160 000 unique, high quality galaxy redshifts. In the final version of the dataset used by Norberg et al. (2009) there are more than 220 000 galaxy redshifts. The solid angle of high spectroscopic completeness regions has also increased, by a larger factor than the change in the total number of redshifts. Hence a more accurate measurement of the clustering in different volume limited samples is now possible. The estimation of errors on the clustering measured for the different samples has also been revisited (Norberg et al. 2008). An internal estimate of the error is made using the bootstrap resampling technique. This has the advantage over the mock catalogues used previously that the change in clustering strength with luminosity is taken into account. The 2dFGRS is selected in the blue $b_{\rm J}$ band. This is more sensitive to recent episodes of star formation in galaxies than the red r band selection used in the SDSS.

We first examine the clustering of L_* galaxies. Fig. 3.2 compares the model predictions and the 2dFGRS measurement for the projected correlation function of L_* galaxies, along with the projected correlation function of the dark matter in the Millennium Simulation. On large scales, $\sigma > 3h^{-1}$Mpc, the models have a similar shape to the observations, but different amplitudes. The Durham models (Bower et al. and Font et al.) have a higher clustering amplitude than the data, but are similar to the dark matter. On small scales, $\sigma \leq 1h^{-1}$Mpc, the Durham models are significantly above the 2dFGRS measurement. The De Lucia & Blaizot prediction is a remarkably good match to the L_* clustering data over the full range of scales plotted. As we will see in the next section, the clustering predictions can be broken down into contributions from the most massive galaxy in each halo, referred to as the central galaxy, and satellite galaxies. The form of the projected correlation function on small scales is driven by the number of satellites in massive haloes. One interpretation of the comparison in Fig. 3.2 is that massive haloes in the Durham

models contain more L_* satellites relative to low mass haloes than in the Munich model. This would also account for the small difference between the predicted clustering amplitudes on large scales.

Over a range of just over two decades in projected pair separation, Fig. 3.2 shows that the clustering amplitude changes by four and a half orders of magnitude. In order to see more clearly the changes in the clustering amplitude with varying galaxy luminosity, in Fig. 3.3 we divide the galaxy correlation functions by the dark matter correlation function. If the Millennium Simulation dark matter was indeed a match to the real Universe, then the ratio plotted in Fig. 3.3 would be the logarithm of the square of the bias, albeit quantified in terms of projected clustering. The departure of this ratio from a constant value would then indicate the presence of a scale-dependent bias. However, it is of course possible that the Millennium Simulation is not quite representative of reality, with recent studies suggesting a lower value of the fluctuation amplitude σ_8 (Sanchez et al. 2009; Li & White 2009). Nevertheless, the Millennium dark matter serves as a useful benchmark, even if these caveats limit the interpretation of the ratio.

Fig. 3.3 shows that the clustering amplitude between the 2dFGRS galaxy samples varies by a factor of between 3 and 5 depending on the projected separation. The two faintest samples plotted have almost the same clustering amplitude. None of the models is able to reproduce both the overall amplitude of clustering and the trend with luminosity. The Durham models do a reasonable job of predicting the order of the clustering strength of the luminosity samples. However, the overall clustering amplitude in these models is too strong. The Munich model, on the other hand, gives a better match to the overall clustering amplitude, but fails to reproduce a dependence of clustering on luminosity. The brightest sample in the DeLucia07 model is more weakly clustered than the corresponding sample in the 2dFGRS and there is essentially no difference in the clustering displayed by the three fainter bins; in the 2dFGRS data, the $-21 < M_{b_\mathrm{J}} - 5\log h < -20$ sample is clearly more strongly clustered than the L_* and fainter galaxies.

The correlation function ratios plotted in Fig. 3.3 show strong scale dependence. On the largest scales plotted, this could indicate that the clustering of dark matter

in the Millennium cosmology is not the same as in the real Universe, as we remarked upon above. However, the 2dFGRS measurements become noisy on the scales on which one would expect the bias to approach a constant value (e.g. Coles 1993). On small scales there is a range of shapes and amplitudes, indicating a wide variety of satellite fractions in the different galaxy samples. Apart from the brightest sample, the Durham models show a higher clustering amplitude on small scales than the Munich model and also a higher amplitude than the observations. This suggests that there are too many satellite galaxies in haloes in the Durham models, a conclusion which we confirm in the next section.

3.4 What drives galaxy clustering?

In this section, we look at the clustering predictions in more detail to identify which galaxies determine the shape and amplitude of the correlation function. This will allow us to identify which model galaxies are responsible for the disagreement found with observational measurements in the previous section, and hence will motivate approaches to altering the model predictions for these objects.

The clustering of dark matter haloes depends on their mass. Haloes which are more massive than the characteristic mass scale at a particular redshift (roughly the location of the break in the halo mass function) will be much more strongly clustered than the overall dark matter (Cole & Kaiser 1989; Mo & White 1996). We start by plotting the relation between galaxy luminosity and the mass of the host dark matter halo in Fig. 3.4. The main panel in each plot shows the median host halo mass and 10-90 percentile range of the distribution as a function of luminosity, for satellite and central galaxies separately. The sub-panel shows the fraction of galaxies that are satellites at each magnitude.

Overall, the host halo mass – galaxy luminosity relations for the different models share the same qualitative behaviour. There is a trend of increasing host mass with increasing central galaxy luminosity which steepens around $M_{b_J} - 5\log h \approx -21$. A magnitude brighter than this, the median host halo mass drops in each case. The scatter in host mass is small at the faintest luminosities plotted (around a factor of

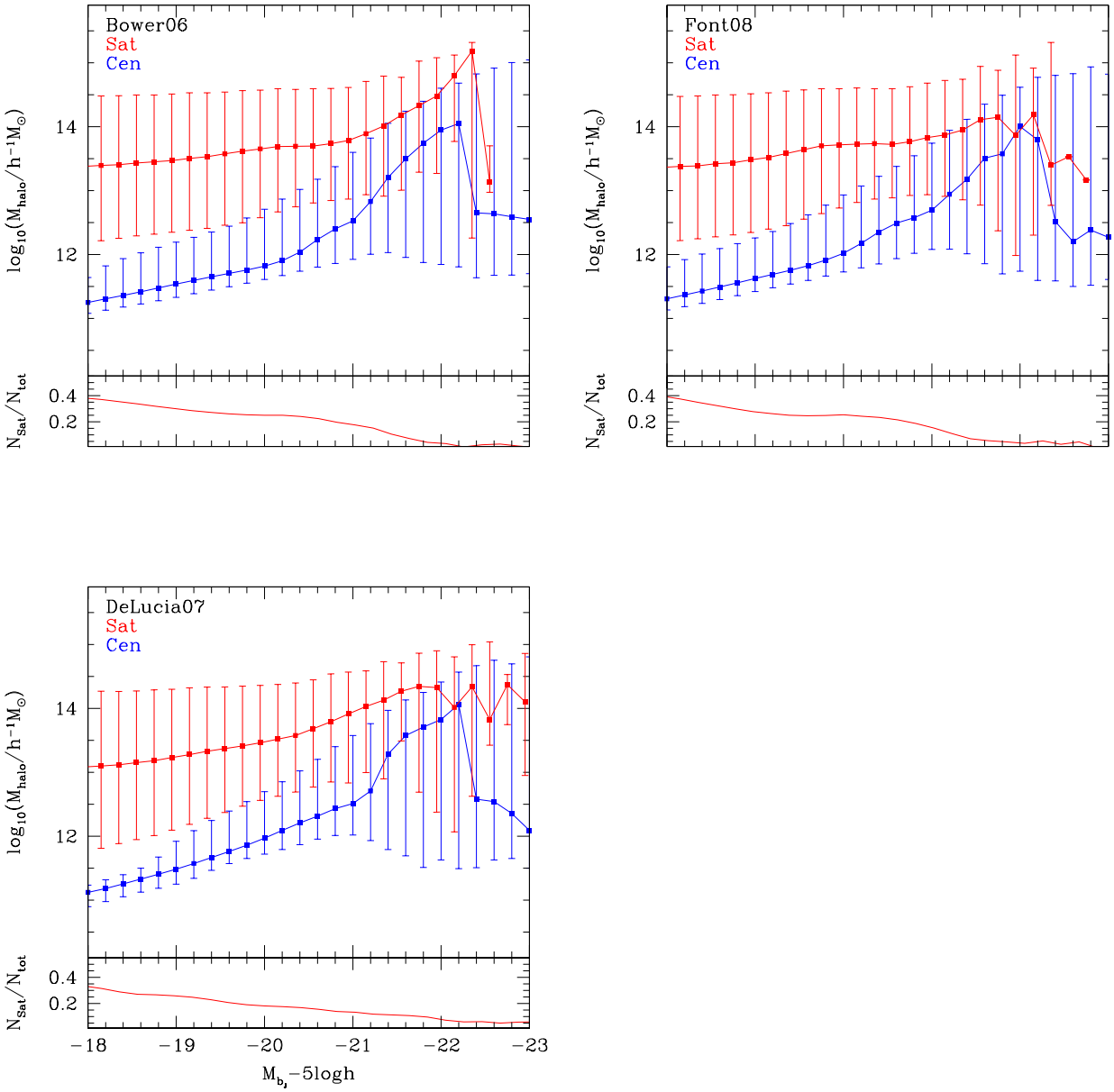

Figure 3.4: The host halo mass for galaxies as a function of luminosity. The main window in each panel shows the predictions for a different galaxy formation model, with Bower et al. shown in the top panel, Font et al. in the middle panel and de Lucia & Blaizot in the lower panel. The median mass and 10-90 percentile ranges are shown separately for central (blue) and satellite (red) galaxies. The small window in each panel shows the fraction of galaxies that are satellites as a function of magnitude.

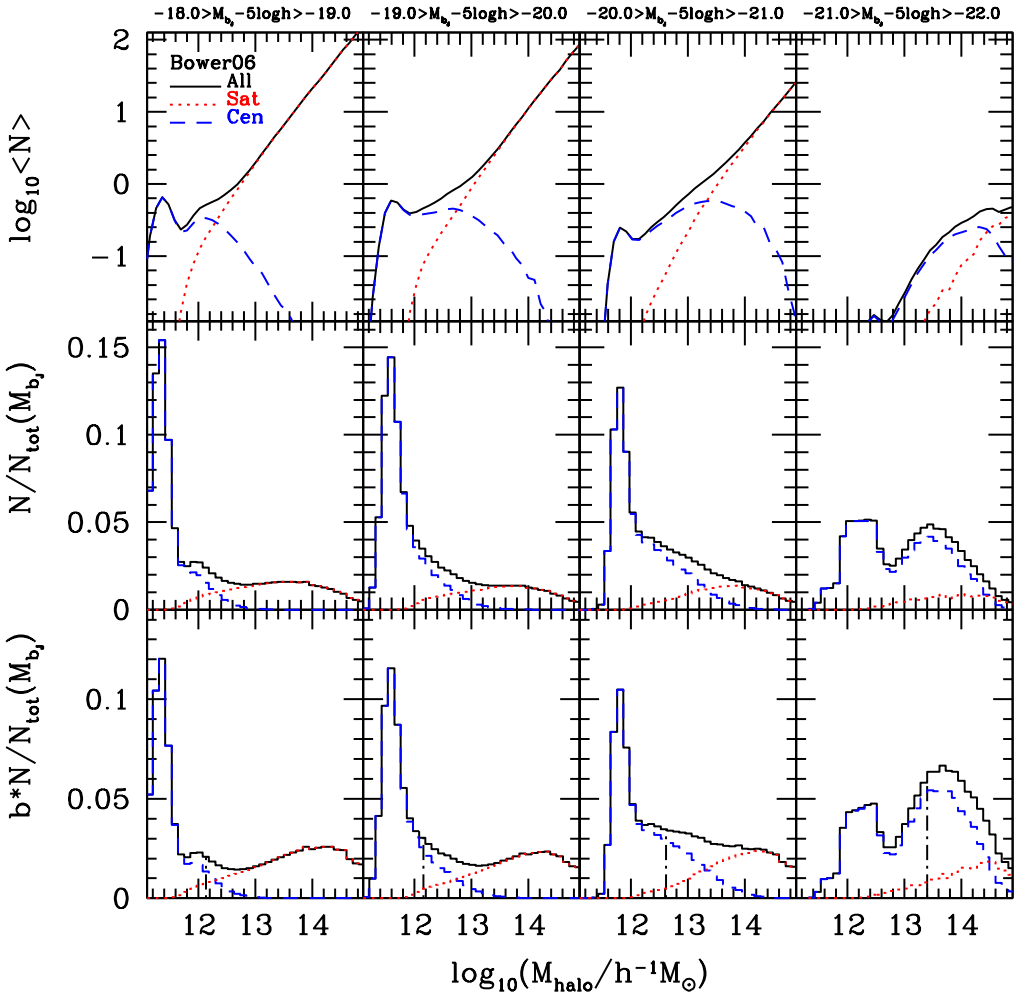

Figure 3.5: The steps connecting the number of galaxies per halo to the strength of galaxy clustering in the Bower et al model. Each column corresponds to a different galaxy sample, as indicated by the label. The blue curves show the contribution from central galaxies, the red dotted curves show satellite galaxies and the black solid curves show centrals plus satellites. The top row shows the galaxy halo occupation distribution. The middle row shows this HOD multiplied by the dark matter halo mass function and normalized by the total number of galaxies in the luminosity bin. The bottom row shows the HOD multiplied by the halo mass function and the halo bias, again normalized by the total number of galaxies in the luminosity bin. In this case the area under the black solid curve is the effective bias of the sample. The vertical dot-dashed line in the lower panels indicate the mass which divides the area under the curve in half.

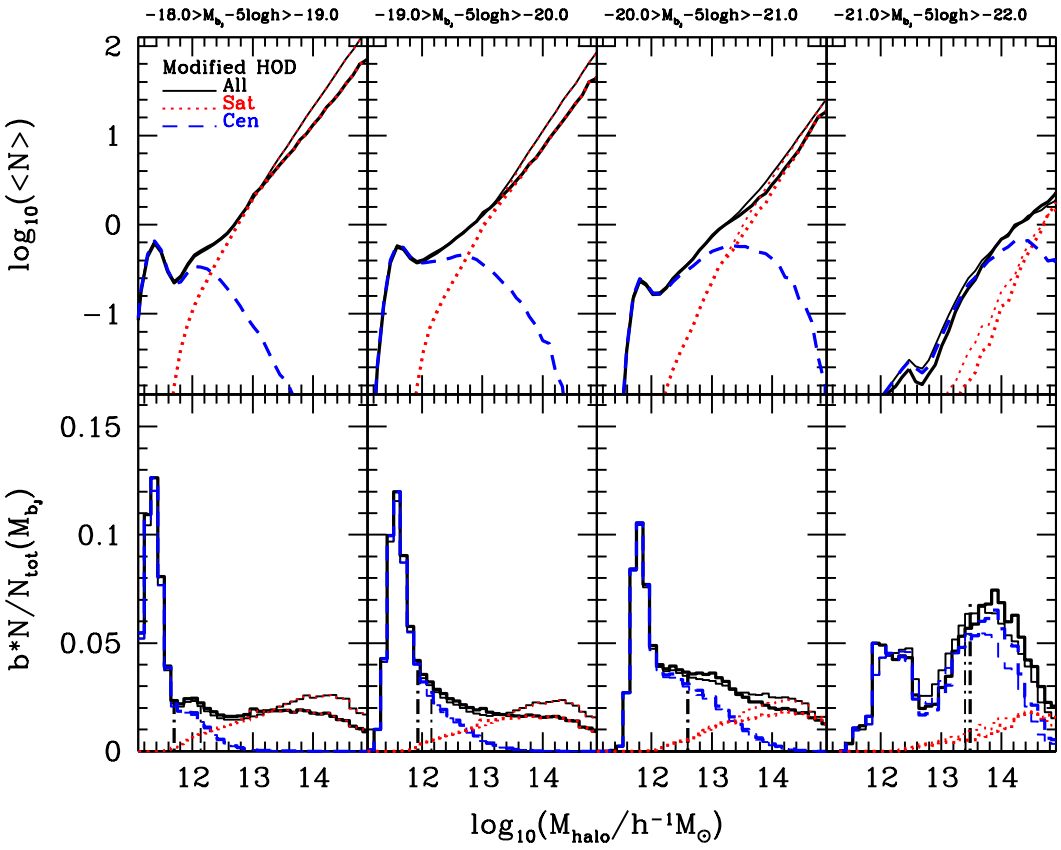

Figure 3.6: *Top row*: A comparison of the modified HOD (thick lines) in which the slope of the satellite HOD has been adjusted to match the 2dFGRS clustering measurements with the original HOD of the Bower et al. model. *Bottom row*: The contribution to the effective bias as a function of halo mass. The quantity plotted is the modified HOD weighted by the halo mass function and the halo bias parameter; the area under the black solid curve gives the effective bias parameter. Each column corresponds to a different luminosity bin as shown by the label. The blue dashed curves show the contribution of central galaxies, satellites are shown in red dotted and the total is shown in black solid. The vertical lines mark the halo mass which divides the contribution to the effective bias integral into two. The dot-dashed lines show this mass for the original Bower et al. model and the thick dot-dashed lines for the modified HOD.

2 in the de Lucia & Blaizot model), and increases with luminosity. For the brightest galaxies shown in Fig. 3.4, the 10-90 percentile range covers more than 2 orders of magnitude in halo mass. The median host mass of satellite galaxies does not increase with luminosity as quickly as it does for the centrals (an order of magnitude increase in host mass over the magnitude range $-18 > M_{b_J} - 5\log_{10} h > -22$, compared with two orders of magnitude for the centrals). The 10-90 percentile range is very broad for faint and intermediate luminosity satellites ($\sim$ 2 orders of magnitude) and shrinks only for the brightest satellites.

The quantitative differences between the models in Fig. 3.4 explain the differences in the predictions for luminosity dependent clustering evident in Fig. 3.3. Firstly, the median host mass relations for the central and satellite galaxies in the de Lucia & Blaizot model are lower than those in the Bower et al. and Font et al. models. This means that the overall amplitude of clustering is lower in the de Lucia & Blaizot model as seen in Fig. 3.3. Secondly, the scatter in the mass – luminosity relation for centrals is substantially smaller in the Munich model than it is in the Durham models, particularly for fainter galaxies. This means that the halo mass – luminosity relation is better defined in the Munich model compared with the Durham models, which explains the somewhat stronger trend of luminosity dependent clustering displayed in the Munich model.

The difference in the width of the distribution for the central galaxies could be driven by the choice of time over which gas is allowed to cool in a halo. In the Munich model, gas is allowed to cool over a dynamical time. In the Durham models, the cooling time depends upon the merger history of the individual trees. For haloes of a given mass, there will therefore be a range of cooling times in the Durham models, but a fixed cooling time in the Munich model.

The subpanels in each part of Fig. 3.4 show the fraction of satellite galaxies as a function of luminosity. For all the models, the fraction declines to brighter magnitudes. Due to the wide range of halo masses occupied by satellites, and the strong dependence of bias or clustering strength on halo mass, it is possible for satellites to make an important contribution to the overall clustering signal, even if they are outnumbered by centrals. We investigate this point in more detail next

in this section. The fraction of satellites in the Durham models is somewhat higher fraction for intermediate luminosities in the Durham model which is not present in the Munich models. This suggests that we should focus on reducing the number of satellite galaxies in order to improve the Durham model predictions for luminosity dependent clustering.

An alternative way to present the information contained in the host mass – luminosity plot is the halo occupation distribution (HOD; for a review see Cooray & Sheth 2002). The HOD gives the mean number of galaxies as a function of dark matter halo mass, divided into the contribution from central and satellite galaxies. The HOD has the advantage over the host halo mass – galaxy luminosity plot that it can be more directly related to galaxy clustering (e.g. Benson et al. 2000; Berlind et al. 2003).

The HOD for the Bower et al. model is shown in the top row of Fig. 3.5, in which each column shows the HOD for galaxies in a different bin in absolute magnitude. The bins are one magnitude wide, whereas in the majority of cases in the literature, cumulative bins are used. The generic form adopted for the HOD is a step function for central galaxies, which makes the transition from 0 to 1 galaxies per halo at some halo mass threshold, which is determined by the galaxy selection (e.g. Zehavi et al. 2002). More gradual forms for the transition from 0 to 1 galaxy per halo have been discussed (Zheng et al. 2005). The HOD for satellites is assumed to be a power-law with slope α; the mean number of satellites per halo reaches unity at a somewhat higher halo mass than that at which the mean number of central galaxies first approaches unity. The satellite galaxy HOD for the Bower et al. model agrees with the standard HOD paradigm. The central galaxy HOD, on the other hand, has a richer structure. The downturn seen at high masses is due to the adoption of a differential, finite width magnitude bin. With increasing halo mass, the central galaxies eventually become too bright to be included in a particular magnitude bin. For all the luminosity bins plotted, the HOD of central galaxies does not reach unity, in contradiction to one of the primary assumptions in HOD modelling. The central HOD rises to a peak just below unity, before showing a dip with increasing halo mass. This feature is due to AGN heating which suppresses

gas cooling above $M \sim 10^{12}h^{-1}M_{\odot}$ at the present day in this model. This spike has a similar appearance in the Font et al. model, even though the "switch-on" of AGN heating feedback is handled in a more gradual way in this case.

The HOD does not tell us the full story about galaxy clustering, but is only the first step. The next relevenat consideration is the abundance of dark matter haloes. The number density of haloes declines exponentially with increasing mass beyond the characteristic mass (see for example Jenkins et al. 2001). The HOD weighted by the halo mass function is shown in the second row of Fig. 3.5. Note that we have now switched to a linear scale on the y-axis. The contribution of satellite galaxies is now much less important than the impression gained from the HOD plot. Next, in the bottom row of Fig. 3.5 we plot, as a function of halo mass, the HOD multiplied by the halo mass function and the bias factor. Again, a linear scale is used for the y-axis. The area under the black curve in this case gives the effective bias of the galaxy sample. The satellites make a larger contribution to the effective bias than they do to the number density. This is because the satellites are preferentially found in high mass haloes which have large bias factors.

3.5 An empirical solution to the problem of luminosity dependent clustering

In this section we find an empirical solution to the problem of matching the observed luminosity dependence of clustering. We do this by changing the HOD of the Bower et al. model by hand. We could equally well have chosen to use the Font et al. model and would have reached similar conclusions. We saw in the previous section that the HOD for central galaxies has a complicated shape which is not well described by the standard HOD parametrizations. This is, in part, due to the physics invoked in the models and to the use of differential rather than cumulative luminosity bins. The satellite galaxy HOD, on the other hand, has a more straightforward power law form, $N_{\rm sat} \propto M_{\rm halo}^{\alpha}$, where $M_{\rm halo}$ is the host halo mass. Moreover, we saw in the previous section that the Durham models have more satellite galaxies than the Munich model and that this could be the reason behind their poorer match to the

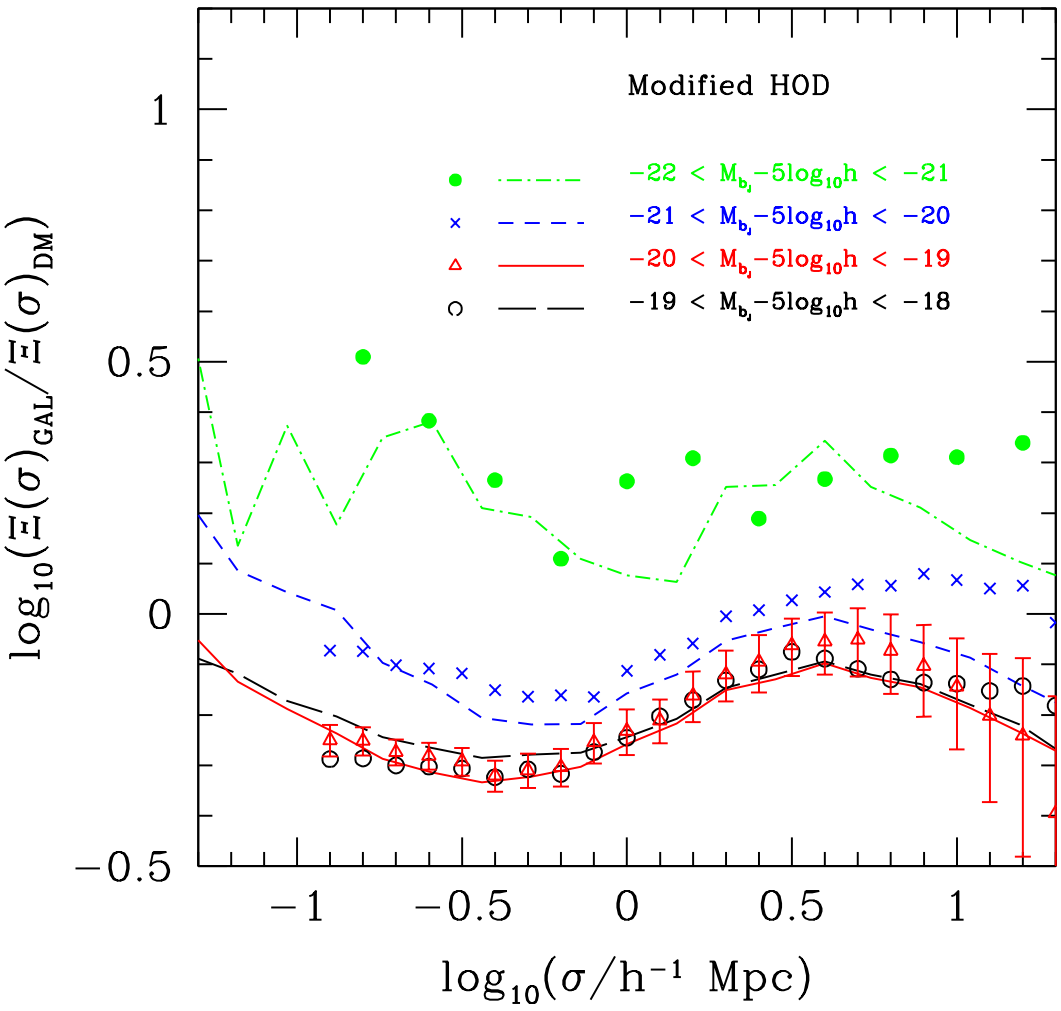

Figure 3.7: The clustering of galaxies after modifying the HOD of the Bower et al. model (lines) compared to the 2dFGRS data (points). We plot the projected correlation function divided by an analytic estimate of the nonlinear projected correlation function of the dark matter in the Millennium simulation cosmology. Different colours and lines show the results for different luminosity bins as indicated by the key.

observed clustering. Here, we establish how the satellite HOD must be changed in order to match the 2dFGRS results better. This will help guide an investigation into changing the physics of the galaxy formation model which is carried out in the next section.

The satellite HOD for the Bower et al. model plotted in Fig. 3.5 has a power law form with slope $\alpha \sim 1$ in each of the luminosity bins. We note that the same slope is generally found for other galaxy selections, such as luminous red galaxies (Almeida et al. 2008; Wake et al. 2008).

The starting point to make a realization of galaxy clustering is the DHalo[3] catalogue of dark matter halo masses and positions constructed from the Millennium simulation (Harker et al. 2006). This is the halo catalogue used in the `GALFORM` model and is somewhat different from the list of haloes generated by the friends-of-friends group finding algorithm. The DHalo catalogue is constructed with reference to the merger histories of the dark matter haloes. In the case of a friends-of-friends merger history, it is possible, occasionally, for the mass of a halo to decrease with increasing time. This happens, for example, when two haloes are either extremely close or overlap to some extent at one timestep, but move apart and are identified as separate haloes at a subsequent output time. The DHalo algorithm "looks ahead" to check if haloes merged by the group finder at one output time stay merged at the next two outputs.

Keeping the same mass at which the mean number of satellites per halo reaches unity as predicted by the fiducial Bower et al. model, we allow the slope of the satellite HOD to vary for each magnitude bin in order to obtain a better match to the 2dFGRS clustering data. The number of galaxies as a function of halo mass is assumed to have a Poisson distribution for $\langle N \rangle > 1$. Semianalytic models and hydrodynamic simulations predict a significantly sub-Poisson $P(N|M)$ distribution at low $\langle N \rangle$ and Poisson distribution at $\langle N \rangle$ >1 (Berlind et al. 2003). We check the distribution of generate galaxies using HOD formalism. The value which indicate the Poisson distribution, $\alpha_M^2 = \langle N(N-1) \rangle / \langle N \rangle^2$, nearly 1 from higher halo mass

[3]http://galaxy-catalogue.dur.ac.uk:8080/Millennium/

than $10^{11.5}h^{-1}M_{\odot}$. For halo masses for which the HOD predicts $\langle N \rangle < 1$, a fraction of haloes is populated with a satellite galaxy at random: i.e. if the random number chosen from a uniform distribution between zero and one, $x < N$, then the halo is assigned a satellite, otherwise it has no satellite. We have tested that this procedure can reproduce the clustering in the Bower et al. model when the Bower et al. HOD is used.

The modified HOD derived as described above is shown by the solid lines in Fig. 3.6. The HOD of the original Bower et al. model is shown by the dashed lines in this plot. In the three faintest luminosity bins, the slope of the modified satellite HOD is shallower than the original i.e. $\alpha < 1$, corresponding to a reduction in the number of satellites in massive haloes. The change in slope is largest in the faintest bin. In the brightest luminosity bin, the trend is reversed and there are slightly more satellites in massive haloes in the modified HOD. By reducing the number of satellites in high mass haloes, two effects are generated in the correlation function. The effective asymptotic bias of the sample is reduced, due to a smaller two-halo clustering term. Also, the one-halo term is suppressed, reducing clustering on small scales, as there are fewer pairs of galaxies within massive haloes. By contrast with the modified HOD, as we remarked upon above, the HOD of the Bower et al. model exhibits the same value of the slope of the satellites in each luminosity bin.

Fig. 3.7 shows that the trend of clustering strength with luminosity displayed by the modified HOD matches that of the 2dFGRS data. Furthermore, the improved level of agreement is seen on both large and small scales. The matching of the asymptotic bias on large scales and the shape of the correlation function on small scales is convincing evidence in support of the modified HOD having the correct number of satellite galaxies in haloes of different masses. The challenge now is to see if the semi-analytical model can reproduce the form of the modified HOD, either by further exploration of the model parameter space or by adding new physical processes.

3.6 Implications for satellite galaxies in galaxy formation models

In the previous section we demonstrated that the clustering properties of the Bower et al. model can be significantly improved if the number of satellite galaxies in massive haloes is reduced. This was achieved by changing the HOD of the Bower et al. model by hand. The clustering predictions subsequently changed on all scales (in HOD terminology, both the one and two halo contributions were changed) to improve the match with the 2dFGRS measurements, which can only be achieved by changing the number of satellites. In this section we try to reproduce the modified HOD in a physical, rather than empirical, way by using the `GALFORM` model.

The first approach we tried was to run variants of the Bower et al. model in which selected parameters were perturbed from their fiducial values. In particular, we varied parameters which we thought would have an impact on the relation between galaxy luminosity and host halo mass, as plotted in Fig. 3.4. These included the strength of supernova feedback, the degree of suppression of gas cooling in massive haloes due to AGN heating and the timescale for galaxy mergers. In the case of each of these variant models, we rescaled the model galaxy luminosities to agree exactly with the observational estimate of the luminosity function from Norberg et al. (2009). The clustering predictions in the variants were different to those of the original Bower et al. model. However, none was able to match the observed clustering. Intriguingly, the slope of the satellite HOD was $\alpha \approx 1$ in all of the models, that is none of the parameter variations was able to change the slope of the satellite HOD in the way suggested by the modified HOD.

The second approach we tried was to change the timescale for galaxies to merge due to dynamical friction. `GALFORM` uses a modified version of the timescale given by the dynamical friction formula of Chandrasekhar (1943; see eqn. 4.16 of Cole et al. 2000). We experimented with adjusting this timescale by allowing an extra scaling based on the ratio of the host halo mass to the mass of the satellite, $M_{\rm H}/M_{\rm sat}$. To solve the problem of too many satellites we needed to reduce the merger timescale for $M_{\rm H}/M_{\rm sat} > 1$. Recent numerical studies of satellite mergers found that the

Chandrasekhar formula needs to be revised but in the opposite sense, i.e. with a somewhat longer merger timescale for objects with $M_{\rm H}/M_{\rm sat} > 1$ (Jiang et al. 2008, 2009). Hence this approach, although viable, was abandoned as requiring an unrealistic change to the prescription for calculating the timescale for galaxy mergers.

In this section, we explore the incorporation of two physical processes into the GALFORM semi-analytical model: the tidal disruption or stripping of mass from satellite galaxies and mergers between satellites. The implementations presented here are exploratory and are meant to give an indication of the likely impact of the new physics on the model predictions. If the changes turn out to be promising, the intention is that this should motivate future, fully self-consistent revisions to the GALFORM machinery.

3.6.1 The dissolution of satellite galaxies

Galaxy clusters contain a diffuse background of light, the intracluster light (ICL), which is not associated with any particular galaxy (e.g. Welch & Sastry 1971). The ICL is thought to result from the disruption of small galaxies and the stripping of stars from larger ones. The measurement of the intracluster light is challenging. Current estimates put the ICL in the range of 5-30% of the total cluster light (Zibetti et al. 2005; Krick & Bernstein 2007; Zibetti 2008).

A number of physical processes could be responsible for the removal of stars from satellite galaxies e.g. tides produced by the cluster potential and successive high speed fly-by encounters between cluster members (Richstone 1976; Aguilar & White 1985). A full treatment of these effects would require a dynamical simulation (e.g. Moore et al. 1996; Gnedin 2003). Attempts have been made to implement analytic descriptions of the phenomena modelled in the simulations into galaxy formation models (e.g. Taylor & Babul 2001; Benson et al. 2004; Yang et al. 2009).

In general, standard semi-analytical galaxy formation codes ignore the tidal disruption of satellite galaxies. A recent exception is the calculation of Henriques, Bertone & Thomas (2008). These authors post-processed the output of the Munich group's semi-analytical model to remove galaxies that they believed should have

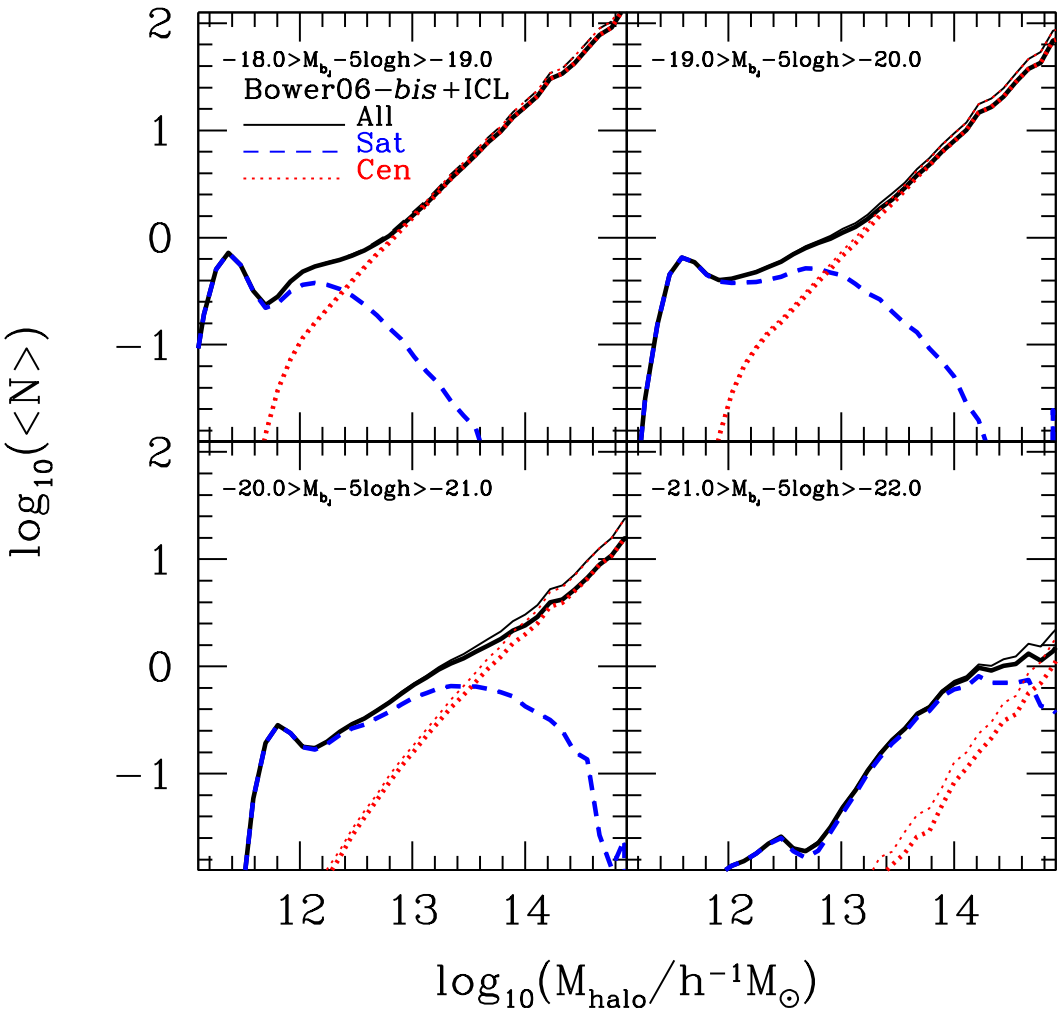

Figure 3.8: The HOD after applying the satellite disruption model of Eq. 3.2 (thick lines). The starting point is the HOD of the Bower06-*bis* model shown by the lines (normal width). Each panel corresponds to a different luminosity bin as indicated by the key.

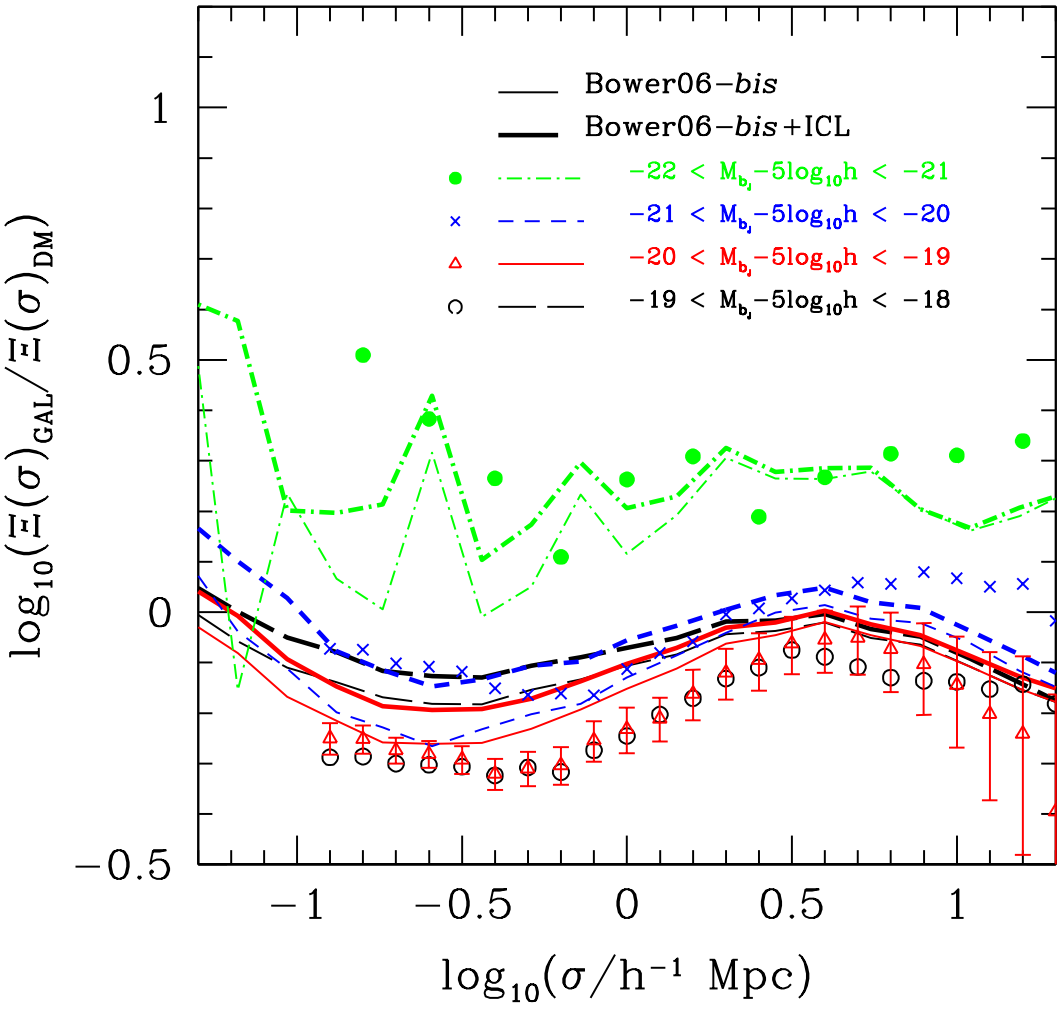

Figure 3.9: The projected correlation function for galaxy samples of different luminosity divided by the dark matter projected correlation function for the Millennium simulation cosmology. The normal width lines show the predictions of the Bower et al. (re-run) model and the thick lines show this model after applying the satellite disruption model of Eq. 3.2. The symbols show the clustering data measured from the 2dFGRS.

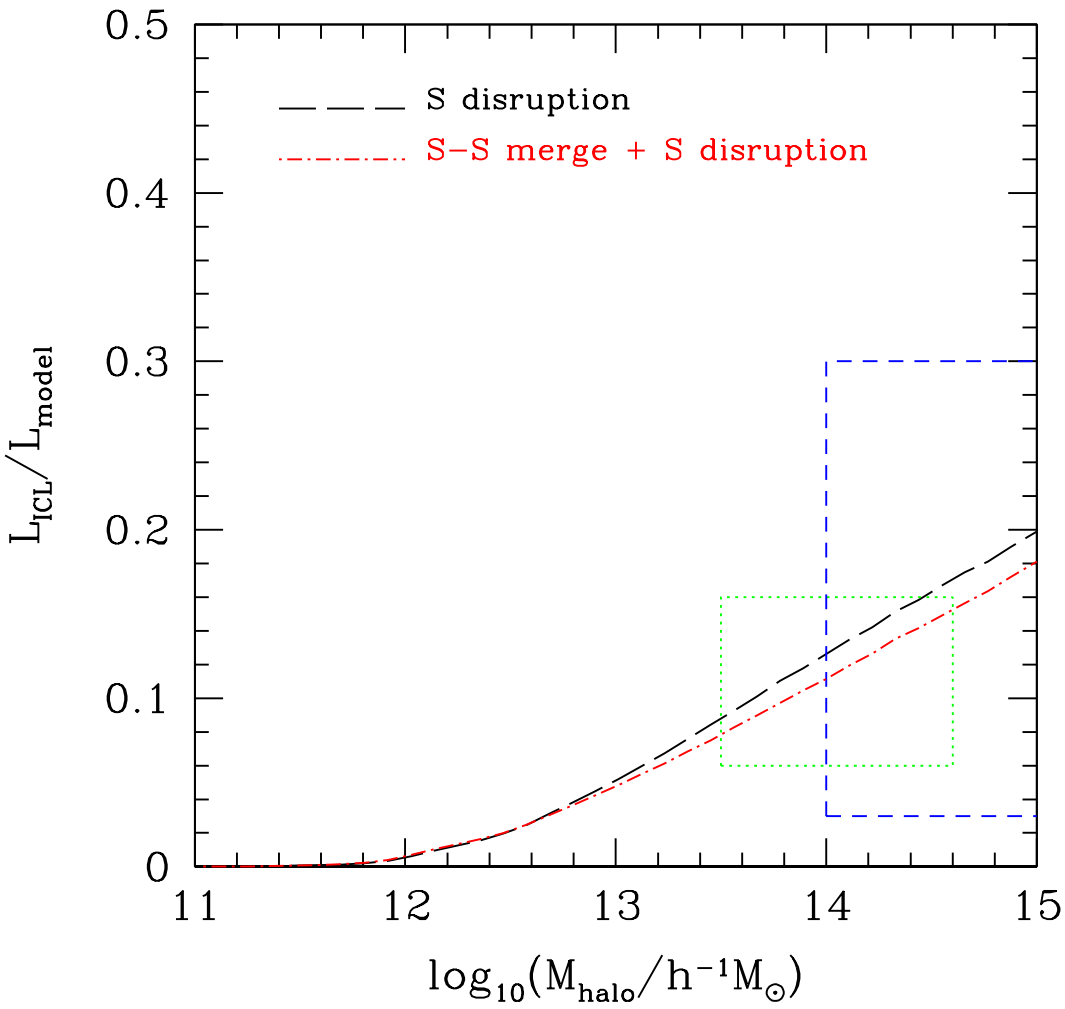

Figure 3.10: The intracluster light as a function of halo mass in the satellite disruption model. The y-axis shows the fraction of the total cluster light which is attached to galaxies. The green dotted box shows the observational estimate of the intracluster light from Zibetti (2008) and the blue dashed box shows the result from Krick & Bernstein (2007). The red dot-dashed line shows the intracluster light predicted by the model with satellite disruption alone (as discussed in Section 3.6.1); the black long dashed line shows a model with satellite-satellite mergers (Section 3.6.2) and disruption of satellites. This hybrid model is discussed in Section 3.6.3.

been tidally disrupted. Galaxies are associated with the dark matter halo in which they first formed as a central galaxy. When this halo merges with a more massive halo, it becomes a satellite halo or substructure, and is stripped of mass through dynamical effects. Eventually, the substructure may fall below the resolution limit of the N-body simulation (in this case the Millennium Simulation). Henriques et al. removed satellites whose host dark matter substructure had dissolved, and added these to the ICL. They found that by adopting this procedure, the model predictions agreed better with the slope of the faint end of the luminosity function and the colour distribution of galaxies. However, this algorithm depends on the resolution of the N-body simulation, which governs when subhalos are destroyed. Moreover, the softening length adopted in the simulation exceeds the scale size of all but the very brightest galaxies. Hence, it is not clear that any of the more condensed baryonic material would have been stripped from the model galaxies, even when the host dark matter halo has been shredded.

Here we adopt a simpler approach which is independent of the resolution of the N-body simulation. We assume that the degree of disruption of a satellite galaxy depends on the ratio of the mass of the main dark matter halo to the mass of the satellite halo at infall, $M_{\rm H}/M_{\rm sat}$:

$$\frac{L_{\rm new}}{L_{\rm orig}} = \beta \left(\frac{M_{\rm H}}{M_{\rm sat}} \right)^{-1}, \tag{3.2}$$

where $L_{\rm orig}$ is the original luminosity of the satellite galaxy predicted by the galaxy formation model, $L_{\rm new}$ is the new luminosity intended to take into account stripping of mass from the satellite and β is an adjustable parameter. We chose this scaling of disrupted luminosity fraction because the galaxy merger timescale essentially scales with the mass ratio $M_{\rm H}/M_{\rm sat}$; objects with large values of $M_{\rm H}/M_{\rm sat}$ will spend longer orbiting within the host dark matter halo and are therefore more susceptible to dynamical disruption. Our satellite disruption prescription involves post-processing the output of the galaxy formation model, to reduce the luminosity of satellite galaxies according to Eq. 3.2. One clear shortcoming of our approach is that we do not take into account the time when the satellite galaxy actually fell into the more massive halo. With our prescription, a satellite could suffer a large luminosity reduction immediately after falling into a larger structure. On the other hand, we

ignore any stripping which may have occurred at earlier stages in the merger history. Hence it is not clear whether our simple model for the disruption of satellites is likely to be an over or underestimate of the actual effect.

The Millennium Archive does not list the satellite galaxy dark halo mass for the Bower et al. model. Hence, it was necessary for us to re-run the Bower et al. model in order to extract the information required to apply the model described by Eq. 3.2. We present the results of rerunning the Bower et al. model, labelled Bower06-it bis, *without* applying any dynamical disruption, in Fig. 3.8 in which we show the HOD and in Fig. 3.9, where we compare the predicted clustering with the 2dFGRS measurements. A comparison of the results presented in these plots with the equivalent results for the version of the Bower et al. model available from the Millennium Archive (Figs. 3.5 and 3.3 respectively) shows a subtle but appreciable change in the model predictions. The re-run version of the Bower et al., which we refer to as Bower06-*bis*, is actually in better agreement with the 2dFGRS clustering results than the Millennium Archive version. The main reason for these differences are small improvements in the model. There has been substantial code development in the three years since the Bower et al. model was originally placed in the Millennium Archive, to incorporate new physical ingredients and to improve the implementation of other processes. Also, improvements have been made to the construction of the dark matter halo merger histories from the Millennium (J. Helly, private communication). The re-run Bower et al. model is available in the Millennium archive as Bower06-*bis*. As we shall see, the changes to the clustering predictions arising from the implementation of new physical processes are, in any case, larger than those between Bower et al. and Bower06-*bis*.

The HOD resulting from applying the satellite disruption model of Eq. 3.2 is compared with the Bower06-*bis* model in Fig. 3.8. The free parameter β in the stripping model was set to 0.9 to produce the best match to the clustering measurements, as plotted in Fig. 3.9. As expected, Fig. 3.8 shows that there are fewer satellites in the model with disrupted satellites. The effect appears largest in the brightest luminosity bin. This is primarily due to the imposed change in the shape of the luminosity function, rather than to a shift in the typical value of $M_{\rm H}/M_{\rm sat}$ for

each galaxy sample. In the brightest bin, since the abundance of galaxies drops exponentially with luminosity, more galaxies are shifted out of the bin in the faintwards direction, after applying the disruption recipe, than are shifted into that bin from brighter luminosities. The change in the HOD generated by applying the satellite disruption model falls short of the target suggested by the modified HOD derived in the previous section. In the intracluster light model, the slope of the satellite HOD is essentially unchanged and the biggest variation in the number of satellites is found in the brightest luminosity bin rather than the faintest. The resulting clustering predictions do not change in the desried way, as shown by Fig. 3.9. Rather than altering the luminosity dependence of clustering, the main effect of disrupting satellites is to reduce the clustering amplitude in all the luminosity bins.

We close this section by showing the model prediction for the fraction of the total light in a cluster that is in the form of a diffuse intergalactic background. Fig. 3.10 shows that the satellite disruption model removes at most 20% of the total cluster light from galaxies, in excellent agreement with the observational estimate from Zibetti (2008). This agreement is encouraging as the parameter in the satellite disruption model was set without reference to the constraint on the background light, but was chosen to improve the match to the observed clustering.

The mass loss on satellites is caused by the tidal forces. The material outside tidal radius of satellite galaxy will be stripped. Assuming a satellite in a circular orbit with spherically symmetric mass distribution, the tidal radius can be identified as the radius in the satellite galaxy beyond which tidal forces become important can be made by equating the tidal force to the self-gravity of the satellite galaxy (King 1962; Binney & Tremaine 1987). Tayor & Babul (2001) described the tidal radius r_t as

$$r_t \approx \left(\frac{G m_{sat}}{\omega_{sat}^2 - \mathrm{d}^2\phi/\mathrm{d}r^2} \right)^{1/3}, \tag{3.3}$$

where ω_{sat} is the orbital angular velocity of the satellite and ϕ is the potential of the halo. Henriques et al. (2010) calculate the mass loss of satellite galaxy using a semi-analytic model adding the effect of tidal disruption described by Eq. 3.3. The median mass loss of satellite galaxies in the $10^{13}h^{-1}M_{\odot}$ mass halo is $\sim 20\%$ in Henriques et al. (2010). This value is similar with our result using simple tidal

dissolution model Eq. 3.2

3.6.2 Mergers between satellite galaxies

Semi-analytical models typically only consider the merger of satellites with the central galaxy in a halo. In general, a timescale is calculated analytically for the orbit of the satellite to decay due to dynamical friction. If this timescale is shorter than the lifetime of the host dark matter halo, then the satellite is assumed to merge with the central galaxy. When a halo merges with a larger structure, the galaxies in the smaller halo are assumed to become satellite galaxies orbiting the new central galaxy. The satellites retain no memory of the fact that they were once members of a common halo. New dynamical friction timescales are calculated for each satellite.

With the advent of ultra-high resolution N-body simulations, there is now convincing evidence that this simple picture is incomplete (Springel et al. 2008; Angulo et al. 2008; Wetzel, Cohn & White 2009). The simulations reveal that, following a merger, the subhaloes of the lower mass halo often remain as a distinct unit, orbiting coherently in the new main subhalo. Indeed, several levels of subhalo hierarchy have been uncovered. By tracing the evolution of the subhaloes in these simulations, their ultimate fate can be determined. A large fraction of the high mass subhaloes which undergo a merger coalesce with the main subhalo of the new halo. However, the probability of a merger with a subhalo other than the main subhalo increases with decreasing subhalo mass. At $z = 0$, Angulo et al. (2008) found that subhaloes with 1% or less of the total mass of the main subhalo were as likely to merge with another subhalo as with the main subhalo. Rather than merging with a random subhalo, the merger is with another subhalo which shared a common parent halo. A merger which started before this parent halo was subsumed by the main halo is being completed inside the new halo.

We added satellite-satellite mergers to `GALFORM` by modifying the prescription for galaxy mergers. Guided by the results obtained by Angulo et al. for the Millennium Simulation, we modified the calculation of the galaxy merger timescale. Depending on the mass ratio, $M_{\rm H}/M_{\rm sat}$, and the redshift, we allowed a fraction of satellite galaxies to be considered for satellite-satellite mergers (see figure 5 of Angulo et al.).

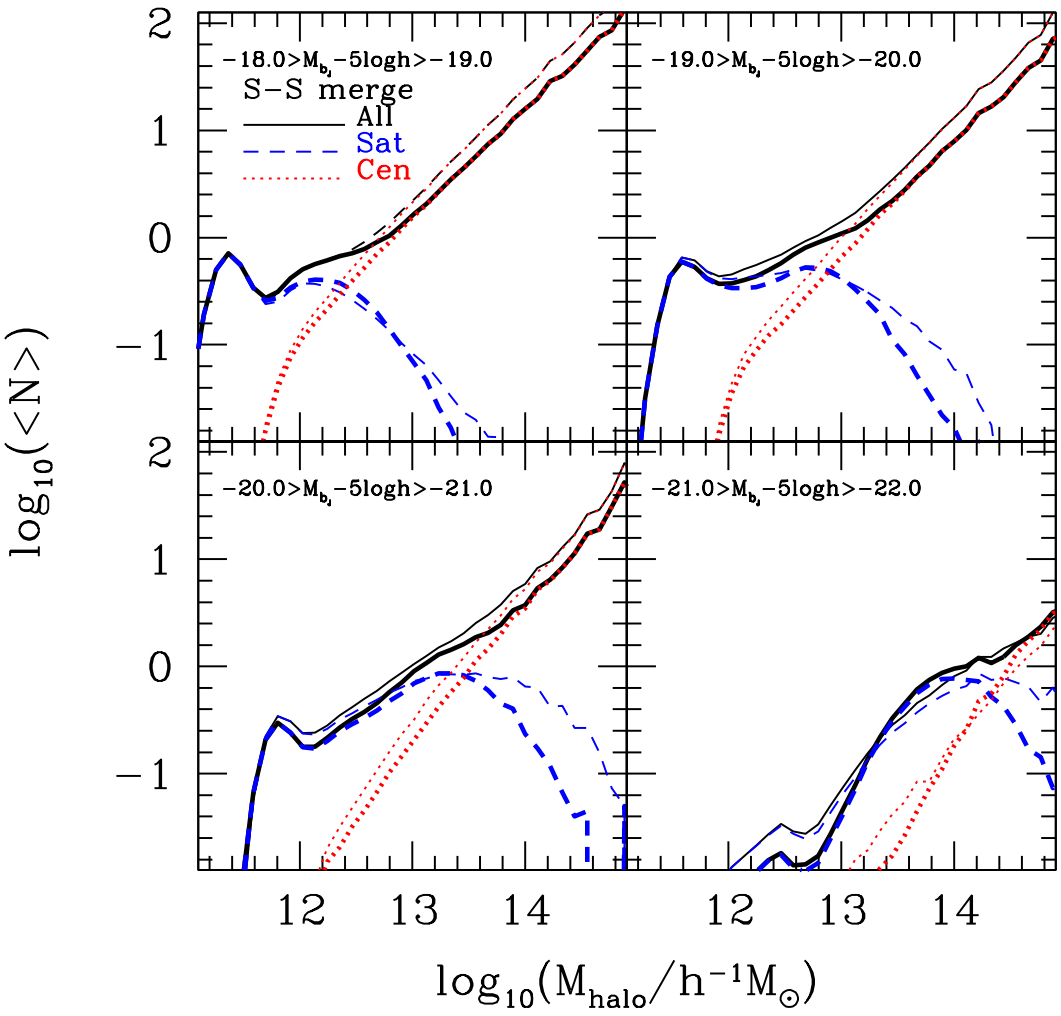

Figure 3.11: The HOD of the model including satellite-satellite mergers (thick lines). For reference, the HOD of the Bower06-*bis* model is shown by the normal width lines. The values of the power-law slope α of the satellite HOD are now different in each luminosity bin.

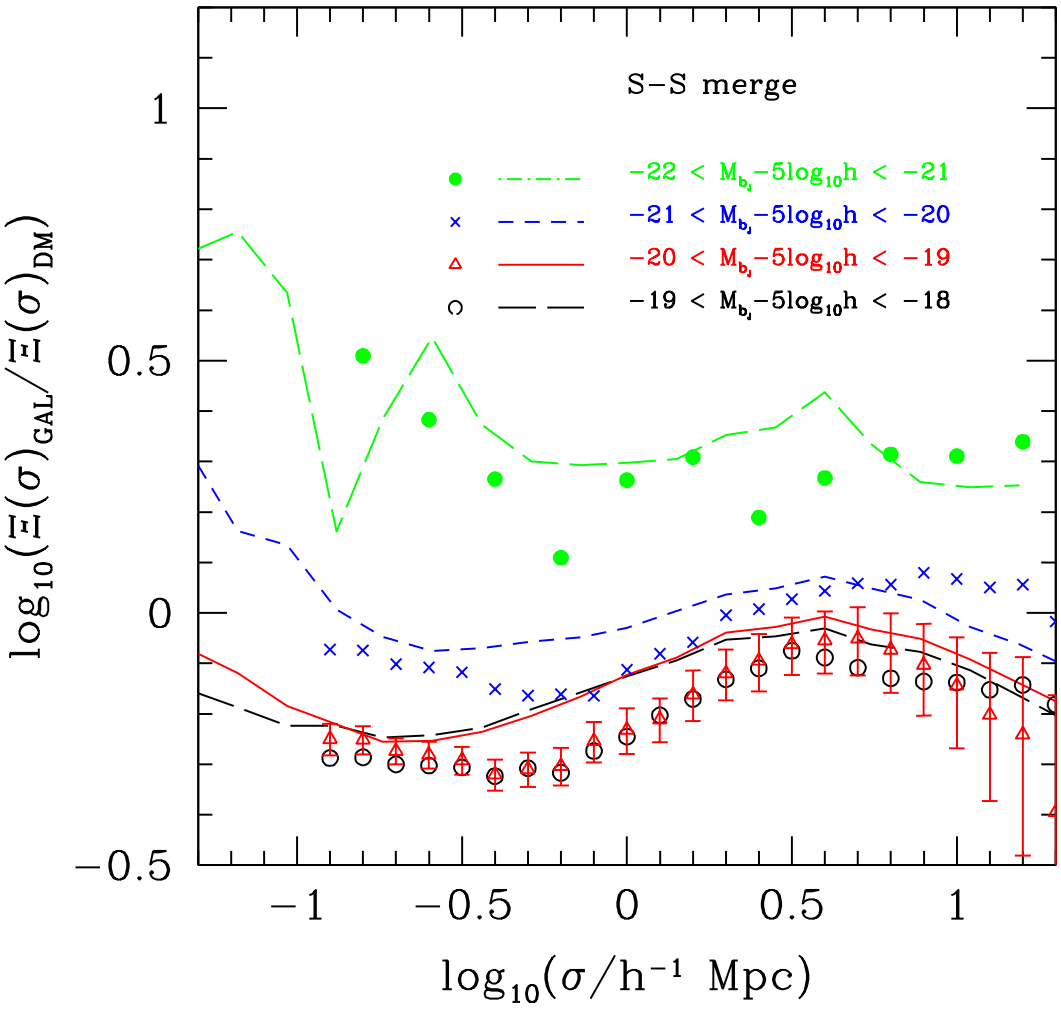

Figure 3.12: The projected correlation functions for galaxies divided by the projected correlation function of the dark matter for the model with satellite-satellite mergers. The symbols show the 2dFGRS measurements. The different colours and lines show the different luminosity bins.

We did this by considering the last but one level of the halo merger history i.e. the progenitor haloes of the present day halo. For a selected satellite in the progenitor halo, we asked if there would be sufficient time for this object to have merged with the central galaxy in the progenitor *by the present day*, rather than by the end of the lifetime of the progenitor. This is equivalent to allowing the merger to continue in the substructure after it becomes part of the larger halo. If there is sufficient time, then we merge the satellite with the central galaxy of the progenitor at the end of the progenitor's lifetime. This means that the merger happens sooner than it would do in practice. If there is a burst of star formation associated with the merger, then this burst will also happen earlier than it should have done. However, in the Bower et al. model there is relatively little star formation in bursts at low redshift. Our scheme does, however, reproduce the number of satellite-satellite mergers implied by the subhalo mergers in the Millennium Simulation.

By allowing satellite-satellite mergers, we are able qualitatively to reproduce the changes suggested by the empirically determined modified HOD, as shown in Fig. 3.11. There are two main reasons for the change in the HOD. Firstly, satellite-satellite mergers reduce the number of satellite galaxies in the model. Secondly, the number of low luminosity satellite galaxies in high mass haloes is reduced because these objects can merge with other satellites; the remnant is also a satellite but it is, of course, brighter than its progenitors. The HOD for central galaxies also changes, with the central galaxies in more massive haloes becoming brighter (and hence moving into a brighter luminosity bin). This is because satellites which have experienced satellite-satellite mergers are more massive than they would have otherwise been and therefore have a shorter dynamical friction timescale. The clustering predictions for the model with satellite-satellite mergers are shown in Fig. 3.12. The model now matches the sequence of luminosity dependent clustering measured in the 2dFGRS, albeit with slightly higher clustering amplitudes overall.

3.6.3 The kitchen sink model

In the previous two subsections we have seen that the satellite disruption and satellite-satellite merger models have appealing features. The satellite disruption

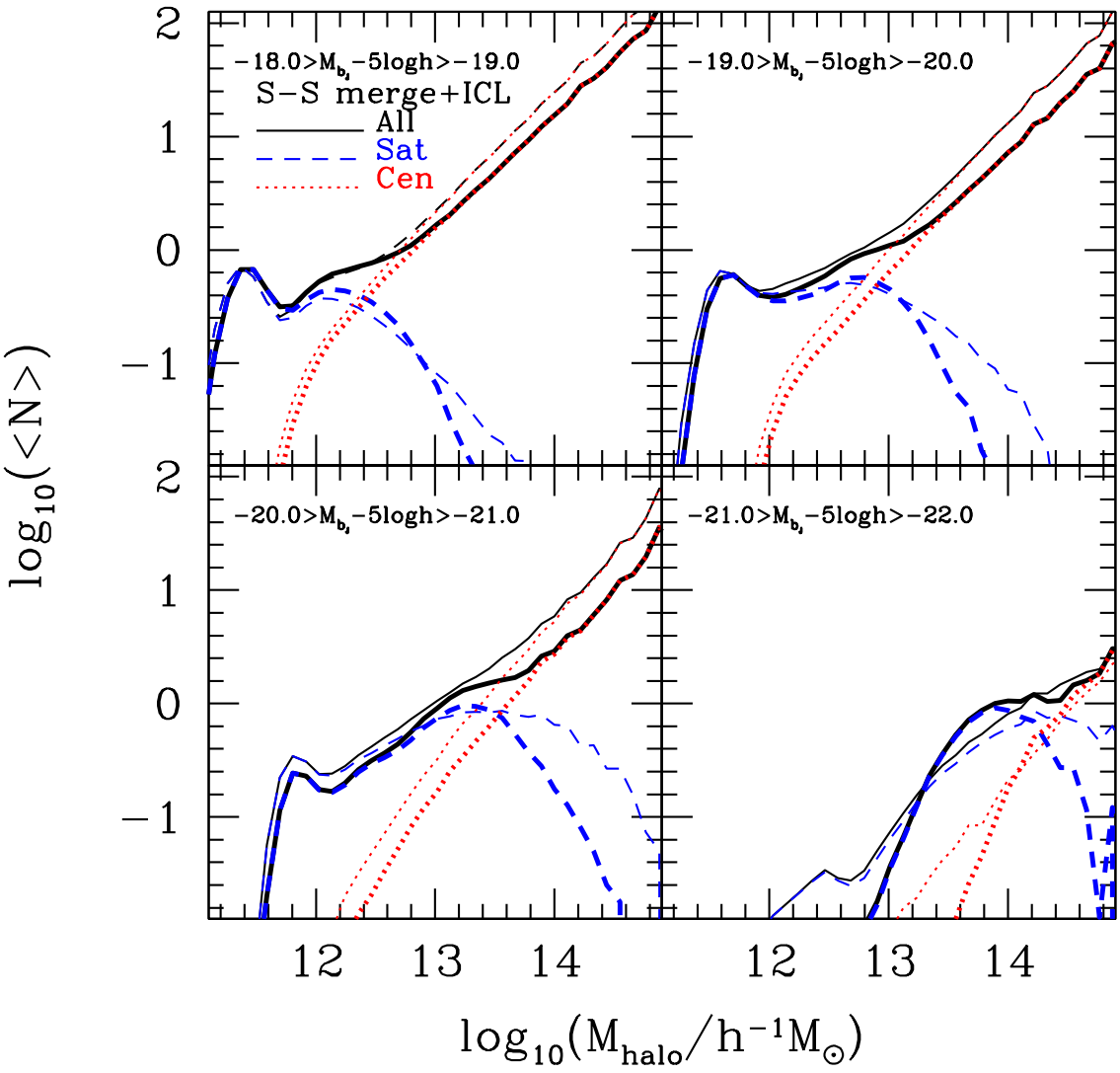

Figure 3.13: The HOD of the hybrid model with satellite-satellite mergers and disruption of satellites (thick lines). The Bower et al. model HOD is shown by the normal width lines.

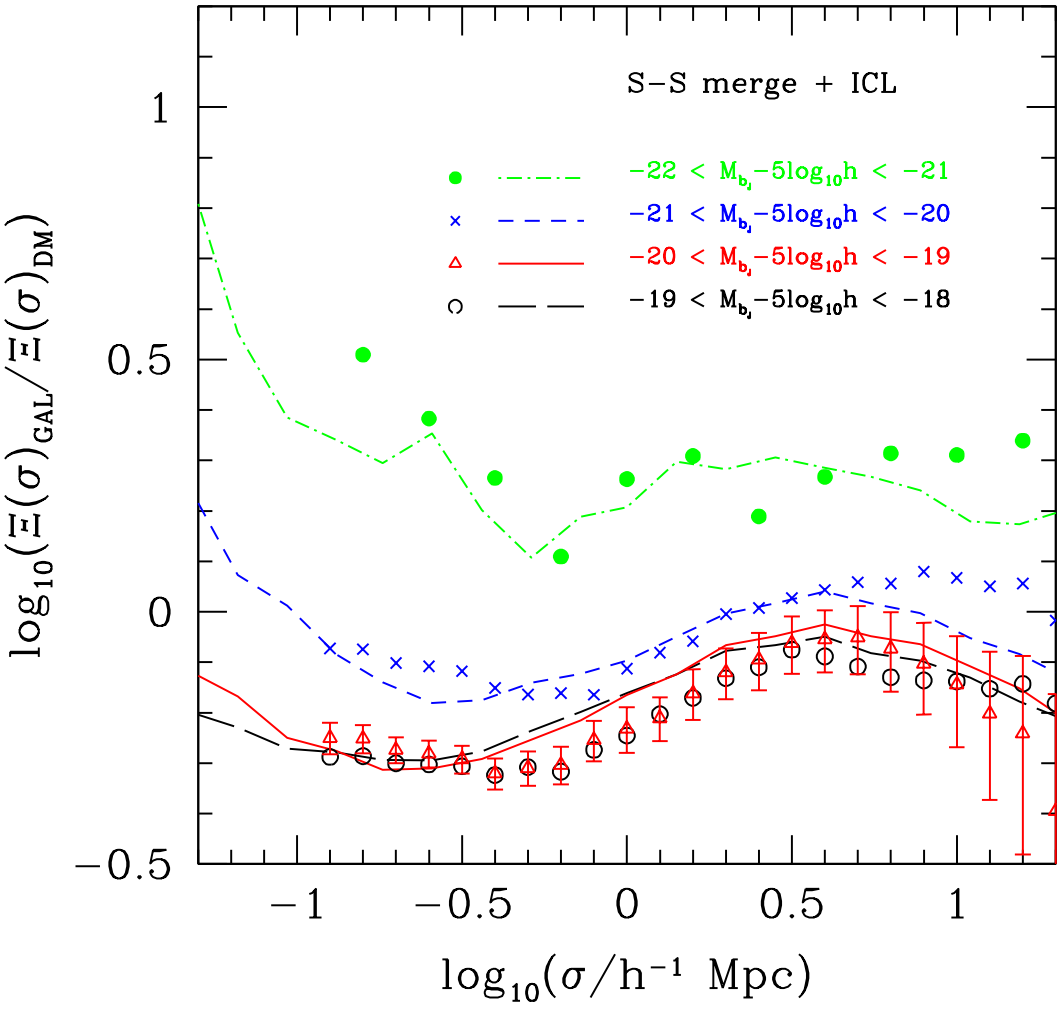

Figure 3.14: The projected correlation functions divided by correlation function of the dark matter. The lines show the predictions for the hybrid satellite-satellite merger and satellite disruption model. The symbols show the 2dFGRS measurements.

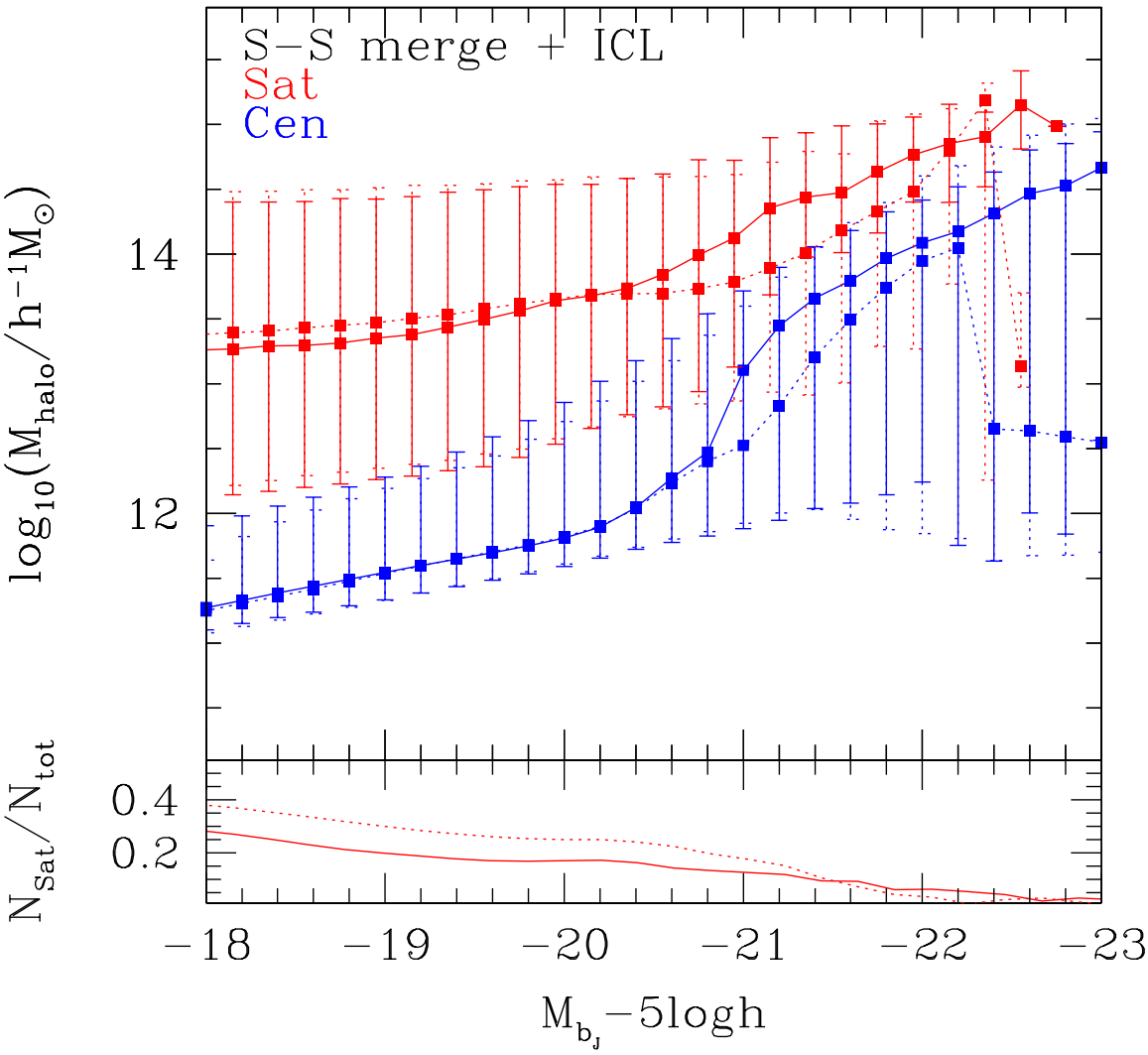

Figure 3.15: The host halo mass - luminosity relation for the hybrid model. The upper panel shows the median halo mass and the 10-90 percentile range. The red points show the relation for satellite galaxies and the blue lines for central galaxies. The lower panel shows the fraction of galaxies which are satellites as a function of magnitude. The dotted lines in both panels show the relations for the original Bower et al. (2006) model.

model can change the overall amplitude of the clustering for different luminosity samples, whereas the satellite-satellite merger model can reproduce the observed trend of clustering strength with luminosity if not the precise amplitude. In isolation, neither model offers a fully satisfactory solution to the problem of matching the luminosity dependent clustering seen in the 2dFGRS. It seems desirable therefore to implement both effects in tandem. We do this by generating a model which incorporates satellite-satellite mergers and post-processing the resulting satellite luminosities using the disruption model of Eq. 3.2.

Fig. 3.14 shows the projected correlation functions predicted by the hybrid model. The model predictions are now in remarkably good agreement with the 2dFGRS measurements. The model matches the amplitude of clustering, the trend and strength of the luminosity dependence of clustering and the shape of the correlation functions. The HOD of this model matches the form of the reference empirical HOD as shown in Fig. 3.13. The slope of the satellite HOD in the hybrid model is influenced by satellite-satellite mergers, whereas its amplitude is determined by satellite disruption.

Fig. 3.10 shows how the predicted intracluster light in the hybrid model with satellite disruption and satellite-satellite mergers compares with the Bower06-*bis* model. Again, the amount by which the plotted halo luminosity ratio deviates from unity shows the fraction of the total light is not attached to galaxies. The fraction of intracluster light depends on halo mass and is in very good agreement with the observational estimates by Zibetti (2008).

Fig. 3.15 shows the relation between host halo mass and galaxy luminosity in the hybrid model. Compared with the Bower et al. and Font et al. models, there is relatively little difference in the median halo mass for either satellite or central galaxies; the changes in the median mass are of the order of 0.1dex. However, the host halo masses of satellite galaxies are large and thus these haloes are highly biased. A small change in the typical host mass will therefore produce an appreciable change in the predicted bias. The key difference is in the fraction of galaxies that are satellites as a function of magnitude, shown in the lower panel of Fig. 3.15. The number of satellites in the hybrid model is down by almost a factor of two from that

in the original Durham models.

3.7 Summary and Conclusions

The dependence of galaxy clustering on luminosity has been measured with high accuracy in the local Universe by the 2dFGRS and SDSS (Norberg et al. 2001, 2002; Zehavi et al. 2002, 2005; Jing & Borner 2004; Li et al. 2006). We have shown that the current "best bet" publicly available galaxy formation models only match the observational results in a qualitative sense. These models fail to match the trend of clustering strength with luminosity. We have demonstrated that the reason for the discrepancy is that the models predict too many satellites in massive haloes. Li et al. (2007) reached a similar conclusion comparing the clustering of galaxies in the red selected SDSS with the semi-analytical models of Kang et al. (2005) and Croton et al. (2006).

One potential way to improve the agreement between the model predictions and the observations, particularly for the Durham models, would be to change the background cosmology. In particular, a reduction in the amplitude of density fluctuations to $\sigma_8 \sim 0.75$ has been suggested as a means to match the observed clustering (Yang et al. 2004). We investigated this using a small volume simulation. Whilst the slope of the correlation function on small scales was less steep than it is for $\sigma_8 = 0.9$, this change did not produce luminosity dependent clustering. Moreover, such a low value of the fluctuation amplitude is now disfavoured by the latest cosmological constraints (e.g. Sanchez et al. 2009).

Li et al. (2007) showed that the match to the observed clustering could be improved if $\approx 30\%$ of the satellite galaxies were removed from the catalogues generated from the semi-analytical models. Li et al. did this by hand without any reference to the mass of the host dark matter halo. This is equivalent to changing the normalization of the halo occupation distribution for satellites, without altering the slope. In this chapter, we first changed the HOD of satellites by hand and found that the agreement with the observed clustering could be improved by changing the slope of the satellite HOD. For galaxy samples close to L_*, satellites have to be preferentially

removed from more massive dark matter haloes.

Out of the original semi-analytical models we considered in this chapter, the de Lucia & Blaizot (2007) model came closest to reproducing the 2dFGRS clustering measurements. This model predicted roughly the correct overall clustering amplitude but failed to show *any* dependence of clustering strength on luminosity except for the very brighest galaxies. This was also the model with the smallest number of satellites. However, the origin of this paucity of satellites is unclear, and it could well be due to approximations used in the model rather than to different physical ingredients. For example, one approximation used in the De Lucia & Blaizot model that is cleary crude is the adoption of a fixed cooling time for all haloes of a given mass (in the Durham models the cooling time depends on the halo lifetime which is set by the merger history). This could lead to a difference in the mass distribution of galaxies between the models, which would in turn alter the time taken for them to merge due to dynamical friciton after they become satellites in a more massive halo. In any case there is a clear need to reduce the number of satellites in *all* the models in order to improve the match with the 2dFGRS clustering measurements.

We next tried to remove satellite galaxies from massive haloes in the Durham semi-analytical models by perturbing the values of the parameters which control certain processes, such as supernova feedback, the suppression of gas cooling by AGN heating and galaxy mergers. When running a variant model, the predicted luminosity function often changes. To ensure that changes in the clustering predictions were robust to the requirement that a model should reproduce the observed galaxy luminosity function, we rescaled the model luminosity functions to agree exactly with the observations. We were unable to find an improved model within the existing framework, which suggests that additional physical processes which mostly affect satellite galaxies need to be considered.

The Durham models have recently been revised as regards the treatment of gas cooling in satellites (Font et al. 2008). Satellite galaxies can now retain some fraction of the hot halo associated with them at infall. The precise fraction depends upon the orbit of the satellite. This improvement of the gas cooling treatment alters the colours of faint satellites in groups and clusters. The galaxies we consider in

this chapter are brighter by comparison and there is little change in the clustering predictions of the Font et al. model compared with those from its predecessor, the Bower et al. (2006) model.

In this chapter, we considered two processes which are not currently included in most galaxy formation models: mergers between satellite galaxies and the tidal disruption of satellites. The first of these processes is motivated by recent high resolution simulations of the formation of dark matter haloes which show that hierarchies of substructures persist (Diemand et al. 2008; Springel et al. 2008). Mergers which started in a progenitor halo can run to completion in the descendant halo. The disruption of satellites has been modelled analytically in the Durham model in a study of the heating of the Milky Way's disk (Benson et al. 2004). Here, we applied a simple prescription to remove luminosity from satellites based on the ratio of the host halo mass to the mass of the halo in which the satellite formed, which is related to the timescale for the satellite's orbit to decay through dynamical friction. Applying the model for the disruption of satellites changes the overall amplitude of clustering without improving the trend of clustering strength with luminosity. Including mergers between satellites, on the other hand, does alter the predictions for the luminosity dependence of clustering. By applying both extensions together, we are able to obtain a significantly improved match to the 2dFGRS measurements (Norberg et al. 2009). The hybrid model matches the observational constraints on the amount of intracluster light.

The differences between the clustering predictions of current galaxy formation models and observations are small. However, the differences can be measured robustly and will become even more apparent when larger surveys become available. These discrepancies limit the usefulness of the models in the construction of mock catalogues needed for the exploitation of future galaxy surveys and suggest the need for new physical processes to be incorporated into the models. The revisions to the galaxy formation models we propose in this chapter are simplistic and are merely intended to highlight promising areas where the models need to be developed in the future, in a self consistent way.

Chapter 4

The spatial distribution of cold gas in hierarchical galaxy formation models

4.1 Introduction

Cold gas is central to galaxy formation yet little is known about how much there is in the Universe at different epochs and how this gas is distributed in dark matter haloes of different mass. The neutral hydrogen content of galaxies has been probed at high redshifts ($z > 2$) using the absorption of the Lymanα line by gas clouds along the line of sight to distant quasars (e.g. Lanzetta et al. 1991; Wolfe et al. 1995; Storrie-Lombardi, Irwin & Wolfe 1996; Peroux et al. 2005; Wolfe et al. 2005). A complementary probe of the atomic hydrogen content of galaxies and the physical state of the gas is the 21cm line. A blind survey of 21cm line absorption of gas illuminated by background radio sources has been proposed as an unbiased probe of damped Lyman-α clouds, which would extend to objects with high dust content, unlike surveys for damped Lyman-α absorbers (Kanekar & Briggs 2004; Kanekar et al. 2009). The 21cm absorption line and absorption lines associated with molecular hydrogen have been detected in the same Damped Lyman-α system (Srianand et al. 2010). The 21cm line is a forbidden transition and the detection of the line in emission for low gas masses and high redshifts is challenging. It is only in recent

years that a robust and comprehensive census of atomic hydrogen (HI) in the local universe has been made possible through the HI Parkes All Sky Survey (Barnes et al. 2001; Zwaan et al. 2003, 2005). This work is being extended to lower mass systems by the ALFALFA survey (Giovanelli et al. 2005). Despite this progress, the highest redshift direct detection of HI in emission is very firmly confined to the local Universe at $z = 0.25$ (Catinella et al. 2008. Lah et al. 2009 have reached $z = 0.34$ by stacking a sample of optically selected galaxies; see also Verheijen et al. 2007). However, over the coming decade, this situation is expected to change dramatically with the construction of new, more sensitive radio telescopes such as the precursors of the Square Kilometre Array, MeerKAT (Booth et al. 2009) and ASKAP (Johnston et al. 2008), and the Square Kilometre Array itself (Schilizzi, Dewdney & Lazio 2008). The Widefield ASKAP L-band Legacy All-sky Blind surveY (Johnston et al. 2008) will reach a median redshift of $z \sim 0.05$, compared with $z \sim 0.009$ reached by HIPASS, putting our knowledge of the local HI universe on a par with that in the optical due to the 2dFGRS and SDSS. MeerKAT will be more sensitive than ASKAP but with a smaller field of view, pushing HI detections to higher redshifts. The SKA will revolutionise our understanding of galaxy formation and cosmology, uncovering the HI Universe out to high redshifts. One of the major science goals is to better characterise the evolution of dark energy with redshift. The SKA is expected to provide competitive constraints on the nature of dark energy through high accuracy measurement of large-scale structure in the galaxy distribution over a lookback time representing a significant fraction of the age of the Universe (Albrecht et al. 2006). This conclusion currently rests on very uncertain calculations which we seek to place on a firmer, more physical footing in this chapter.

Modelling the abundance and clustering of HI sources is challenging. A number of possible approaches have been tried; empirical modelling, which relies upon the observations of HI in the Universe, the fully numerical approach, which uses cosmological gas dynamics simulations to model the HI content of galaxies from first principles and semi-analytical modelling, which we use in this chapter. Empirical estimates have been attempted despite the paucity of observational results for guidance (Abdalla & Rawlings 2005; Abdalla, Blake & Rawlings 2010). Such

calculations require an assumption about the evolution of the HI mass function over a broad redshift interval. The only constraint on this assumption is the integrated density of HI, which can be compared with the results inferred from quasar absorption features, which themselves require corrections for unseen low column density systems and dust extinction (Storrie-Lombardi et al. 1996). The empirical approach does not predict the clustering of HI sources. Further assumptions and approximations are necessary to extend this class of modelling so that predictions can be made for galaxy clustering. Another layer of approximation in this class of modelling has been motivated by observations which suggest that HI sources tend to avoid the centres of clusters and that clusters do not boast an important population of satellites (e.g. Waugh et al. 2002; Verheijen et al. 2007). This led Marin et al. (2009) to make a one-to-one connection between halo mass and HI mass. However, the nature of the relation is uncertain and several possibilities are explored by Marin et al. based on different assumptions about the evolution of the HI mass function.

Ideally, a physically motivated model which follows the sources and sinks of cold gas is needed. Gas dynamic simulations are computationally expensive and are typically restricted to small computational volumes, which makes it impossible to accurately follow the growth of structure to the present day. An example is provided by Popping et al. (2009), who carry out a smoothed particle hydrodynamics simulation in a $32h^{-1}$Mpc box. The HI mass function in the simulation is in very poor agreement with the observational estimate of Zwaan et al. (2005), underpredicting the abundance of galaxies of HI mass $10^{10}M_{\odot}$ by a factor of 30, which the authors put down to the small computational volume, and overpredicting low mass systems by a factor of two. Clustering predictions are limited to scales smaller than a few Mpc due to the small box size. Furthermore, it is important to be aware that gas dynamic simulations do not have the resolution to follow all of the processes in galaxy formation directly and in all cases resort to what are essentially semi-analytical rules to treat sub-resolution physics.

Currently the most promising route to making physical and robust predictions for the HI in the Universe is semi-analytical modelling of galaxy formation (see Baugh 2006; Benson 2010). This type of model includes a simplified but physi-

cally motivated treatment of the processes which control the amount of cold gas in a galaxy: gas cooling, galaxy mergers, star formation and reheating of gas by supernovae. These calculations are quick and can rapidly cover the haloes in a cosmological volume. Baugh et al. (2004) presented predictions for the mass function of cold gas galaxies in the `GALFORM` semi-analytical model of Cole et al. (2000) (see also Rawlings et al. 2004). One issue which must be dealt with is that the models predict only the total mass of cold gas, which includes helium, and both atomic and molecular hydrogen. Baugh et al. assumed a fixed ratio of molecular to atomic hydrogen. Obreschkow & Rawlings (2009a,b) developed an empirical model based on observations and theoretical arguments by Blitz & Rosolowsky (2006) in which this ratio could vary from galaxy to galaxy. Obreschkow & Rawlings (2009b) applied this ansatz to the semi-analytical model of de Lucia & Blaizot (2007; see also Obreschkow et al. 2009).

In the first paper in this series, we compared the predictions of a range of semi-analytical models for the mass function of HI and show the global mass density of cold gas Ω_{cold}=ρ_{cold}/ρ_{crit} as a function of redshift for the model predictions have shown in Fig. 2 (Power et al. 2010). Despite the different implementations of the physical ingredients used in the models and the different emphasis placed on various observations when setting the model parameters, the predictions show generic features. Power et al. found that there is surprisingly little variation in the predicted HI mass function with redshift, and that the models make similar predictions for the rotation speed and size of HI systems. The models predict the mass of cold gas and so a conversion is required to turn this into a HI mass. Currently the most uncertain step is the assumption about what fraction of hydrogen is in atomic form and what fraction is molecular. Power et al. presented predictions for two cases, one in which all model galaxies are assumed to have a fixed molecular to atomic hydrogen ratio (H_2/HI) and the other in which this ratio varies from galaxy to galaxy, depending upon the local conditions in the galactic disk (Blitz & Rosolowsky 2006). The assumption of a variable H_2/HI ratio results in a dramatic reduction in the number of HI sources in the tail of the redshift distribution.

In this chapter we look at the distribution of cold gas in galaxies as a function of

halo mass. In particular we look at the halo occupation distribution (HOD) for HI galaxies, which gives the mean number of galaxies of a given HI mass as a function of dark matter halo mass, and the clustering of HI galaxies. Using this information, we assess the potential of the SKA to measure the baryonic acoustic oscillation (BAO) signal. We briefly review the GALFORM model in Section 4.2, explaining the differences between the four models that we consider. We then look at the halo occupation distribution of cold gas galaxies in Section 4.3, in which we also present predictions for the clustering of cold gas galaxies at different redshifts and compare to measured clustering at the present day. In Section 4.4 we compare the performance of future redshift surveys in the optical and using HI emission for measuring the properties of the dark energy. We present a summary along with our conclusions in Section 4.5.

4.2 Galaxy formation models and basic predictions

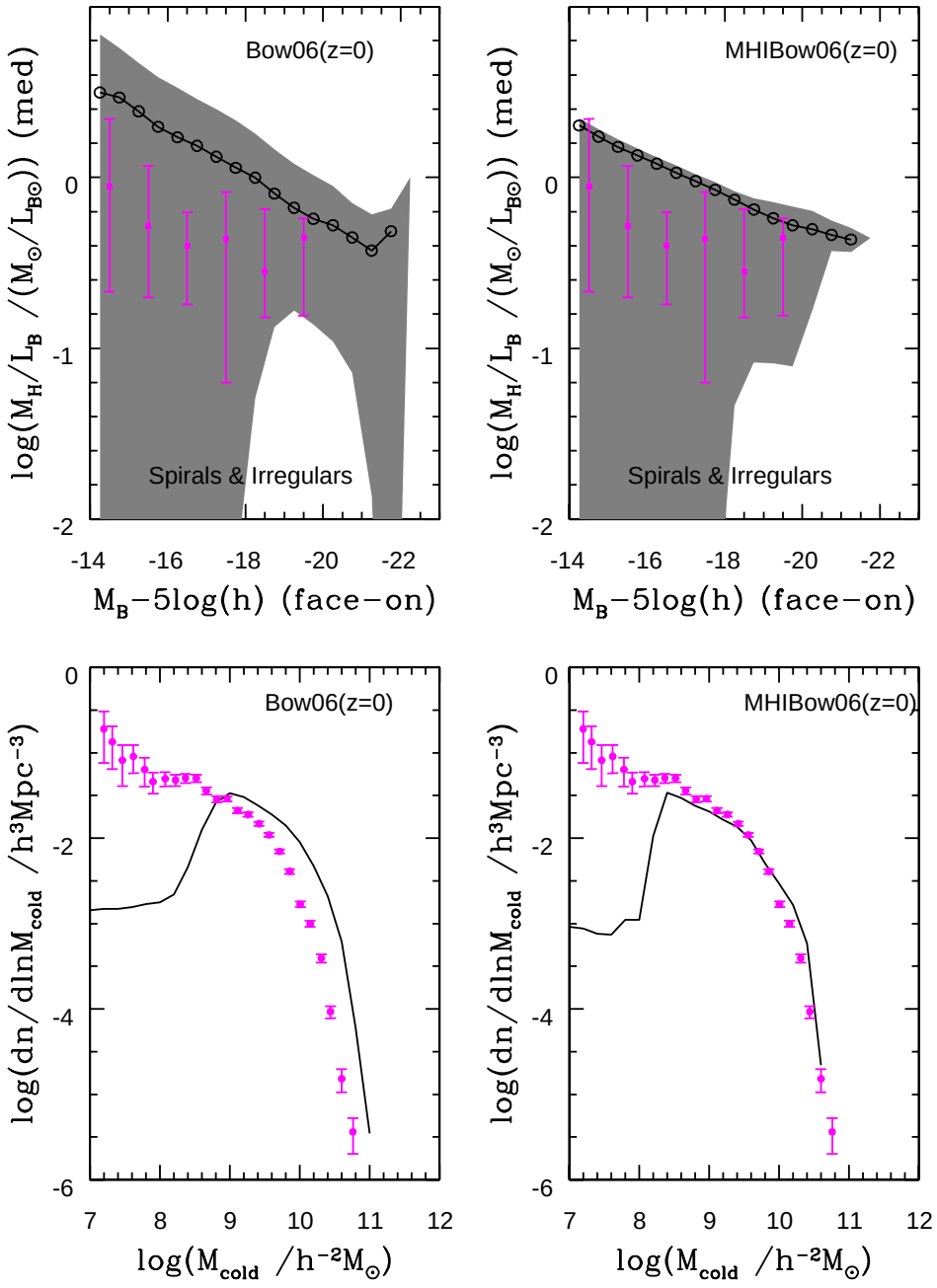

Figure 4.1: The predicted ratio of neutral hydrogen mass to B-band luminosity (upper panels) and the cold gas mass function (lower panels) in the Bow06 (left panels) and MHIBow06 models (right panels). In the upper panels, the magenta points show observational estimates of the hydrogen mass to luminosity ratio using data from Huctmeier & Richter (1988) (HI) and Sage (1993) (H_2). The black points show the median ratio predicted by the models and the grey shading shows the 20 - 80 percentile range of the predicted distribution. We assume that 76% by mass of the cold gas predicted by the models is neutral hydrogen. In the lower panels, the magenta points show the cold gas mass function derived from the HI mass function estimated by Zwaan et al (2005). Here, a constant H_2/HI ratio of 0.4 has been assumed to convert the HI measurement into a cold gas mass.

Table 4.1: The values of selected parameters which differ between the models. The columns are as follows: (1) The name of the model. (2) The equation used to calculate the star formation timescale, $\tau_\star$. (3) The value of $\epsilon_\star$ or $\tau_0^\star$ used in the star formation timescale. (4) The AGN feedback parameter, $\alpha_{\rm cool}$, (Eq. 4.1) (5) The supernova feedback parameter,$V_{\rm hot}$ (Eq. 4.2). (6) The source of halo merger histories. (7) Comments giving model source or key differences from published models.

	$\tau_\star$	$\epsilon_\star$ or $\tau_\star^0$ [Gyr]	$\alpha_{\rm cool}$	$V_{\rm hot}$[kms^{-1}]	Merger tree	Comments
Bow06	Eq. 4.3	0.0029	0.58	485	N-body	Bower et al. (2006)
Font08	Eq. 4.3	0.0029	0.70	485	N-body	Font et al. (2008) Modified cooling recipe in satellites from Bow06
MHIBow06	Eq. 4.4	8	0.62	485	N-body	Modified star formation recipe from Bow06
GpcBow06	Eq. 4.4	4	0.72	390	Monte Carlo	Different background cosmology and modified star formation recipe from Bow06

Semi-analytical models of galaxy formation invoke simple, physically motivated recipes to follow the fate of the baryons in a universe in which structure in the dark matter grows hierarchically (White & Rees 1978; White & Frenk 1991; Kauffmann et al. 1993; Cole et al. 1994; for a review of this approach see Baugh 2006). The current generation of models include a wide range of phenomena, ranging from the heating of the intergalactic medium, which affects the cooling of gas in low mass haloes, to the suppression of cooling flows in massive haloes due to heating by accretion of matter onto supermassive black holes (e.g. Bower et al. 2006; Croton et al. 2006; Cattaneo et al. 2007; Monaco et al. 2007; Lagos, Cora & Padilla 2008). In this chapter, we use the Durham semi-analytical galaxy formation code `GALFORM` to make predictions for the amount of cold gas in dark matter haloes of different masses. This code was introduced by Cole et al. (2000) and has been developed in a series of papers (Benson et al. 2003; Baugh et al. 2005; Bower et al. 2006; Font et al. 2008). The code predicts a wide range of properties for the galaxy population in the context of a spatially flat cold dark matter cosmology with a cosmological constant.

In this chapter we consider four different models run using `GALFORM`. Two of these are available from the Millennium Archive[1]; these are the Bower et al. (2006; hereafter Bow06) and Font et al. (2008) models (hereafter Font08). The third model is a modified version of the Bow06 model (which we label as MHIBow06), which is discussed in more detail below. In this model a small number of parameters have been adjusted from the values used in Bow06 in order to produce a better match to the cold gas mass function estimated by Zwaan et al. (2005). The fourth model (denoted by GpcBow06) is set in a different background cosmology from the other three, which adopt the cosmology of the Millennium simulation (Springel et al. 2005). The cosmology of the GpcBow06 model is in better agreement with recent measurements of the cosmic microwave background and the large-scale structure of the Universe (Sanchez et al. 2009).[2] The Bow06, Font08 and MHIBow06 models

[1] http://galaxy-catalogue.dur.ac.uk:8080/Millennium/

[2] The cosmological parameters used in the Millennium simulation are a matter density $\Omega_0 = 0.25$, a cosmological constant $\Lambda_0 = 0.75$, a Hubble constant $H_0 = 73\,\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, a primordial

use merger histories extracted from the Millennium Simulation. The GpcBow06 model uses Monte Carlo generated merger trees as described below. When we make predictions for the spatial distribution of galaxies in the GpcBow06 model, we use the GigaParsec simulation run at the Institute for Computational Cosmology (GPICC; Baugh et al. in preparation), which uses 10 billion particles to model the hierarchical clustering of mass in a simulation cube $1000h^{-1}\,\mathrm{Mpc}$ on a side. To keep the number of models manageable, we do not consider the Baugh et al. (2005) model in this chapter. This model was included in the study by Power et al. (2010). The star formation recipe used in the MHIBow06 model is based on that used in Baugh et al. (2005).

scalar spectral index $n_\mathrm{s} = 1$, baryon density $\Omega_\mathrm{b} = 0.045$ and fluctuation amplitude $\sigma_8 = 0.9$. In the Sanchez et al. (2009) best fitting model these parameters become $\Omega_0 = 0.26$, $\Lambda_0 = 0.74$, $H_0 = 71.5\,\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, $n_\mathrm{s} = 0.96$, $\Omega_\mathrm{b} = 0.044$, and $\sigma_8 = 0.8$.

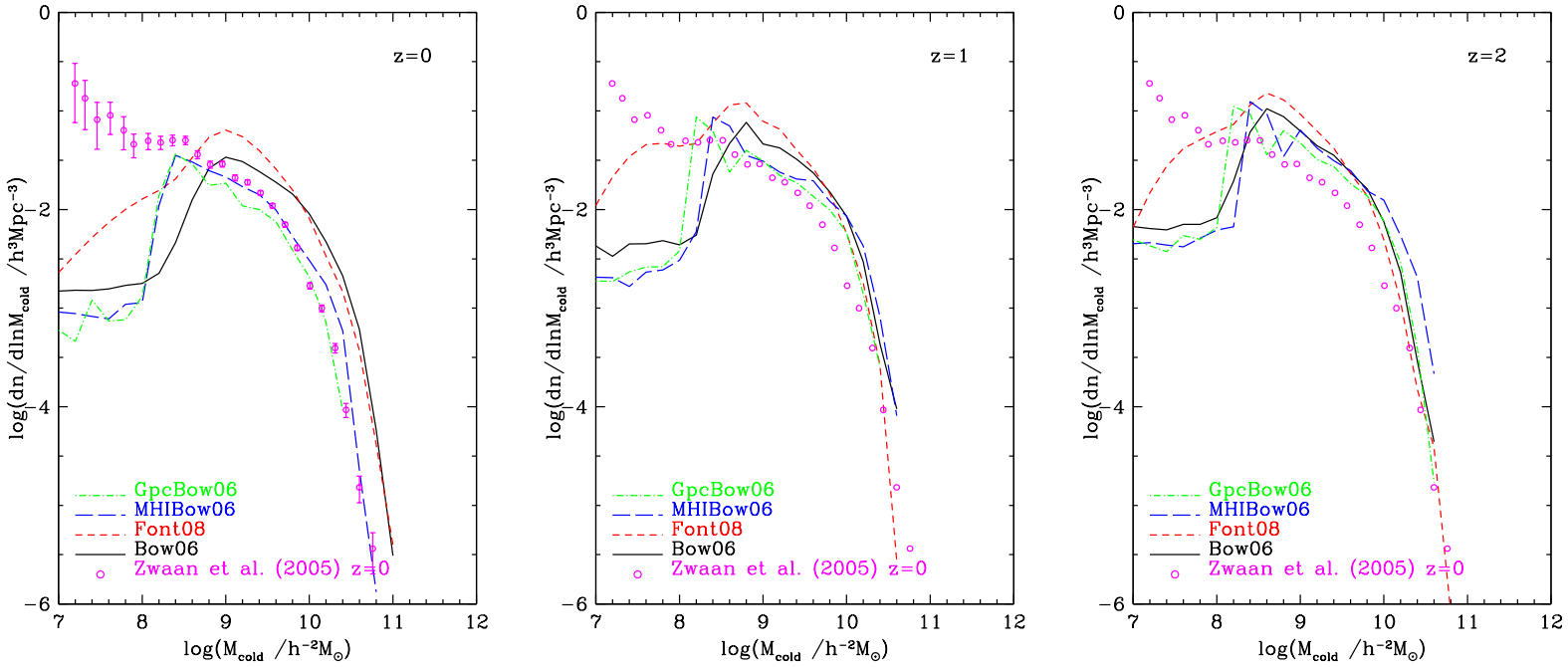

Figure 4.2: The cold gas mass function predicted in the four models at z=0 (left), $z = 1$ (middle) and z=2 (right). Different colours and line types correspond to different models as indicated by the legend. The points show the local ($z = 0$) observational estimate of the cold gas mass function inferred from the HI mass function of Zwaan et al. (2005) (see text in Section 4.3.2 for details of the conversion). These data are reproduced without error bars in the z=1 and z=2 panels as a reference from which to illustrate the evolution of the mass function.

The MHIBow06 and GpcBow models use the same basic physical ingredients as the Bow06 model but with a change in the assumed star formation timescale. The Font08 model is based on Bow06, with a modification to the cooling prescription. We discuss these differences in more detail below. We first discuss some of the ingredients which are varied between the models, in order to introduce some of the parameter definitions used in `GALFORM`.

All of the models we consider in this chapter include the suppression of cooling flows in massive haloes, as a result of the energy released following accretion of matter onto a central supermassive black hole (Bower et al. 2006; Malbon et al. 2007; Fanidakis et al. 2009). A halo is assumed to be in quasi-hydrostatic equilibrium if the time required for gas to cool at the cooling radius, $t_{\rm cool}(r_{\rm cool})$, exceeds a multiple of the free-fall time at this radius, $t_{\rm ff}(r_{\rm cool})$:

$$t_{\rm cool}(r_{\rm cool}) > \frac{1}{\alpha_{\rm cool}} t_{\rm ff}(r_{\rm cool}), \tag{4.1}$$

where $\alpha_{\rm cool}$ is an adjustable parameter, whose value controls the sharpness and position of the break in the optical luminosity function. The cooling flow in the halo is then shut down completely if the luminosity released by accretion of matter onto the supermassive black hole (SMBH) exceeds the cooling luminosity. The energy released by accretion depends on the mass of the SMBH (see, for example, Fanidakis et al. 2009).

The models also include the ejection of cooled gas into the hot halo due to heating by supernovae. The strength of supernovae feedback is defined by the factor β:

$$\beta = (V_{\rm hot}/V_{\rm disk})^{\alpha_{\rm hot}}. \tag{4.2}$$

The rate at which gas is reheated is β times the star formation rate. Here $V_{\rm disk}$ is the circular velocity of the disk at its half mass radius, and $V_{\rm hot}$ and $\alpha_{\rm hot}$ are parameters. A similar equation holds for supernova feedback in the galactic bulge. In `GALFORM`, the parameters $V_{\rm hot}$ and $\alpha_{\rm hot}$ are set without reference to the number of supernovae. The primary constraints on these parameters are the shape of the luminosity function, the slope of the disk rotation speed - luminosity relation and the scale size of disks (see Cole et al. 2000).

The Bow06 and Font08 models use a star formation timescale in disks, $\tau_\star$, which is proportional to the galactic dynamical time, $\tau_{\rm dyn}$, and is given by :

$$\tau_\star = \epsilon_\star^{-1} \tau_{\rm dyn} (V_{\rm disk}/200\,{\rm kms}^{-1})^{\alpha_\star}, \tag{4.3}$$

where $\alpha_\star$ and $\epsilon_\star$ are adjustable parameters ($\alpha_\star = -1.5$ in both cases). The dynamical time is defined as $\tau_{\rm dyn} = r_{\rm disk}/V_{\rm disk}$. In contrast, the MHIBow06 and the GpcBow08 models adopt a star formation timescale which does not depend on the galactic dynamical time. Instead, in these cases, the timescale is given by :

$$\tau_\star = \tau_\star^0 (V_{\rm disk}/200\,{\rm kms}^{-1})^{\alpha_\star}, \tag{4.4}$$

where $\tau_\star^0$ and $\alpha_\star$ are adjustable parameters (again, in both cases, $\alpha_\star = -1.5$); this parameterization was used in Baugh et al. (2005).

The Font08 model includes an improved treatment of the ram-pressure stripping of hot-gas atmospheres of satellite galaxies, motivated by the hydrodynamic simulations of McCarthy et al (2008). Also in this model, the yield of metals per solar mass of stars formed is increased by a factor of two over the default but rather uncertain value expected for a standard solar neighbourhood stellar initial mass function. These changes are motivated in part by the desire to improve the predictions of the Bow06 model for the colour magnitude relation of central and satellite galaxies in groups. The revision to the stellar yield reddens the colour of all galaxies in the Font08 model compared with Bow06. The change in the cooling model changes the relative abundance of galaxies in the red and blue populations at low luminosities. In the Font08 model, there are more faint blue satellite galaxies than in the Bow06 model. These galaxies are starved of freshly cooled gas in Bow06 and so had redder stellar populations. The predicted colours in the Font08 model are in much better agreement with the observed colour magnitude relation measured by Weinmann et al. (2006).

The motivation for the MHIBow06 model is clear from Fig. 4.1. This plot shows the galactic neutral hydrogen mass to optical luminosity ratio and the cold gas mass function at the present day. Note that when we plot the mass function (lower panels of Fig. 4.1, cold gas masses are plotted in units of $h^{-2}M_\odot$ rather than $h^{-1}M_\odot$, which is the unit used in the simulation. This ensures that the observational units (which

depend upon the square of the luminosity distance) are matched. The Bow06 model predicts a gas mass to luminosity ratio with the wrong zeropoint and slope. Since this model gives an excellent match to the local optical luminosity function, the discrepancy in the gas to luminosity ratio results in a poor match to the cold gas mass function. The MHIBow06 model uses the star formation timescale given by Eq. 4.4 and also adopts a different value for the AGN feedback free parameter, $\alpha_{\rm cool}$ (Eq. 4.1; see Table 4.1). The right hand panels of Fig. 4.1 show that the MHIBow06 model is in much better agreement with the observed gas to luminosity ratio and cold gas mass function for cold gas masses in excess of $\sim 3 \times 10^8 h^{-2} M_{\odot}$. Note that the models predict the mass of cold gas, which includes helium, atomic hydrogen and molecular hydrogen. The observed mass function in the lower panels of Fig. 4.1 is measured in terms of the atomic hydrogen (HI) content of galaxies. To convert this into a cold gas mass, we have assumed a fixed ratio of molecular to atomic hydrogen and corrected for the mass fraction of helium (see Power et al. 2010). We have also tried applying a variable H_2/HI conversion, as advocated by Blitz & Rosolowsky (2006) and Obreschkow & Rawlings (2009a). Whilst this improves the comparison of the Bow06 HI mass function with the observational estimate of Zwaan et al. (2005), to obtain a satisfactory match it is still necessary to change the model parameters, as in the MHIBow06 model. We shall return to this point in Section 4.5. Note that while the cold gas mass function of the MHIBow06 model differs from the Bow06 model by changing the star formation time scale, the expected optical luminosity functions from the Bow06 model and MHIBow06 model are very similar and fairly well describe the observation luminosity function. Because the cold gas mass reduces by changing star formation time scale from Eq. 4.3 to Eq. 4.4 and the instantaneous star formation rate increase. However, the increasing instantaneous of star formation rate leads more active supernovae feedback process which suppress the star formation. Therefore, the changing star formation time scale not affects on the optical luminosity function, but on the cold gas mass function.

The GpcBow06 model starts from the Bow06 model, with small adjustments made to the galaxy formation parameters to obtain a good match to the optical luminosity function (this is required because the cosmological model has changed

from that used in Bow06) and also to reproduce the observed HI mass function. The GpcBow06 model uses Monte-Carlo merger trees generated using the improved algorithm devised by Parkinson et al. (2008).

Fig. 4.2 shows the cold gas mass function predicted by the four models at z=0, 1 and 2. The Bow06 and Font08 models overpredict the abundance of galaxies with a given cold gas mass at $z = 0$ compared with the observational estimate by Zwaan et al. (2005). On the other hand, the cold gas mass functions of the MHIBow06 and GpcBow06 models agree well with the local observational estimate for masses in excess of $10^{8.5}h^{-2}M_{\odot}$. The discrepancy between the predictions and observations at lower masses is not due to the finite resolution of the N-body halo merger trees. The turnover can be traced back to the modelling of the photoionisation of the intergalactic medium and the impact this has on the cooling of gas in low mass haloes. In all cases a particularly simple approach is taken to model this effect, whereby cooling in low circular velocity haloes ($v_{\rm c} < v_{\rm cut}$) is suppressed below the redshift at which the universe is assumed to have been reionised ($z_{\rm cut}$). The parameters adopted ($v_{\rm cut} = 50\,{\rm km\,s^{-1}}$ and $z_{\rm cut} = 6$) may overestimate the impact of this effect according to recent simulations by Okamoto, Gao & Theuns (2008). The form of the observed HI mass function at low masses could give interesting contraints on the modelling of photoionisation and supernova feedback (Kim et al, in preparation). Here we focus on the more massive galaxies which dominate the overall HI content of the Universe.

4.3 The spatial distribution of cold gas

We now compare the predictions of the four galaxy formation models for the spatial distribution of cold gas with one another and with observations. To understand the spatial distribution of cold gas, we first look at the halo occupation distribution (HOD; Benson et al. 2000; Peacock & Smith 2000; Seljak 2000; Berlind & Weinberg 2002). This quantifies the number of galaxies above a given cold gas mass, as a function of dark matter halo mass (Sec 4.3.1). We present predictions for the correlation function of galaxies selected by their cold gas mass in Sec 4.3.2.

4.3.1 The halo occupation distribution

Variation of Cold Gas Mass with Halo Mass

Before considering the halo occupation distribution directly, it is instructive to first look at how the cold gas mass of galaxies varies with the mass of their host dark matter halo, which we plot in Fig. 4.3 for the Bow06 model. The median cold gas mass as a function of host halo mass is plotted separately for central and satellite galaxies. There is a tight correlation between the mass of cold gas of a central galaxy and its host halo mass for galaxies in haloes less massive than $\sim 3 \times 10^{11} h^{-1} M_{\odot}$. In haloes more massive than this, AGN feedback suppresses gas cooling and there is a dramatic break in the galaxy cold gas mass - halo mass relation, with an accompanying increase in the scatter. The galaxies with the largest mass of cold gas do not lie in the most massive dark matter haloes, but reside instead in haloes with masses $\sim 10^{12} h^{-1} M_{\odot}$. The predicted cold gas mass - halo mass relation is remarkably similar to that inferred observationally (Wyithe et al. 2009a). Another conclusion that is readily apparent from Fig. 4.3 is that the bulk of the baryons associated with a dark matter halo are not in the form of cold gas. The solid line in this plot shows the mass a galaxy would have if all of the available baryons in the halo were in the form of cold gas in one object, assuming the universal baryon fraction. The points are some way below this line for two reasons: 1) in most haloes, the bulk of the baryons are in the hot phase and 2) there is more than one galaxy in most haloes.

Cold Gas Halo Occupation Distributions

We now examine the predictions for the halo occupation distribution (HOD) of galaxy samples constructed according to cold gas mass. The HOD gives the mean number of galaxies which satisfy a given selection criterion as a function of halo mass, and can be broken down into the contribution from the central galaxy in a halo and its satellite galaxies. In the case of optically selected galaxy samples, the HOD is commonly described by a step function for central galaxies and a power law for satellite galaxies (Peacock & Smith 2000; Seljak 2000; Berlind & Weinberg 2002; Zheng 2004). Many attempts have been made to interpret the clustering of

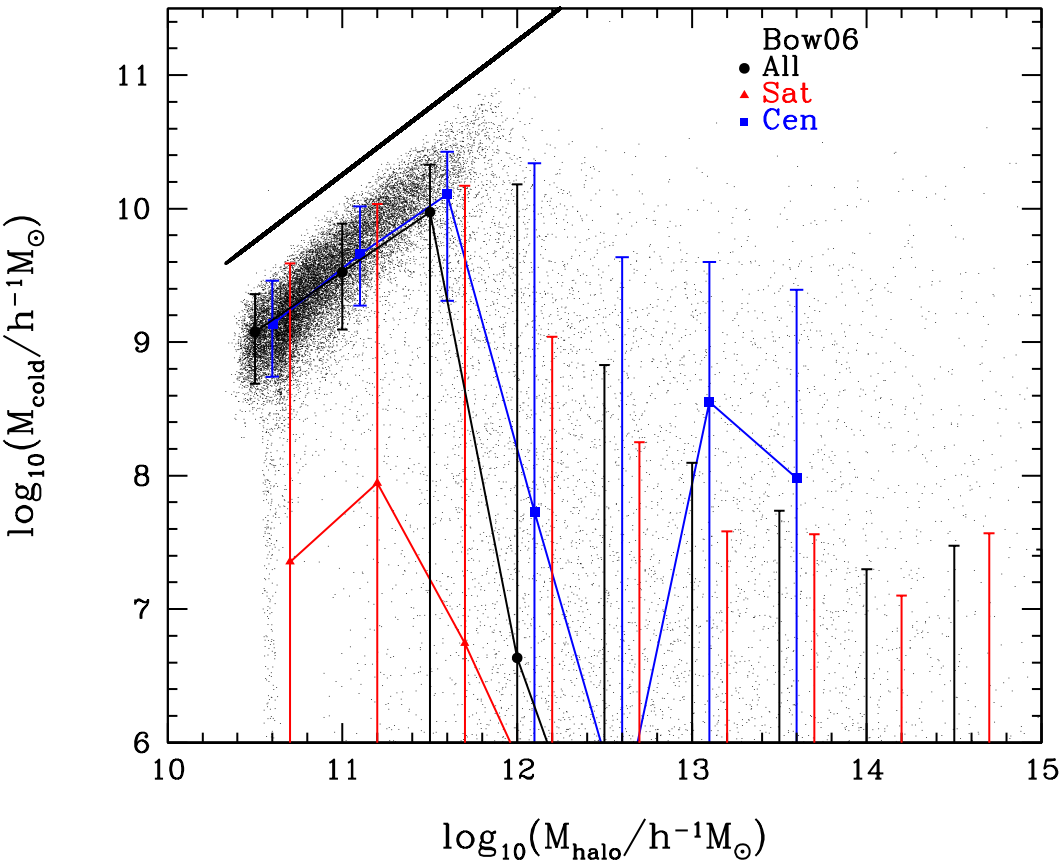

Figure 4.3: The cold gas mass of galaxies in the Bow06 model as a function of the mass of their host dark matter halo. The black points show individual galaxies. The symbols joined by lines show the median cold gas mass as a function of halo mass, for central galaxies (blue), satellite galaxies (red) and all galaxies (black). The bars show the 10-90 percentile range of the distribution of cold gas masses. All galaxies, including those with zero cold gas mass are included when computing the median and percentile range. The solid black line shows the cold gas mass a galaxy would have if all the available baryons in its halo were in the form of cold gas in one object.

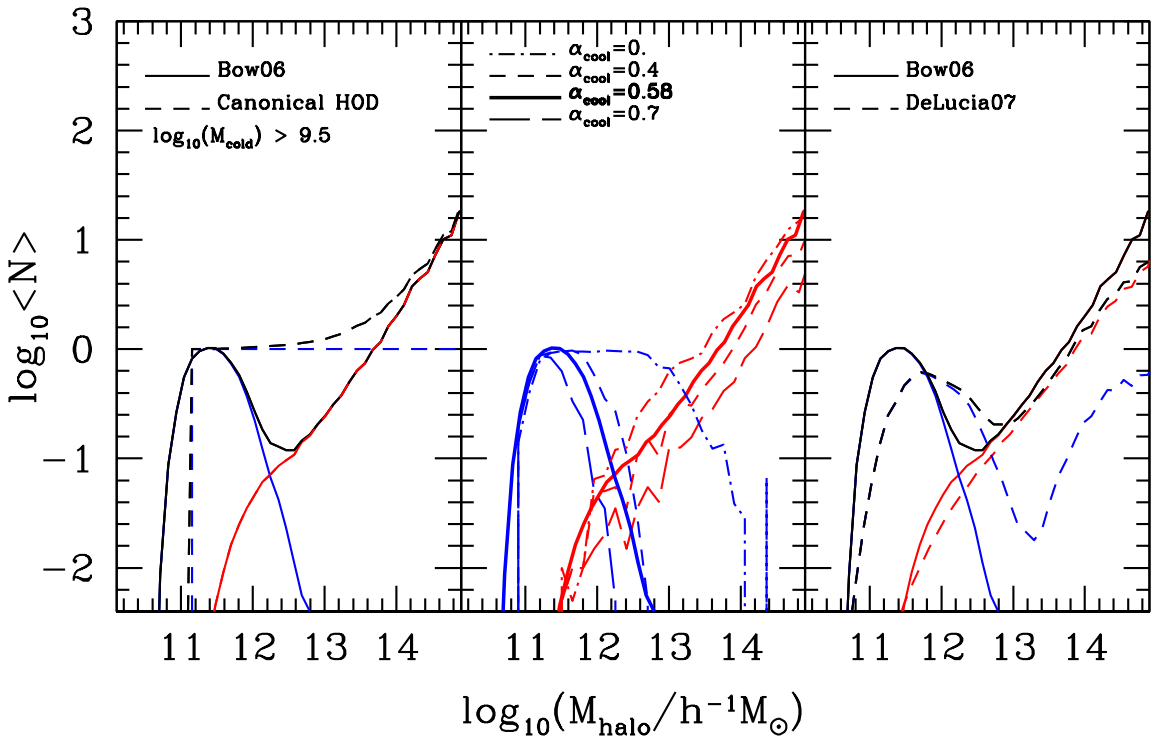

Figure 4.4: The predicted halo occupation distribution (HOD) of galaxies with cold gas mass in excess of $10^{9.5}h^{-2}\mathrm{M}_{\odot}$, chosen to match the sample of galaxies for which Wyithe et al. (2009) estimated the HOD for in HIPASS. The left panel shows the HOD predicted in the Bow06 model (solid lines: blue shows the central galaxy HOD, red shows satellites and black shows the overall HOD). The dashed blue line shows a step function designed to reproduce the number of central galaxies in Bow06. The dashed black line shows this step function combined with the model HOD for satellites. The central panel shows the impact on the HOD of changing the halo mass above which AGN feedback stops the cooling flow. The fiducial Bow06 model corresponds to $\alpha_{\rm cool} = 0.58$. In a model with a larger value of $\alpha_{\rm cool}$, the onset of cooling suppression can shift to lower mass haloes; reducing $\alpha_{\rm cool}$ means that cooling is only switched off in more massive haloes. The right hand panel compares the HOD predicted by Bow06 (solid lines) with that in the model of DeLucia & Blaizot (2007), for the same cold gas mass threshold (dashed lines). The colour coding is the same in each panel.

optically selected galaxy samples using the HOD formalism (van den Bosch et al. 2003; Magliocchetti & Porciani 2003; Zehavi et al. 2005; Yang et al. 2005; Tinker et al. 2007; Wake et al. 2008; Kim et al. 2009). In contrast, there are few studies of the clustering of galaxies selected on the basis of their atomic hydrogen mass using the HOD formalism (Wyithe et al 2009a, 2009b; Marin et al. 2009).

Fig. 4.4 shows the typical form predicted by the models for the HOD of galaxies selected by their cold gas mass. The left panel shows the HOD for galaxies in the Bow06 model which have cold gas masses in excess of $3 \times 10^9 h^{-2} \mathrm{M}_\odot$, chosen to have the same HI mass cut as HIPASS. For this mass threshold, the abundance of central galaxies is sharply peaked around a halo mass of $\sim 2 \times 10^{11} h^{-1} M_\odot$. The HOD of satellite galaxies reaches unity in haloes which are a hundred times more massive. In these haloes, the central galaxy has a cold gas mass below the cut-off; there is essentially zero chance of finding a halo which contains a central galaxy and a satellite galaxy above this cold gas mass threshold. However, this does not imply that it is impossible to find more than one galaxy per halo with cold gas masses above the threshold, simply that when this occurs (i.e. once a sufficiently massive halo is considered), both galaxies will be satellites.

For comparison, we also plot in the left hand panel of Fig. 4.4 the traditional form adopted for the HOD of central galaxies (i.e. a step function). The minimum halo mass in this case is set by the requirement that the step function reproduces the number of central galaxies in the Bow06 model. The step function HOD is markedly different to the predicted HOD, which is closer to a Gaussian. A similar conclusion about the peaked form of the central galaxy HOD was postulated by Zehavi et al. (2005) for blue central galaxies. Wyithe et al. (2009a) model the clustering of galaxies in the HIPASS survey by adopting a step function for the central galaxy HOD and a truncated power law for satellite galaxies, such that haloes above some mass cut contain no satellites. The truncation point lies in the halo mass range 10^{14}-$10^{15} h^{-1} M_\odot$, depending on the slope of the satellite HOD. As we shall see later on, whilst this truncation is not predicted by any of the models, this has little impact on the abundance or clustering of the galaxies.

In Fig. 4.4, the HOD of central galaxies in the Bow06 model drops far below

unity above a halo mass of $\sim 10^{12} h^{-1} \mathrm{M}_{\odot}$. In this model there is very little cold gas in haloes more massive than this due to the shut down of the cooling flow by AGN heating. To illustrate this, in the middle panel of Fig. 4.4 we vary the halo mass which marks the onset of AGN heating by changing the value of the $\alpha_{\rm cool}$ parameter (see Eq. 4.1). Reducing the value of $\alpha_{\rm cool}$ results in the halo mass in which cooling stops being shifted to higher masses. In the absence of AGN heating (i.e. $\alpha_{\rm cool}$=0), the central galaxy HOD still drops below the unity in the most massive haloes (M_{halo}>$10^{13} h^{-1} \mathrm{M}_{\odot}$) due to the longer cooling time of the gas in these haloes. These haloes typically have a lower formation redshift and thus a lower gas density and are also hotter; hence they have a longer cooling time. Cold gas is depleted by star formation in such massive haloes.

We shall see later that the peaked HOD for central galaxies is common to all of the `GALFORM` models considered, particularly at low redshift. We now examine whether or not this feature is peculiar to the way AGN feedback is implemented in `GALFORM` by comparing the Bow06 predictions with those of De Lucia & Blaizot (2007; hereafter the DeLucia07 model). The right hand panel of Fig. 4.4 shows that the central galaxy HOD in the DeLucia07 model is somewhat broader than that predicted in Bow06, and even increases beyond a halo mass of $\sim 2 \times 10^{13} h^{-1} M_{\odot}$. However, as we shall demonstrate further on in this section, this upturn has little impact on the predicted clustering. The suppression of gas cooling in the DeLucia07 semi-analytical model is smoother than in `GALFORM` (see Croton et al. 2006 for a description of the implementation of radio mode feedback). Some gas is permitted to cool in haloes with hot gas atmospheres in the DeLucia07 model, with the cooling rate modified by accretion onto the central SMBH. In `GALFORM`, the cooling flow and heating rate are assumed to balance exactly whenever there is a quasi-hydrostatic hot halo and the Eddington luminosity of the black hole exceeds the cooling luminosity.

Figs. 4.5, 4.6 and 4.7 show the HOD in the four Durham models at $z = 0, 1$ and 2. Each column shows the HOD predicted for a different cold gas mass threshold, with the mass cut increasing to the right. The rows show the different models introduced in Sec. 4.2. For the most massive cold gas mass threshold plotted in Fig. 4.5, the mean occupation number in the MHIBow06 and GpcBow06 models

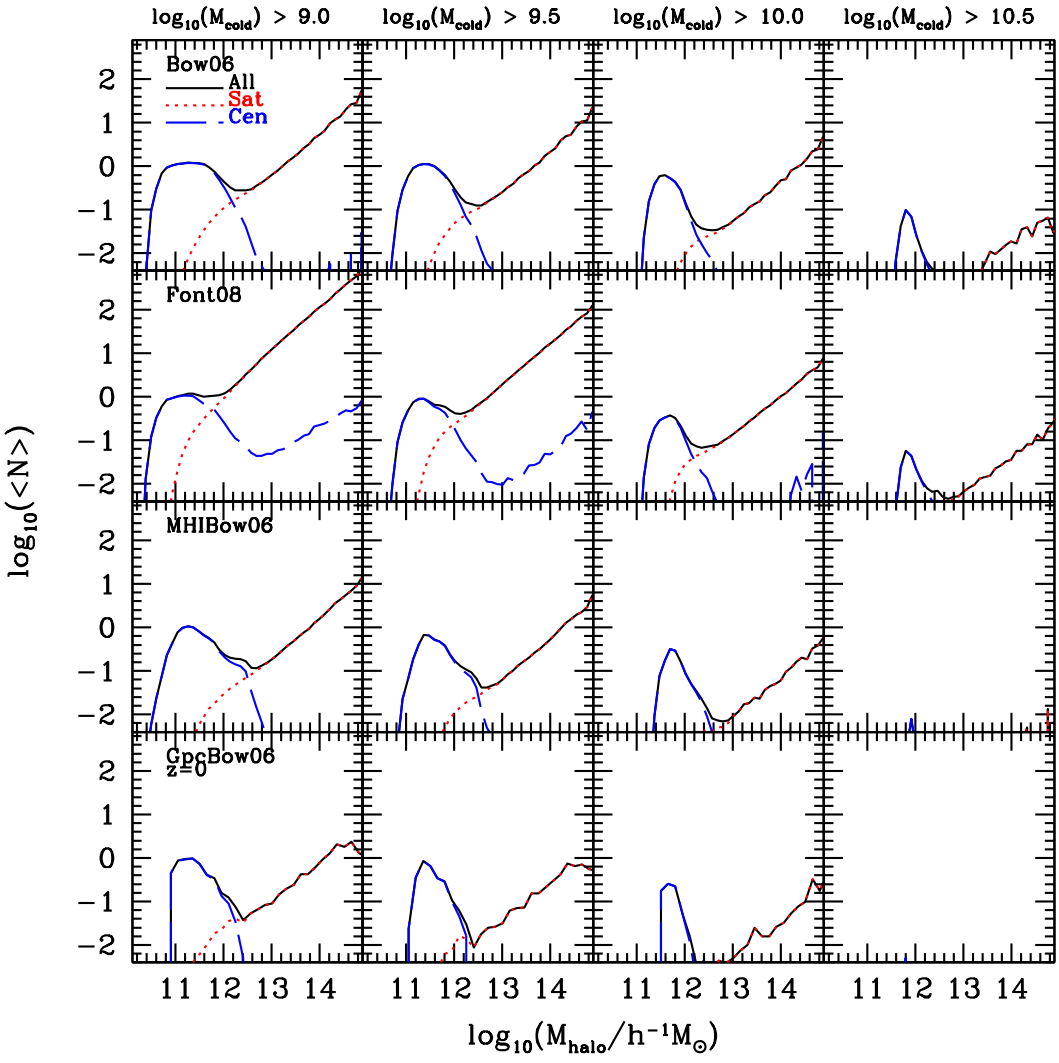

Figure 4.5: The halo occupation distribution, i.e. the mean number of galaxies passing the selection labelled per halo, at $z = 0$ for galaxy samples defined by cold gas mass thresholds. The blue dashed curves show the contribution from central galaxies, the red dotted curves show satellite galaxies and the black solid curves show the overall HOD. Each row corresponds to a different model, and each column to a different cold gas mass threshold, as labelled.

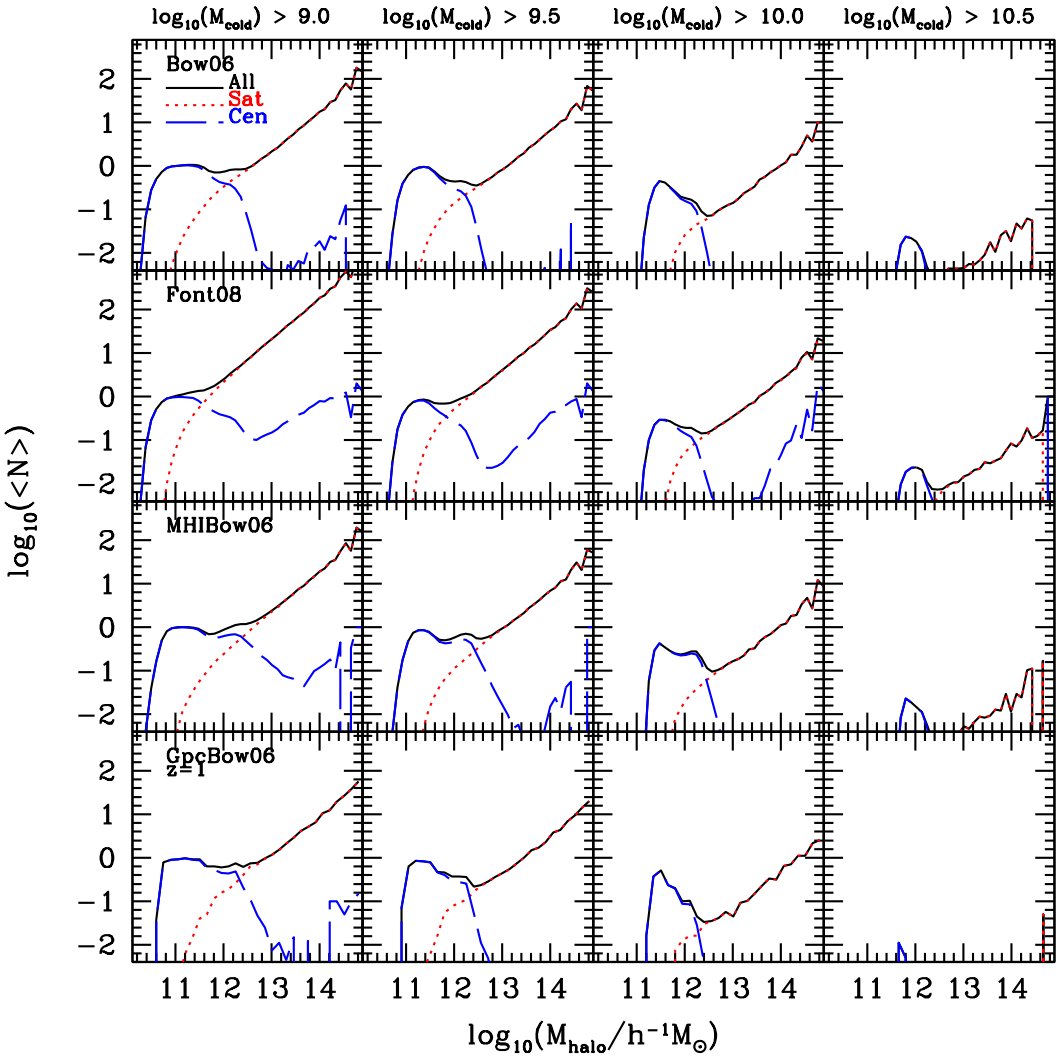

Figure 4.6: The halo occupation distribution at $z = 1$ for samples defined by a threshold cold gas mass. The blue dashed curves show the contribution from central galaxies, the red dotted curves show satellite galaxies and the black solid curves show all galaxies. Each row shows a different model as labelled, using the notation set up in Section 4.2.

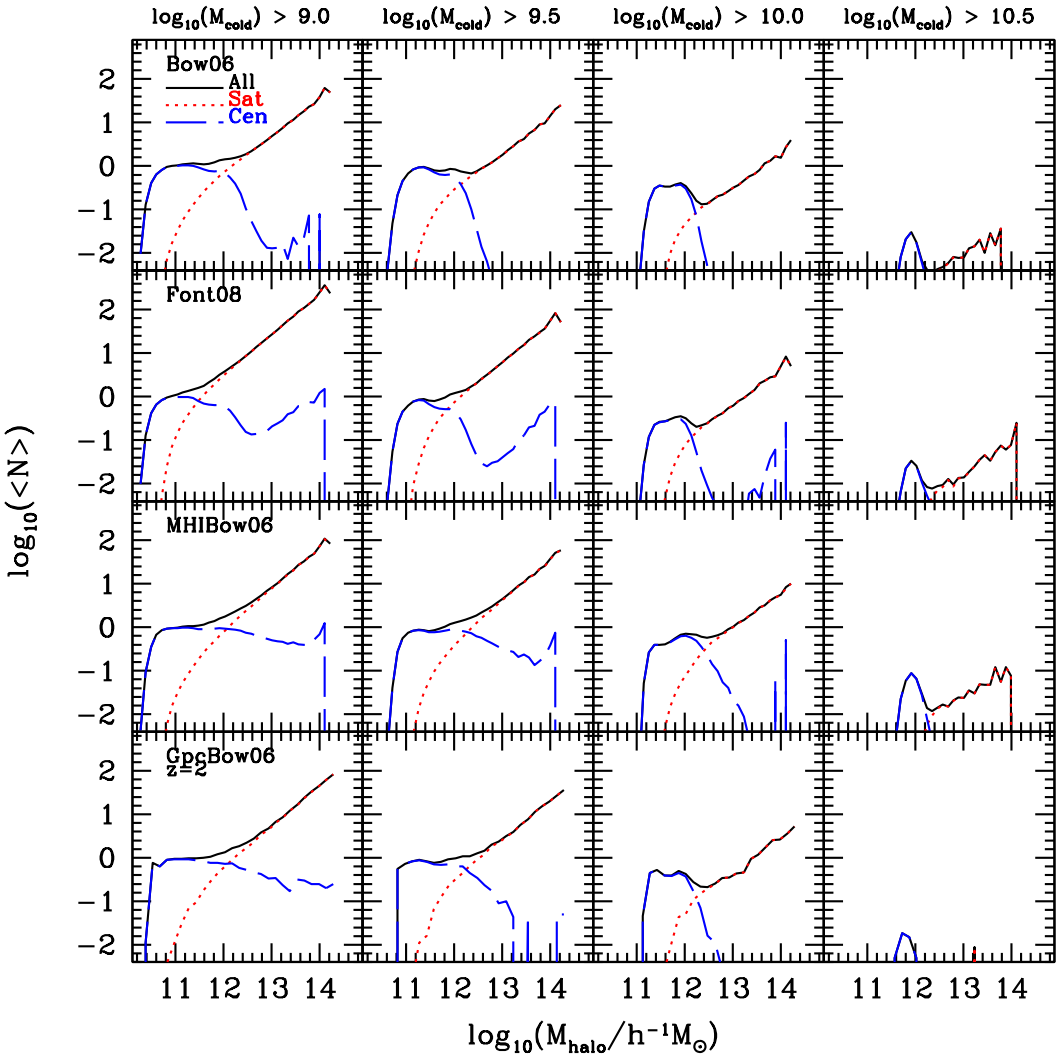

Figure 4.7: The halo occupation distribution at $z = 2$. As before, the blue dashed curves show the contribution from central galaxies, the red dotted curves show satellite galaxies and the black solid curves show all galaxies. Each row shows a different model as described in Section 4.2. Each column corresponds to a different cold gas mass threshold as labelled.

is less than 1 galaxy per 100 haloes. In the Bow06 model, the HOD peaks at a halo mass just under $10^{12}h^{-1}M_{\odot}$, with around 1 in 10 such haloes hosting a central galaxy with cold gas mass above the threshold.

The size of the departure from the traditionally assumed step function HOD for central galaxies at $z = 0$ in Fig. 4.5 varies in proportion to the "strength" of AGN feedback for the Bow06, Font08 and MHIBow06 models (see Table 4.1). Although the GpcBow06 model has the weakest AGN feedback, the deviation from a step function is largest in this case since this model adopts weaker supernovae feedback than the other models (as a result of being set in a different cosmology, with a lower density fluctuation amplitude). The departure of the central galaxy from a step function form is less pronounced at $z = 1$ (Fig. 4.6). This is because fewer haloes have hot gas haloes and those which do host lower mass SMBH (see Fanidakis et al. 2009 for plots showing how the mass of SMBH is built up over time in the models). These trends continue in Fig. 4.7, which shows the HOD for the `GALFORM` models at $z = 2$. The HOD of central galaxies is now better approximated by a step function. The HODs become noisy for massive haloes as such objects are extremely rare at this redshift. The central galaxy HOD in the Font08 model has a Gaussian form centered on halo masses of a few times $10^{11}h^{-1}M_{\odot}$. The HOD displays an upturn for more massive haloes which is reminiscent of the HOD in the DeLucia07 model. In Font08, this mass could be brought in by merging satellites, which will have a higher cold gas mass than in the other Durham models. The central galaxy HOD becomes closer to the canonical step function form with increasing redshift.

Fig. 4.5 shows that the amplitude of the HOD for satellite galaxies in the Bow06 model is higher than in the MHIBow06 and GpcBow06 models. This is due in part to the Bow06 model predicting a higher abundance of galaxies by cold gas mass than is observed (see Fig. 4.2). The Font08 model predicts many more satellite galaxies than the other models ($\sim$10 times more for the two lowest cold gas mass thresholds). This can be traced back to the modified cooling model in Font08, which means that satellites accrete gas that cools from their incompletely stripped hot haloes. Also some of the gas which is reheated by supernovae in the satellite is allowed to recool onto the satellite rather than being incorporated into the main

hot halo. The amplitude of the HOD for satellite galaxies at $z = 1$ (Fig. 4.6 in the Bow06, MHIBow06, and GpcBow06 models is higher than predicted at $z = 0$. Star formation depletes the cold gas by $z = 0$. The power law slope of the satellite HOD is remarkably constant regardless of cold gas mass threshold, redshift or galaxy formation model, with $N_{\rm sat} \propto M_{\rm halo}^{0.8}$. The predicted slope is in good agreement with the best fitting value determined from clustering in the HIPASS sample, with Wyithe et al. (2009a) reporting a slope of 0.7 ± 0.4.

Comparing HODs for optical and cold gas mass selection

We next compare the model predictions for the HOD of an optically selected galaxy sample with those of cold gas mass selected samples. Fig. 4.8 shows the HOD for samples defined by a cold gas mass threshold of $10^{10}h^{-2}M_{\odot}$ in the first four columns, with each column showing the predictions for a different model. In the right hand column, we plot the HOD for a sample in which galaxies are selected on the basis of their r-band luminosity in the GpcBow06 model. The optical luminosity cut is chosen such that the galaxies brighter than the limit ($M_r - 5\log h < -21.06$) have the same number density as the sample selected by cold gas mass in the GpcBow06 model. As we have already remarked, the HODs for the cold gas samples have similar properties, with a peaked HOD for central galaxies which declines rapidly with increasing halo mass, and a power law HOD for satellites. The HOD for central galaxies in the optical sample shows a local bump for haloes masses just below $10^{12}h^{-1}M_{\odot}$, but overall rises gradually, reaching unity at a halo mass of $\sim 3 \times 10^{14}h^{-1}M_{\odot}$. The bump is due to the implementation of AGN feedback. The central galaxy HOD drops after the bump as AGN feedback "switches on" in these haloes. Central galaxies hosted by massive haloes are bright in the r-band, whilst possessing too little gas to be included in the cold gas sample.

The remaining rows of Fig. 4.8 show the steps which connect the HOD predictions to the effective bias of the galaxy samples, which tells us the clustering amplitude. In the lower two rows of this plot we have switched to plotting quantities on a linear scale. In the second row of Fig. 4.8, the HOD is multiplied by the abundance of the host dark matter haloes, giving the contribution to the number density of galaxies as

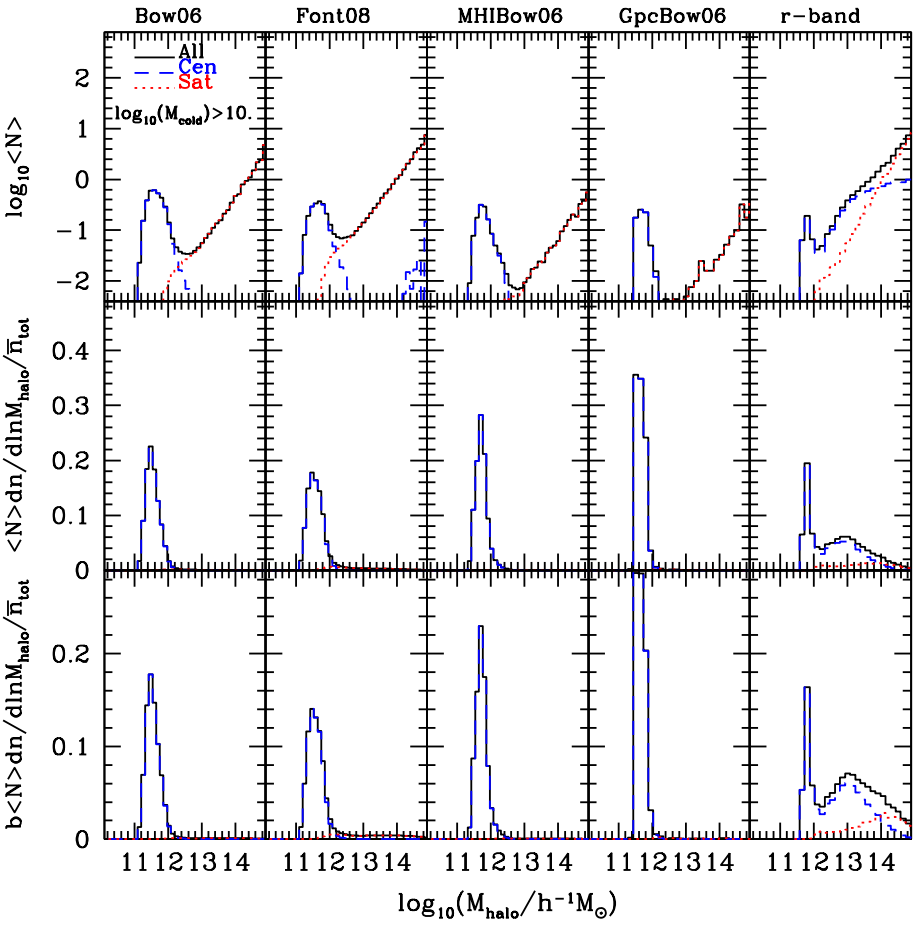

Figure 4.8: The steps relating the number of galaxies per halo to the strength of galaxy clustering in the GALFORM models. The first row shows the HOD as a function of halo mass. The second row shows the HOD multiplied by the abundance of dark matter haloes as a function of halo mass, $dn/d\ln M_{halo}$, (as computed using the prescription of Sheth, Mo & Tormen 2001) with the y-axis plotted on a linear scale. $\bar{n}_{\rm tot}$ is the number density of galaxies which satisfy the selection criteria (i.e. in cold gas mass or r-band luminosity). The integral of these curves is proportional to the number density of galaxies. The bottom row shows the HOD times the halo mass function times the bias factor as a function of halo mass . The area under the curves in this case gives the effective bias of the galaxy sample. The first four columns show the model predictions for galaxies with cold gas mass in excess of $\rm M_{cold} > 10^{10}h^{-2}M_{\odot}$. The fifth column shows an r-band selected sample in the GpcBow06 model, with the magnitude limit ($M_r - 5\log h < -21.06$) chosen such that the number of galaxies matches that in the cold gas sample in this model. As before, the contribution of central galaxies is shown by blue dashed lines, satellite galaxies by red dotted lines and all galaxies by black solid lines.

a function of halo mass. The abundance of the host dark matter haloes is computed using the prescription of Sheth, Mo & Tormen (2001), which gives a good match to simulation results. Beyond the break in the mass function, the number of haloes per unit volume drops exponentially. This means that satellite galaxies, whose HOD is described by a moderate power law, do not contribute significantly to the number of galaxies per unit volume. This is true for samples defined either by cold gas mass or r-band luminosity. The abundance of galaxies is sharply peaked for the cold gas samples. For the optical sample, the galaxy number density has a sharp peak just below a halo mass of $10^{12}h^{-1}M_{\odot}$ and then shows a broad distribution and an appreciable contribution from more massive haloes. In the bottom row of Fig. 4.8, we plot the number density of galaxies multiplied by the bias factor as a function of halo mass (as computed using the prescription of Sheth, Mo & Tormen 2001). The square of the bias gives the factor by which the auto-correlation function of haloes is boosted on large scales relative to the correlation function of the dark matter. The halo bias increases rapidly beyond the break in the mass function, which increases the influence of satellite galaxies on the effective bias (e.g. Angulo et al. 2008b). Nevertheless, for the cold gas mass samples satellite galaxies still make a negligible contribution to the clustering amplitude on large scales, as quantified by the effective bias. Satellite galaxies make a modest contribution to the effective bias in the r-band sample. This contribution increases if the luminosity cut is made fainter. In summary, the models predict that galaxies with cold gas mass in excess of $10^{10}h^{-2}\mathrm{M}_{\odot}$ are predominately central galaxies hosted by dark matter haloes of mass $10^{12}h^{-1}\mathrm{M}_{\odot}$. These haloes are less massive than the characteristic halo mass at $z = 0$ in the cosmologies used and so the bias factor of these samples is below unity; they are sub-clustered compared to the dark matter. In contrast, the r-band sample has an effective bias with a significant contribution from more massive haloes which have a larger bias factor. The bias factor for the r-band selected samples is therefore greater than unity and clustering length is larger than it is for cold gas sample (see Fig. 4.10 later).

Finally, in Fig. 4.9 we compare the spatial distribution of r-band selected galaxies with that of galaxies chosen on the basis of their cold gas mass ($\mathrm{M}_{\mathrm{cold}} > 10^{10}h^{-2}\mathrm{M}_{\odot}$

in the GpcBow06 model). Again the r-band magnitude limit ($M_r - 5\log h < -21.06$) is chosen to match the abundance of galaxies in the cold gas sample. The grey circles represent dark matter haloes. The circle radius and darkness are proportional to halo mass. The cold gas selected galaxies follow the filamentary structure and tend to avoid high density regions. The difference in the number of satellite galaxies (red circles) is obvious between the cold gas and optical samples. The satellites are found in more massive haloes. This difference in the spatial distributions provides a visual impression of the differences in the HODs plotted in Fig. 4.8. The stronger clustering of the optical samples in principle means that it should be easier to measure the power spectrum of galaxy clustering using these tracers. However, the key consideration, as we shall see in Section 4.4, is how the product of the number density of galaxies and their power spectrum amplitude changes with redshift. This quantity controls the "contrast" of the power spectrum signal against the noise which arises from having discrete tracers of the density field.

4.3.2 Predictions for the clustering of cold gas

In this section we present the predictions of the galaxy formation models introduced in Section 4.2 for the two point correlation function. To predict the galaxy distribution of the GpcBow06 model, we generated galaxy samples using the GPICC simulation.

We start in Fig. 4.10 by comparing the spatial two point autocorrelation function of a galaxy sample defined by a threshold cold gas mass ($M_{\rm cold} > 10^{10} h^{-2} M_\odot$) in real (solid black line) and redshift space (dashed black line). The correlation function is computed in redshift space using the distant observer approximation. In this approximation, one of the coordinate axes is chosen as the line of sight and the peculiar velocity of the galaxy in that direction is added to the real space position, after applying a suitable scaling to convert from velocity units to distance units. For the largest pair separations plotted, the correlation function in redshift space has the same shape as the real space correlation function, but a larger amplitude. The magnitude of the shift in amplitude agrees very closely with the expectation of Kaiser (1987). This effect is caused by coherent bulk flows towards overdense regions. On

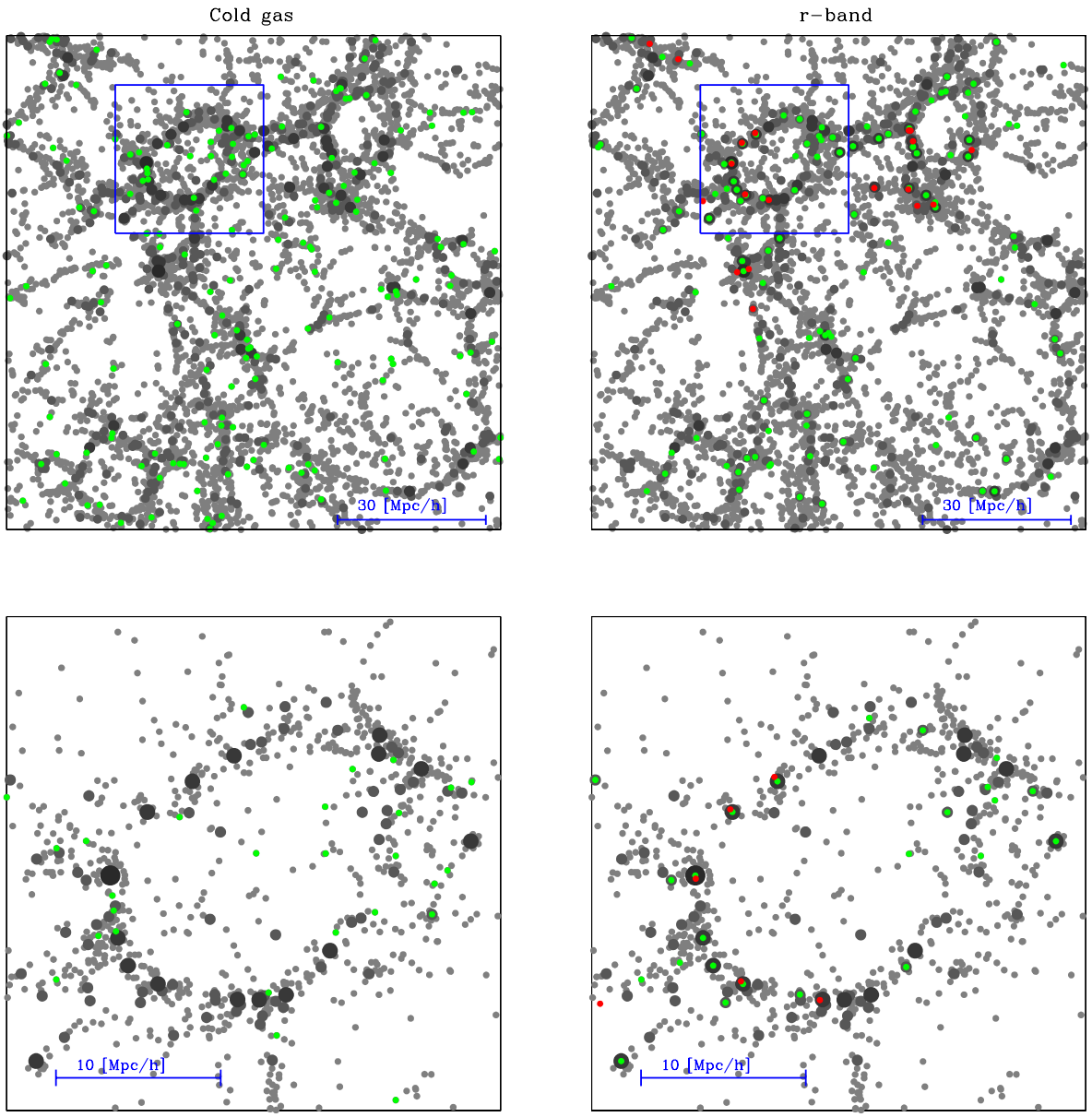

Figure 4.9: The spatial distribution of galaxies and dark matter haloes in the GpcBow06 model at $z = 0$. Dark matter is shown in grey and the size and darkness of the circle used to plot the dark matter halo increase with mass. Galaxies selected by cold gas mass ($M_{\rm cold} > 10^{10}h^{-2}M_{\odot}$) and r-band luminosity ($M_r - 5\log h < -21.06$) are plotted in the left and right hand panels respectively. The top row shows a slice of 100 h^{-1}Mpc on a side and $10h^{-1}$Mpc thick. The bottom row shows a zoom into a region of $30h^{-1}$Mpc on a side and $10h^{-1}$Mpc thick, which corresponds to the blue square in the top row. The green circles represent central galaxies and the red circles show satellite galaxies.

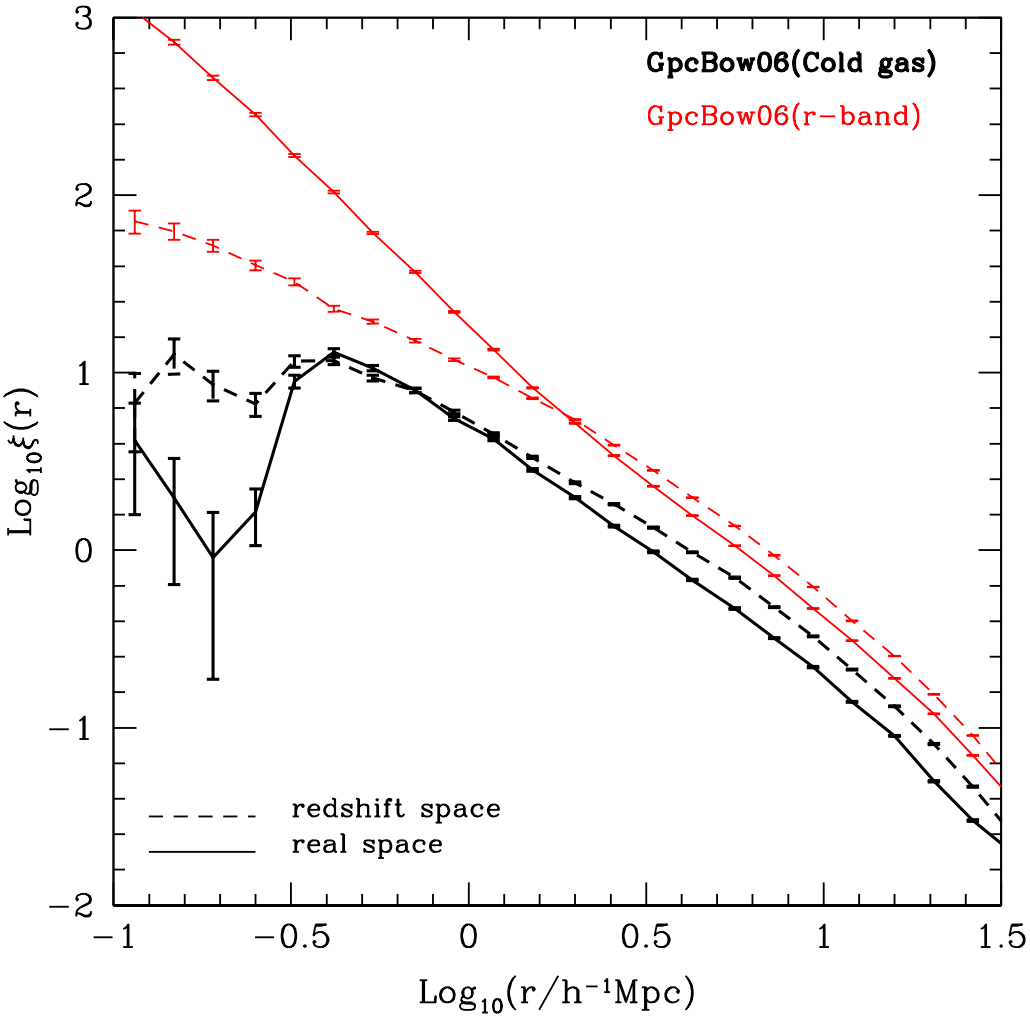

Figure 4.10: The real space (solid) and redshift space (dashed) correlation function predicted for galaxies in the GpcBow06 model at $z = 0$. The black lines show the correlation function of galaxies with cold gas mass $M_{\rm cold} > 10^{10}h^{-2}M_{\odot}$ and the red lines show the clustering of galaxies selected to be brighter than a threshold r-band luminosity, with the limit chosen to match the abundance of galaxies in the cold gas sample. The errorbars show the Poisson error on the pair count in each bin of radial separation.

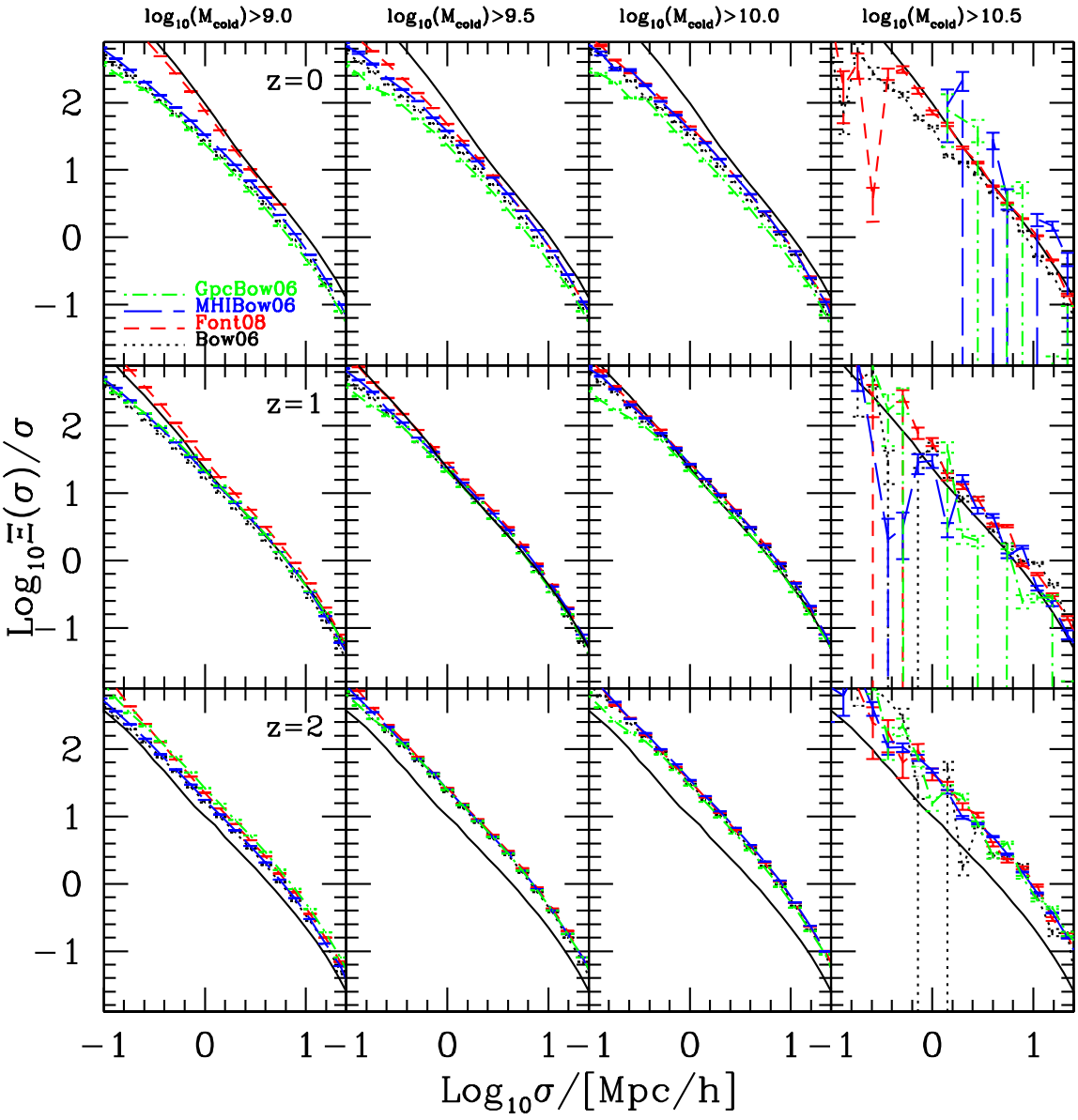

Figure 4.11: The projected correlation function for cold gas mass selected samples at $z = 0$ (top), 1 (middle) and 2 (bottom). Each column shows the predictions for a different cold gas mass threshold, as indicated by the label. The predictions of the models are distinguished by different line types and colours, as shown by the key in the upper left panel. The solid black lines in each panel show the projected correlation function of the dark matter measured in the Millennium simulation (note that the GpcBow06 model uses a different cosmology and has different dark matter correlation functions).

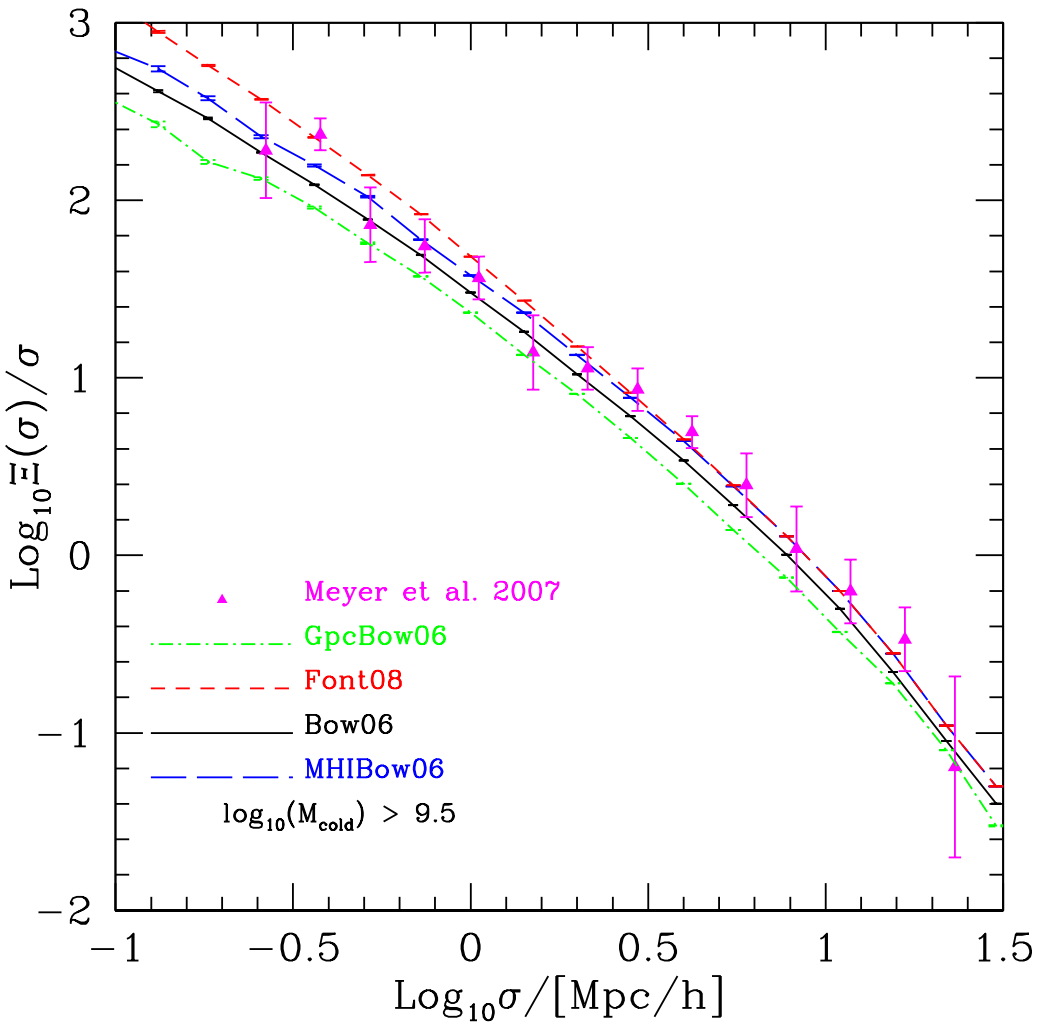

Figure 4.12: The projected galaxy correlation function at $z = 0$. The points with errorbars show an observational estimate made from the HIPASS catalogue by Meyer et al. (2007). The lines show the model predictions for galaxies more massive than $M_{\rm cold} > 10^{9.5} h^{-2} M_{\odot}$, a threshold chosen to match the selection of galaxies in the HIPASS sample. The results for different models are shown by lines with different colours and line types, as indicated by the key.

pair separations between 0.3 and $1h^{-1}$Mpc, the real and reshift space correlation functions are very similar. They diverge on smaller scales, where the predictions are noisy simply because there are few galaxies pairs at these separations.

This behaviour can be contrasted with the clustering in the optically selected sample, which is shown by the red lines in Fig. 4.10. As with the cold gas sample, there is a shift in the clustering amplitude when measured in redshift space for pair separations $r > 3h^{-1}$Mpc. However, the size of the shift is smaller for the optically selected sample, which is consistent with the bias of this sample being greater than unity and larger than the bias of the cold gas selected sample. The real-space correlation function of the optical sample is steep on small scales, reflecting the contribution of satellite galaxies within common dark matter haloes. There is a substantial reduction in the clustering amplitude in redshift space on these scales in the optical sample, again driven by satellite galaxies. This is the so-called "fingers of God" redshift space distortion, whereby randomised peculiar velocities of the satellites within the gravitational potential of the cluster make the cluster appear elongated.

The real space correlation function cannot be estimated directly from a galaxy redshift survey. A related quantity is the projected correlation function which can be estimated from the two point correlation function measured in bins of pair separation parallel (π) and perpendicular (σ) to the line of sight, $\xi(\sigma, \pi)$ (e.g. Norberg et al. 2001):

$$\frac{\Xi(\sigma)}{\sigma} = \frac{2}{\sigma} \int_0^\infty \xi(\sigma, \pi) \mathrm{d}\pi. \tag{4.5}$$

In the limit that the integral over the radial pair separation can be taken to infinity, this quantity is free from redshift space distortions (see Norberg et al. 2009 for an illustration of the impact of imposing a finite upper limit on the integral).

Fig. 4.11 shows the projected correlation function predicted in the four models for a range of cold gas mass samples at $z = 0$, 1 and 2. The columns show the results for different cold gas mass thresholds, and the rows correspond to different models. The solid black lines in each panel show the projected correlation function measured for the dark matter in the Millennium Simulation (recall that the GpcBow06 model has a different cosmology and so should be compared to a consistent dark matter

correlation function which will be slightly different from that in the Millennium simulation on these scales). Overall, the three lowest mass samples at $z = 0$ are less clustered than the dark matter. The most massive threshold sample we consider at this redshift has a similar clustering amplitude to the dark matter. At $z = 1$, the bias of the three lowest mass samples is close to unity, with the projected clustering of galaxies being very close to that of the dark matter. At $z = 2$, the cold gas samples are more clustered than the dark matter and correspondingly have effective biases greater than unity. This evolution in the bias is due to the adoption of a fixed cold gas mass threshold. At high redshift, galaxies with a large cold gas mass will tend to be found in more massive haloes.

Across the different models there is a small spread in clustering amplitude for a given cold gas mass sample, with remarkably similar predictions made for the projected correlation function. Fig. 4.11 shows that the differences start to appear at $z = 1$ and become larger by $z = 0$. The model which shows the largest difference from the others is Font08. On small scales in the two lowest mass threshold samples, this model has an appreciably higher amplitude projected correlation function than the other models. This feature can be traced back to the HODs plotted in Fig. 4.5. Due to the revised cooling model used in Font08, there are more satellite galaxies in the low mass samples in this model, which boosts the one halo term in the correlation function.

Finally, we compare the predicted correlation functions with an observational estimate from Meyer et al. (2007), which was made using the HI Parkes All-Sky Survey (HIPASS) Catalogue (HICAT; Meyer et al. 2004). In order to make this comparison, we need to convert the cold gas mass output by the models into an atomic hydrogen mass. We assume that 76% by mass of the cold gas is hydrogen. Here we adopt a fixed ratio of molecular (H_2) to atomic (HI) hydrogen of H_2/HI=0.4 (see Power et al. 2010 for a discussion). The HI mass, $M_{\rm HI}$, is therefore obtained from the cold gas mass $M_{\rm cold}$ by applying the conversion:

$$M_{\rm HI} = 0.76 M_{\rm cold}/(1 + 0.4). \tag{4.6}$$

With this relation, the sample analyzed by Meyer et al. is equivalent to a cold gas mass threshold of $M_{\rm cold} > 10^{9.5} h^{-2} M_{\odot}$. The comparison between the model pre-

dictions and the observational estimate is presented in Fig. 4.12. The correlation function predicted by the MHIBow06 model agrees remarkably well with the observational estimate. The GpcBow06 and Bow06 models predict too low a clustering amplitude. The Font08 model gives a reasonable match on intermediate and large scales, but somewhat overpredicts the clustering amplitude on small scales, hinting that this model has too many gas rich satellites in massive haloes.

A model for variable H_2/HI ratio in a galaxy is derived by Obreschkow et al. (2009) :

$$M_{\rm HI} = 0.76 M_{\rm cold}/(1 + R_{mol}^{gal}), \tag{4.7}$$

where $R_{mol}^{gal} = M_{H_2}/M_{HI}$:

$$R_{mol}^{gal} = \left(3.44 {R_{mol}^{c}}^{-0.806} + 4.82 {R_{mol}^{c}}^{-1.084}\right)^{-1}, \tag{4.8}$$

where R_{mol}^{c} is the H_2/HI ratio at the centre of the disk, given by

$$R_{mol}^{c} = [K r_{disk}^{-4} M_{gal}(M_{gas} + \langle f_\sigma \rangle M_{disk}^{\star})]^{0.8}, \tag{4.9}$$

where $M^\star$ and M_{gas} are the masses of stars and gas in the disk, and r_{disk} is the exponential scale length of the gas. The adopted values of K is $11.3 \mathrm{m}^4\mathrm{kg}^{-2}$ and $\langle f_\sigma \rangle$ is 0.4.

4.4 Measuring dark energy with future HI redshift surveys

In this section we show how redshift surveys of HI selected galaxies can be used to detect baryonic acoustic oscillations (BAO) in the galaxy power spectrum (Sec 4.4.1), and we assess the relative performance of HI and optical surveys in measuring the large scale structure of the Universe, first by contrasting the effective volumes sampled by different survey configurations (Sec 4.4.2) and then by applying a Fisher matrix calculation to compare the error forecast on the equation of state parameter for different surveys (Sec 4.4.3).

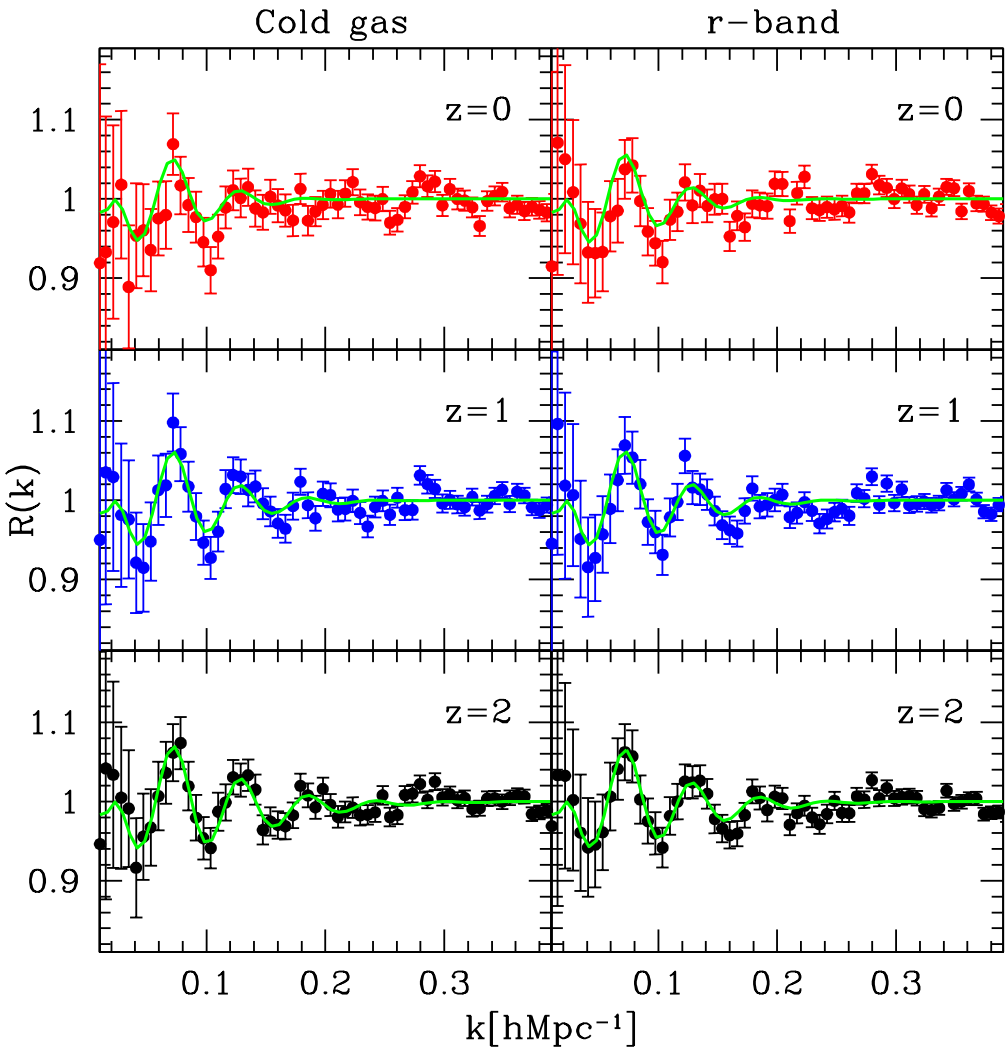

Figure 4.13: The baryonic acoustic oscillations in the galaxy power spectrum. To display the BAO more clearly, we have divided the predicted spectra by smooth fits, as described in the text. The points show the power spectra predicted by the GpcBow06 model at $z = 2$ (bottom), $z = 1$ (middle) and $z = 0$ (top). The left hand column shows the power spectra measured for galaxies with cold gas mass ($M_{cold} > 10^{10} h^{-1} M_{\odot}$). The right hand columns shows the BAO in a sample selected in the r-band with the same number density of galaxies as the cold gas sample at that redshift. The smooth green line shows the linear theory power spectrum, divided by a smooth reference power spectrum, after filtering or "de-wiggling" to damp the higher harmonics (see text). The errors plotted on the power spectrum depend on the number density of galaxies and the simulation volume (see eq. 3 in Angulo et al. 2008a).

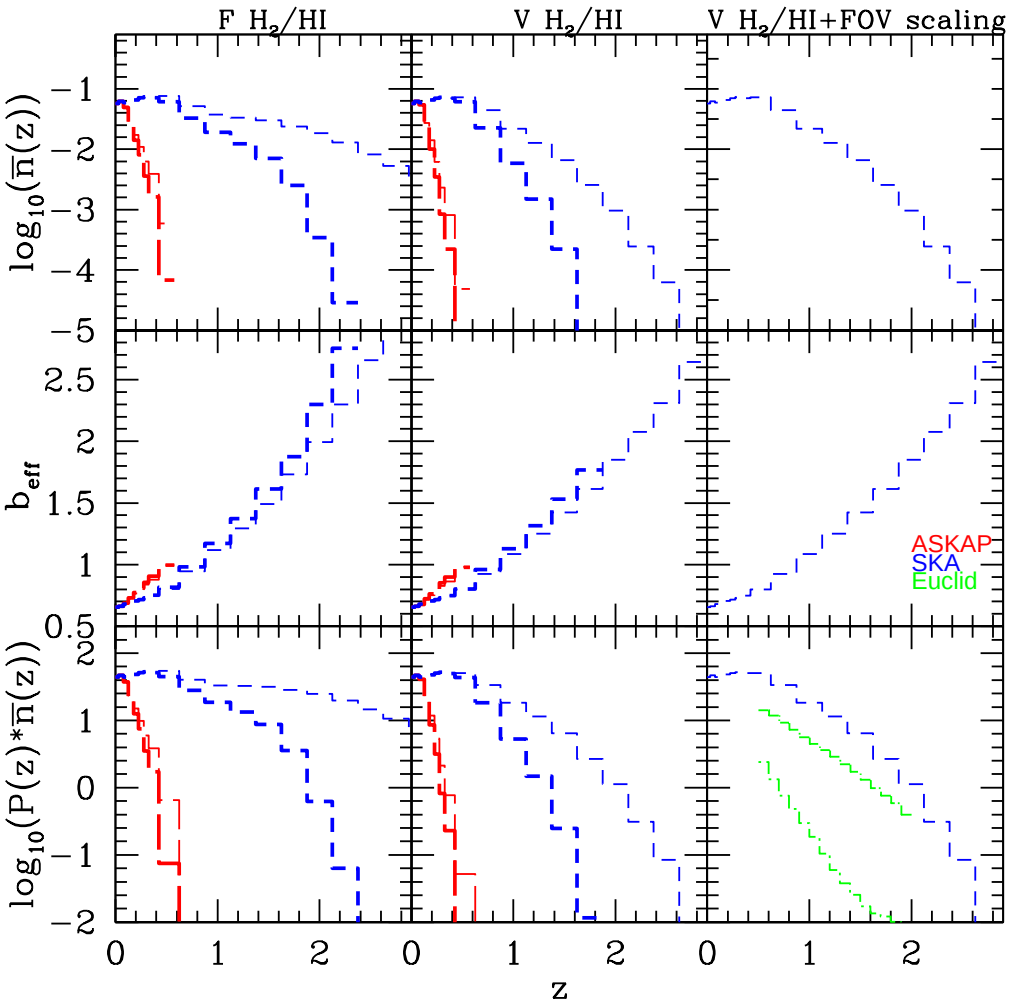

Figure 4.14: The quantities needed to compute the effective volume of a redshift survey, as predicted in the GpcBow06 model. We show predictions for the fiducial SKA configuration (blue) for ASKAP (red) and for Euclid (green; lower right panel). The top row shows the number density of galaxies as a function of redshift, the middle row the effective bias of the galaxy samples and the bottom row shows the product of the galaxy number density and galaxy power spectrum, $P_{\rm gal}(k)$ at $k = 0.2h\,{\rm Mpc}^{-1}$. The first column shows the predictions in the case that the ratio H_2/HI is held fixed and the second column shows how the predictions change when this ratio is allowed to vary between galaxies and with redshift. The thin lines show the results when the FoV is assumed to scale with redshift; the lines corresponding to FoV scaling with redshift and a variable H_2/HI ratio are shown in the third column. Predictions are shown for the fiducial SKA (blue) and ASKAP (red) configurations. The green curves show the predictions for a spectroscopic survey down to $H = 22$ (assuming a 33% redshift success rate; green dot-long-dashed line) and a slitless survey of Hα down to a flux limit of 5×10^{-16}erg s^{-1} cm^{-2}, again with a 33% redshift measurement rate (green dot-short-dashed line); both these results are taken from Orsi et al. (2009).

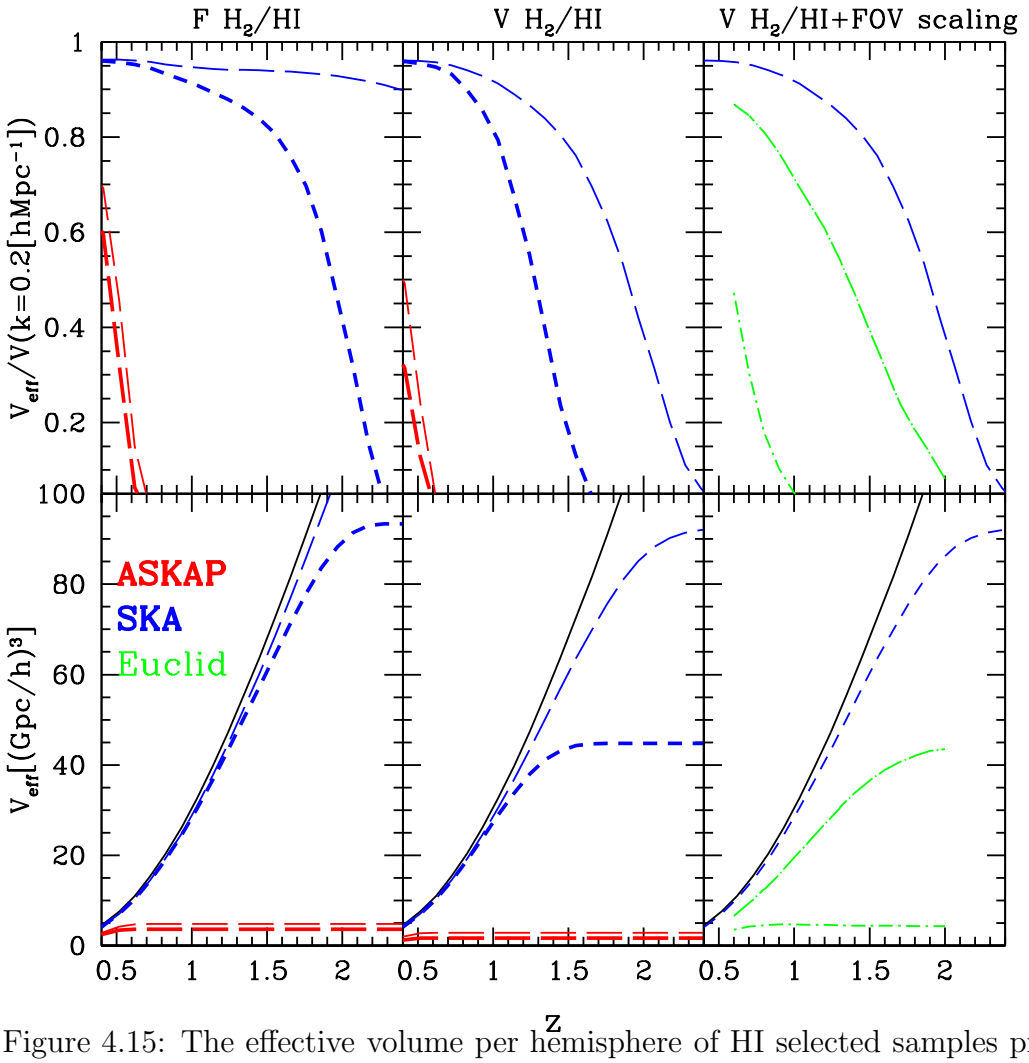

Figure 4.15: The effective volume per hemisphere of HI selected samples predicted by the GpcBow06 model. The upper panel shows the differential effective volume divided by the geometrical volume for narrow bins in redshift. The lower panel shows the cumulative volume out to a given redshift. The results for different telescope configurations are shown by different line styles and colours as indicated by the key. The green curves show the predictions for a spectroscopic survey down to $H = 22$ (33% redshift success rate; green dot-long-dashed line) and a slitless survey down to an Hα flux limit of 5×10^{-16} erg s^{-1} cm^{-2} (green dot-short-dashed line), as computed by Orsi et al. (2009). The black solid line in the lower panel shows the available geometrical volume per hemisphere. The optical and HI surveys are assumed to cover approximately the same solid angle, one hemisphere. The thick curves show the effective volume calculated for a telescope sensitivity given by Eq. 4.14. The thin lines show the results when the integration time scales as $(1 + z)$ due to the oversampling of the FoV by individual dishes.

4.4.1 The appearence of baryonic acoustic oscillations

The BAO signal measured in a sample defined by cold gas mass is shown in Fig. 4.13. We use the galaxy distribution in the GPICC simulation generated using the GpcBow06 model. The GPICC simulation box is $1h^{-1}$Gpc on a side which allows the evolution of the BAO to be modelled accurately. To show the BAO more clearly, we have divided the measured spectrum by a smooth reference power spectrum which contains no BAO features. This "no wiggle" power spectrum is defined differently for the linear theory power spectrum and for the spectra measured from the GPICC simulation. For the linear theory prediction, which is shown by the curves in Fig. 4.13, the reference is based on the "no wiggle" parametrization of the power spectrum given by Eisenstein & Hu (1998). The no wiggle prediction includes the impact of a non-zero baryon component on the width of the turn-over in the matter power spectrum, but does not contain BAO. The ratio of the linear theory power spectrum, $P^{\rm L}(k)$, to the no wiggle prediction, $P^{\rm L}_{\rm nw}$, is "de-wiggled" by damping the oscillations to represent the impact of nonlinear growth and redshift-space distortions (e.g. Blake & Glazebrook 2003; Eisenstein et al. 2005; Sanchez, Baugh & Angulo 2008):

$$R_{\rm lin}(k) = \left(\frac{P^{\rm L}}{P^{\rm L}_{\rm nw}} - 1\right) \times \exp\left(-\frac{k^2}{2k^2_{\rm nl}}\right) + 1, \tag{4.10}$$

where $k_{\rm nl}$ is the damping scale and is treated as a free parameter.

The overall shape of the power spectra measured from the simulation is different from the linear theory prediction due to the nonlinear growth of fluctuations and redshift-space distortions (see Angulo et al. 2008a for a step by step illustration of these effects). We model this change in shape by multiplying the no-wiggle version of the linear theory spectrum by a third order polynomial:

$$P_{\rm g}(k) = (1 + Ak + Bk^2 + Ck^3)P^{\rm L}_{\rm nw}(k). \tag{4.11}$$

The free parameters A, B and C are chosen to give the best match to the overall shape of the measured power spectrum. All points up to $k = 0.4h$Mpc^{-1} were included in the fit and given equal weight. This approach is more straightforward and robust than using a spline fit to a coarsely binned measured spectrum, which is sensitive to the number of k-bins used.

We show in Fig. 4.13 the BAO signal in the GpcBow06 model at $z = 0, 1$ and 2 for galaxies selected by their cold gas mass ($\mathrm{M_{cold}} > 10^{10}h^{-1}\mathrm{M_{\odot}}$; left column), as an illustration of how a cold gas mass selected sample traces this large-scale structure feature. In the right-hand column of Fig. 4.13, we compare this with the BAO signal expected for an r-band selected sample of galaxies that have same number density at each redshift as the cold gas sample. The reference power spectrum is defined as described above, using the third order polynomial fit to the measured spectrum in each case. Fig. 4.13 shows that we should be able to measure the BAO feature just as well using a sample selected by cold gas mass as with an optically selected sample.

4.4.2 The effective volumes of different survey configurations

Many ongoing and proposed redshift surveys have the goal of determining the nature of dark energy by measuring the BAO signal in the galaxy power spectrum. A powerful way to compare the expected performance of different surveys for measuring large-scale structure is to estimate their effective volumes (see, for example, Orsi et al. 2009). This is essentially an indicator of the "useful" survey volume which determines the size of the errorbar on the measured power spectrum. The effective volume is defined as (Feldman, Kaiser & Peacock 1994)

$$V_{\mathrm{eff}}(k,z) = \int_{z_{\mathrm{min}}}^{z_{\mathrm{max}}} \left[\frac{\bar{n}(z)P_{\mathrm{g}}(k,z)}{1+\bar{n}(z)P_{\mathrm{g}}(k,z)}\right]^2 \frac{\mathrm{d}V}{\mathrm{d}z}\mathrm{d}z \qquad (4.12)$$

where all quantities are expressed in comoving coordinates and $\mathrm{d}V/\mathrm{d}z$ is the differential comoving volume. To calculate the effective volume, we therefore need to know the number density of galaxies ($\bar{n}(z)$) down to a given survey flux limit and the effective bias ($b(z)$), both as functions of redshift. In this calculation, we obtain the galaxy power spectrum using the linear relation between galaxy bias and the dark matter power spectrum: $P_{\mathrm{g}}(k,z) = P_{\mathrm{dm}}(k,z=0)b^2(z)D^2(z)$, where P_{g} is the galaxy power spectrum, $P_{\mathrm{dm}}(k,z=0)$ is the linear theory dark matter power spectrum at $z = 0$, $b(z)$ is the effective bias, and $D(z)$ is the growth factor of the dark matter.

To make predictions for the effective volume of the SKA, we need to convert the

cold gas mass predicted by the models into an HI line flux, which we do following the prescription set out in Power et al. (2010). A key step is the assumption about the fraction of neutral hydrogen which is in molecular form as opposed to atomic hydrogen. Power et al. (2010) adopted two prescriptions: a fixed fraction of 40% as used by Baugh et al. (2005) and a variable fraction, as used by Obreschkow & Rawlings (2009a). In the latter case, the H_2/HI ratio varies from galaxy to galaxy as a power-law in gas pressure, as found observationally by Blitz & Rosolowsky (2006). We shall refer to these two scenarios as the fixed and variable H_2/HI ratio cases. Power et al. (2010) showed that the high redshift tail of the count distribution in the variable H_2/HI case is substantially suppressed compared with the fixed H_2/HI ratio case.

Following Power et al. (2010) the observed flux from a galaxy at 21 cm is given by:

$$S_{\rm obs} = \frac{3}{16\pi}\frac{hcA_{12}}{m_{\rm H}}M_{\rm HI}\frac{1}{D_L^2(z)}\frac{1}{\Delta V_{\rm los}}(1+z), \tag{4.13}$$

where A_{12} is the Einstein coefficient which gives the spontaneous rate of transition from the upper to lower hyperfine levels of the hydrogen ground state, $m_{\rm H}$ is the mass of the hydrogen atom, $M_{\rm HI}$ is the mass of HI in the galaxy, $D_L(z)$ is the luminosity distance to redshift z and $\Delta V_{\rm los}$ is the rest-frame line-of-sight velocity width of the galaxy. We assume the line width of the 21 cm emission is sampled by many channels and these are combined to yield the flux integrated over the full rotationally broadened width of the line. The model disks are assumed to have random inclinations. The sensitivity of a single dual polarization radio receiver can be written in terms of a limiting *rms* flux as

$$\frac{S_{\rm rms}}{1.626\mu{\rm Jy}} = \left(\frac{A_{\rm eff}}{{\rm km}^2}\right)\left(\frac{T_{\rm sys}}{50{\rm K}}\right)\left(\frac{\Delta\nu_{\rm rec}}{{\rm MHz}}\right)^{-1/2}\left(\frac{\tau}{\rm hr}\right)^{-1/2}, \tag{4.14}$$

where $A_{\rm eff}$ is the effective area of the telescope, $T_{\rm sys}$ is the system temperature which is a measure of the instrument noise, $\Delta\nu_{\rm rec}$ is the frequency bandwidth and τ is the integration time (in hours).

The instantaneous field of view of a radio telescope is proportional to the wavelength of observation. To detect emission at 21 cm in the rest frame of a galaxy which is at $z > 0$, the telescope samples a longer wavelength in the observer frame

with a correspondingly larger field of view. In the case of a telescope made up of multiple dishes, a survey strategy can be devised which takes advantage of the overlap between the fields of view of the individual telescopes. This results in the effective integration time of the survey *increasing* with redshift in proportion to $1+z$. Hence, in the case of multiple dishes, the sensitivity of the telescope can be modified to by rescaling the integration time in Eq. 4.14 from t to $t(1+z)$.

We base the specifications of the next generation radio telescopes on the SKA "strawman" presented by Carilli & Rawlings (2004; see also Abdalla, Blake & Rawlings 2010). We assume an effective area to system temperature ratio of $A_{\rm eff}/T_{\rm sys} = 20000{\rm m}^2{\rm K}^{-1}$ (i.e. for $A_{\rm eff} = 1{\rm km}^2$, this corresponds to $T_{\rm sys} = 50{\rm K}$). We adopt a fiducial field of view (FoV) of 100 square degrees. We consider survey durations of 1 year. For a one year survey, 20 000 degrees can be covered in 40 hour integrations with the above fiducial specification. We also consider an ASKAP configuration which corresponds to a FoV of 30 square degrees and an effective area to system temperature ratio of $A_{\rm eff}/T_{\rm sys} = 82{\rm m}^2{\rm K}^{-1}$ (Johnston et al. 2007).

Fig. 4.14 shows the model predictions for the quantities needed to calculate the effective volume for various telescope configurations, calculated using the GpcBow06 model. The predictions are divided into three columns to show the impact of different assumptions. A comparison of the first and second columns of Fig. 4.14 shows how the assumption about the H_2/HI ratio affects the predictions: the first column shows the predictions for a fixed H_2/HI ratio and the second column shows the output for a variable ratio. The thick curves show the predictions using a sensitivity given by Eq. 4.14 whereas the thin lines include the scaling of the integration time with redshift mentioned above (due to the increase in the single dish FoV with redshift). The third column shows the predictions with a variable H_2/HI ratio and a FoV which depends on redshift. The number density of galaxies in the ASKAP configuration falls rapidly with redshift (red lines), independently of the use of a FoV scaling with redshift. The decline in the abundance of ASKAP galaxies is slightly more modest in the case of a fixed H_2/HI ratio as opposed to a variable ratio. In the case of the SKA configuration, the number density of galaxies drops by around a factor of two by $z = 1$ in the fixed H_2/HI ratio case. The middle panels

of Fig. 4.14 show that the bias changes by a much more modest amount than the number density of galaxies does, increasing by factor of four with redshift over the redshift range plotted. The increase in effective bias cannot therefore compensate for the dramatic drop in the abundance of galaxies in the high redshift tails of the distributions plotted in the upper panels of Fig. 4.14. The effective volume of a survey configuration no longer increases with redshift once the product of the galaxy number density and the galaxy power spectrum drops below unity. In this regime, the power spectrum signal is swamped by shot noise ($P_{\rm shot} = 1/\bar{n}$) arising from the use of discrete galaxies to trace the continuous density field, and this volume does not contribute to the statistical power of the survey. The product $\bar{n}P$ is plotted in the lower panels of Fig. 4.14.

In the lower right panel of Fig. 4.14, we include two predictions for redshift surveys conducted in the near-infrared taken from Orsi et al. (2009), who followed the same procedure we have set out above, but for different galaxy selection criteria. Orsi et al. consider surveys covering 20 000 square degrees, with two different constructions. The predictions for a redshift survey to $H = 22$ (the DMD based slit spectrometer) with a 33% redshift sampling rate (green dot-long-dashed line) and for a slitless survey to an Hα flux limit of 5×10^{-16} erg s^{-1} cm^{-2}, again with a 33% redshift measurement rate (green dot-short-dashed line), are plotted for comparison. Leaving aside cost considerations and technical feasibility, this comparison shows that the case of FOV=100deg^2(1-yr) with variable H_2/HI has around two times more effective volume than the volume of the one third sampling $H = 22$ survey, and samples over 10 times the volume of the Hα survey (see more detail description for spectrometer in Laureijs et al. 2009).

Our predictions can be compared with those of Abdalla, Blake & Rawlings (2010). These authors adopt a slightly different definition of the *rms* flux limit, based on a study of how the sky background, and hence $T_{\rm sys}$ varies with frequency (F. Abdalla, private communication). In practice, our predictions for a $10-\sigma$ flux limit should be compared with the $5-\sigma$ predictions of Abdalla et al. These authors also assume that the velocity width which appears in the definition of the flux limit (see their Eq. 3) corresponds to a channel width (typically $\Delta V = 30{\rm km\,s^{-1}}$),

whereas we assume the full width of the line emission is sampled by many channels. Abdalla et al. make empirical assumptions about the evolution of the HI mass function with redshift. The predictions of our semi-analytical model with a variable H2/HI ratio are similar to their model C. Lastly, Abdalla et al. assume an effective bias of unit for HI emitter, independent of redshift. Our predictions for the effective survey volume are similar given these caveats.

The effective volumes for the different survey configurations are plotted in Fig. 4.15. The upper panel shows the fraction of the geometrical volume in differential redshift shells that is sampled by the effective volume probed by the different surveys and the lower panels show the cumulative effective volumes. The surveys track the geometrical volume available until the redshift at which $\bar{n}P < 1$. This is clear from the lower panel of Fig. 4.15, in which the effective volume curves flatten once this redshift is reached. The effective volume covered by the fiducial design of the SKA is substantially larger than that expected from the Euclid survey, provided that the SKA FoV scales with redshift and a fixed H_2/HI ratio is assumed. In the event that a variable H_2/HI ratio is adopted, then the Euclid and SKA effective volumes are comparable.

4.4.3 The forecast error on the dark energy equation of state

The effective volume gives a broad brush view of the potential performance of a survey. In order to get a more quantitative impression, we need to make a forecast of the error on the parameter of interest, which in our case is the dark energy equation of state parameter, w. This will allow us to assess if the volume sampled by the survey is at a redshift which is useful for constraining the value of w. The conclusions will depend to some extent on the dark energy model adopted. The fiducial model we use is a flat cold dark matter universe with a cosmological constant. The cosmological constant has little influence over cosmological distances above $z \approx 1.5$-2. Hence, a difference in effective volume between survey configurations at these redshifts is likely to have little impact on how well w can be measured. This behaviour could change if we adopted a different dark energy model, such as one with appreciable amounts of dark energy at early epochs (see, for example, the plots

of Hubble parameter and luminosity distance in Jennings et al. 2010).

To make the forecast of the error on w for a particular survey configuration, we use a Fisher matrix approach, closely following the calculation in Seo & Eisenstein (2003). The observable quantity is the ratio of the measured power spectrum, including the BAO signal, to a smooth reference spectrum, which has no BAO features, defined as outline in Sec 4.4.1. The Fisher matrix encodes the sensitivity of this ratio to different cosmological parameters, assuming Gaussian errors. Our goal is to compare the different survey configurations, so we use a number of approximations to simplify the calculation. In particular, we work in the flat sky approximation, ignore the impact of redshift space distortions on the appearance of the BAO and neglect any evolution of the power spectrum over bins of redshift of width 0.1. Under these assumptions, the Fisher matrix (for arbitrary parameters) obtained from the power spectrum is given by (Tegmark, Taylor & Heavens 1997; Seo & Eisenstein 2003),

$$\begin{aligned} F_{ij} &= \sum_{i=1}^{N_z} \int_{k_{\rm min}}^{k_{\rm max}} \frac{\partial \ln R(k,z_i)}{\partial p_i} \frac{\partial \ln R(k,z_i)}{\partial p_j} \\ &\quad \times V_{\rm eff}(k,z_i) \, \frac{4\pi k^2 dk}{2(2\pi)^3}, \end{aligned} \tag{4.15}$$

where R is the measured power spectrum divided by a smooth reference, as given by Eq. 4.10 and the effective volume, $V_{\rm eff}(k,z)$ is given by Eq. 4.12. The integration is over the wavenumber interval $k_{\rm min} = 0.02h{\rm Mpc}^{-1}$ to $k_{\rm max} = 0.2h$ Mpc^{-1}. To isolate the cosmological constraints which come from the BAO scale, we ignore any information stored in the amplitude of the power spectrum and assume the power spectrum is sensitive to the cosmological parameters only through the observed angular and radial distance scales. The explicit dependence, as given in Seo & Eisenstein (2003) is,

$$\begin{aligned} P_{\rm obs}\left(k_{{\rm ref}_\perp}, k_{{\rm ref}_\parallel}, z\right) &= \frac{D_A^2(z)_{\rm ref}\, H(z)}{D_A^2 z\, H(z)_{\rm ref}} \\ &\quad P_{\rm true}\left(\sqrt{k_\perp^2 + k_\parallel^2}, z\right), \end{aligned} \tag{4.16}$$

where $k_{{\rm ref}_\perp} \equiv k_\perp D_A(z)/D_A(z)_{\rm ref}$ and $k_{{\rm ref}_\parallel} \equiv k_\parallel H(z)_{\rm ref}/H(z)$ relate the wavenumbers inferred via an assumed cosmological model and the true physical scales in the power spectrum. The power spectrum defined by Eq. 4.16 is then divided by a smooth reference spectrum to form the ratio R.

Table 4.2: The forecast constraints on a constant dark energy equation of state, w, when marginalizing over Ω_m for effective survey volumes corresponding to the different assumptions about the redshift scaling of the FOV and H_2/HI ratio. The numbers shown are ratios of $\sigma(w)$, the $1-\sigma$ error on w, forecast relative to those obtained for an $H\alpha$ survey predicted by Orsi et al. (2009).

	Fixed H_2/HI		**Variable** H_2/HI	
	WMAP5 prior	Planck prior	WMAP5 prior	Planck prior
ASKAP non-redshift scaling of FOV	3.963	1.657	5.710	1.711
ASKAP redshift scaling of FOV	2.591	1.539	4.352	1.674
SKA non-redshift scaling of FOV	0.245	0.293	0.386	0.453
SKA redshift scaling of FOV	0.162	0.195	0.228	0.274

Table 4.3: The forecast constraints on a constant dark energy equation of state, w, when marginalizing over Ω_m for effective survey volumes corresponding to the different assumptions about the redshift scaling of the FOV and H_2/HI ratio. The numbers shown are ratios of $\sigma(w)$, the $1-\sigma$ error on w, forecast relative to those obtained for an $H = 22$ survey predicted by Orsi et al. (2009).

	Fixed H_2/HI		**Variable** H_2/HI	
	WMAP5 prior	Planck prior	WMAP5 prior	Planck prior
ASKAP non-redshift scaling of FOV	12.405	4.377	17.870	4.518
ASKAP redshift scaling of FOV	8.109	4.065	13.620	4.422
SKA non-redshift scaling of FOV	0.768	0.775	1.210	1.197
SKA redshift scaling of FOV	0.507	0.516	0.715	0.723
Hα	0.319	0.379	0.319	0.379

Tables 4.2 and 4.3 show the relative constraints forecast on a constant dark energy equation of state, w, when marginalizing over the matter density, Ω_m, for different survey configurations. The numbers shown are the ratios of $\sigma(w)$, the $1-\sigma$ error on w, forecast for different radio surveys relative to those forecast for the $H\alpha$ and $H = 22$ redshift surveys predicted by surveys Orsi et al. (2009) (tables 4.2 and 4.3 respectively). These calculations include a prior on Ω_m derived from a Fisher matrix for either the WMAP5 or Planck CMB missions. We marginalize the CMB constraints over the spectral index, n_s, Hubble parameter h, spectrum normalization, A_s, baryon density Ω_b, and optical depth to reionization, τ. We consider only constraints from the temperature-temperature CMB power spectra, but include an additional constraint on τ from the WMAP5 temperature-polarization cross power spectra (with no such constraint for Planck). We model only the cosmic variance and beam smearing (at 90 GHz for WMAP and 143 GHz for Planck) in the power spectrum noise models. The resulting priors are $\sigma(\Omega_m) = 0.0108$ for WMAP5 and $\sigma(\Omega_m) = 0.00203$ for Planck. Note it is not necessary to include h as a parameter in the BAO forecasts because when we consider only distance measurements, as the h dependence factors out by choosing to measure distances in units of, e.g., h^{-1} Mpc. This is why we marginalize over h to get the prior on Ω_m from the CMB.

4.5 Summary and conclusions

The cold gas content of galaxies and its variation with halo mass lie at the core of the galaxy formation process. The amount of cold gas in a galaxy is set by the balance between a number of competing processes. The cold gas supply comes from the cooling of gas from the hot halo and the accretion of cold gas following mergers with other galaxies. Star formation and supernova feedback act as sinks of cold gas. Semi-analytical simulations model all of these processes in the context of structure formation in the dark matter and so are ideally suited to make predictions for the distribution of cold gas between haloes of different mass. Since the models can make a wide range of predictions, their parameters are set by the requirement that a variety of observed galaxy properties be reproduced, not just the local HI data. The model

predictions can be tested by measurements of the clustering of HI-selected galaxy samples, and are invaluable to plan surveys to measure the large-scale structure of the Universe with the next generation of radio telescopes.

In this chapter we have compared the predictions for the distribution of cold gas in dark matter haloes of four versions of the Durham semi-analytical galaxy formation model, GALFORM. The Bower et al. (2006) and Font et al. (2008) models are publicly available from the Millennium Archive. Currently, most semi-analytical galaxy formation models predict the total cold gas mass and do not make a distinction between atomic or molecular hydrogen (for exceptions to this, see Fu et al. 2010 and Lagos et al., in preparation). In order to compare with observations of the HI content of galaxies, an assumption is needed for the molecular to atomic hydrogen ratio (Blitz & Rosolowsky 2006; Obreschkow & Rawlings 2009a). Despite this uncertainty, the Bower et al. and Font et al. models still overpredict the local abundance of galaxies as a function of their cold gas mass. This excess is straightforward to fix, with the primary adjustment made to the model star formation timescale. This modified model, based on Bower et al. (2006) is still able to reproduce the optical luminosity functions at the level enjoyed by Bower et al. We also considered a galaxy formation model set in a different cosmology, to take advantage of a N-body simulation with a large enough box size to accurately model baryonic acoustic oscillations. This model also adopted a modified star formation timescale to better match the local HI mass function.

The model predictions have several features in common. In agreement with observations, satellite galaxies are relatively unimportant in samples selected by cold gas mass. This is true even in the Font et al. (2008) model in which satellites retain some of their hot haloes, depending on their orbit within the main halo, and can hence continue to accrete cooling gas. Samples constructed according to a cold gas mass threshold are dominated by central galaxies in haloes around $10^{11}h^{-1}M_{\odot}$. The halo occupation distribution of central galaxies is peaked in halo mass, rather than being a step function as is the case for optical samples. As the cold gas mass cut is increased, the width of the central galaxy HOD increases and the amplitude drops. The peaked nature of the HOD of central galaxies is due to the suppression of

gas cooling in massive haloes following heating by AGN. We found the same general form for the HOD in an independent model by de Lucia & Blaizot (2007), in which the implementation of AGN/radio mode feedback is different from that in GALFORM.

The relative importance of central and satellite galaxies has an impact on the form of the predicted correlation function. The correlation function of a galaxy sample selected by cold gas mass is remarkably similar on small scales in real and redshift space. For pair separations in excess of a few Mpc, the redshift space correlation function has a higher amplitude than in real space, as expected given the effective bias of the sample (Kaiser 1987). In contrast, for an optically selected sample with the same number density of galaxies, the correlation is steeper in real space for $r < 1h^{-1}$Mpc and is damped in redshift space on these scales, due to the greater influence of satellite galaxies in massive haloes. On larger scales there is a more modest boost in the clustering amplitude in redshift space, due to the larger effective bias of the optical sample. The clustering predictions for HI selected galaxies are in reasonable agreement with the measurements by Meyer et al. (2007). The clustering in the modified version of the Bower et al. model (MHIBow06) best agrees with the HIPASS results.

One of the primary science goals of the Square Kilometre Array (SKA) is to make a high precision measurement of large-scale structure in the galaxy distribution. By measuring the apparent size of baryonic acoustic oscillations (BAO) at a particular redshift, the cosmological distance to that redshift can be derived, thereby constraining the equation of state of the dark energy. By combining the galaxy formation model with a very large volume N-body simulation ($1h^{-3}$Gpc^3), we have been able to demonstrate that galaxy samples constructed on the basis of cold gas mass can trace the BAO with the same fidelity as an near-infrared selected sample with the same number density of galaxies.

The key remaining question is how effectively do HI and optical redshift surveys sample the available geometrical volume and how does this translate into an error on the dark energy equation of state parameter? The effective survey volume varies substantially between HI surveys of different duration and for different assumptions about the split between atomic and molecular hydrogen. However, at least for the

case of a cosmological constant, these differences occur in a redshift range which has little impact on the derived error on the equation of state. We find that HI surveys are comparable in accuracy to the most ambitious near-infrared spectroscopic surveys currently under discussion, particularly if the oversampling needed to obtain a field of view which increases with redshift can be achieved, and will give a factor of ≈ 3 times smaller error on w than a slitless Hα redshift survey; all are bone fide Stage IV experiments in the Dark Energy Task Force nomenclature (Albrecht et al. 2006). The assumption about the ratio of molecular to atomic hydrogen is one of the major uncertainties at present, and leads to larger differences in the predicted counts of HI emitters than the choice of galaxy formation model. The fraction of molecular hydrogen is thought to depend upon the local conditions in the interstellar medium. This question requires further modelling (e.g. Krumholz, McKee & Tumlinson 2009), augmented by observations of the HI and CO distribution in nearby galaxies, for example by HI surveys on the SKA pathfinder MeerKAT and CO measurements using the Atacama Large Millimeter/submillimeter Array (Wootten 2008).

Chapter 5

Clustering of extragalactic sources below the Planck detection limit

5.1 Introduction

Planck was launched on 14 May 2009 and aims to provide answers to some of the most important questions in modern cosmology. Planck's primary science objectives are to map CMB anisotropies, to test inflationary models of the early universe, to measure the amplitude of structures in the CMB and to perform measurements of the Sunyaev-Zeldovich effect (Sunyaev & Zel'dovich 1980); for detail see the Planck bluebook (2005). The CMB anisotropy map produced by Planck will be markedly superior to those currently available and will be used to set constraints on the values of the basic parameters that govern the large scale structure and evolution of the Universe. Planck has the ability to detect much smaller temperature variations in the CMB (of the order of one part in 10^6) than WMAP or COBE, and to perform measurements with an angular resolution better than 5 arcminutes. The availability of a wider range of frequencies (from 30GHz to 857GHz) will improve the separation of the primordial CMB signal from interfering foreground signals (Planck bluebook, 2005).

In the Planck frequency range (30GHz $\sim$ 857GHz) there are several important foreground signals due to the Galaxy and extragalactic sources, which are poorly understood. One of the most important steps the analysis of in cosmic microwave

background (CMB) data is the isolating of the primordial signal. This is essential for the robust estimation of cosmological parameters. In order to identify contamination in the CMB maps at high l (>500), we need to consider the intensity fluctuations of undetected galaxies at the Planck frequencies.

The importance of foreground removal has been recognized (Banday & Wolfendale 1991; Readhead & Lawrence 1992; Brandt et al. 1994; Tegmark & Efstathiou 1996; Tegmark 1998; Tegmark et al. 2000), as has the preferred method for discriminating against such contamination, namely multi-frequency observations. The internal linear combination map (ILC) have been used for the cleaned CMB map which is reconstructed by co-adding the data at the five frequencies with a set of weights that minimizes the final variance of the map by WMAP team.

To better understand the nature of dusty extragalactic sources, several surveys in the Far-IR and sub-mm have been carried out using e.g. ISO (Genzel & Cesarsky 2000; Elbaz et al. 2002; Dole et al. 2001), SPITZER (Papovich et al. 2004; Dole et al. 2004), SCUBA (Holland et al. 1998), and MAMBO (Bertoldi et al. 2000). In addition to extending our knowledge of galaxy formation and evolution, these surveys have revealed the nature of the cosmic infrared background.

There are have been a few attempts to build theoretical models to explain the abundance and redshift distribution of dusty galaxies. Predictions using empirical approaches have been made by Haiman & Knox (2000), Lagache et al. (2003) and Fernandez-Conde et al. (2008). More theoretical approaches have been used by Song et al. (2003), Vielva et al. (2003), Negrello et al. (2007) and Righi et al. (2008)

Arguably the most sophisticated predictions are those by Negrello et al. (2007) who combined the galaxy formation model of Granato et al. (2004) with phenomenological models for the spectral energy distributions (SEDs) of starburst and normal late-type galaxies and of radio sources to predict the angular power spectrum of extra-galactic sources. Negrello et al. predicted the angular power spectrum of intensity fluctuations for the HFI instrument of Planck (353GHz, 527GHz, and 857GHz) an approximation for the two-point correlation function of the undetected galaxies. They assumed that the correlation function on large scales is determined

by combining of dark matter correlation function with a redshift dependent linear bias factor (Sheth & Tormen 1999; Percival et al. 2003) with the small scale correlation function forced to be a power law.

Here we use the GALFORM semi-analytical galaxy formation model (Cole et al. 2000) to predict the physical properties of the galaxy population at different redshifts and the GRASIL spectrophotometric model to predict the detailed SEDs of the model galaxies (Silva et al. 1998). We use the model of Baugh et al. (2005) which reproduces the observed number counts and redshift distribution of the faint sub-mm galaxies by invoking a the top-heavy IMF in bursts. We use the galaxies predicted by combining GALFORM with the GRASIL code. The spatial distribution of galaxies is obtained by combining the galaxy formation model with the Millennium N-body simulation (Springel et al. 2005). The galaxy formation model has been tested at a range of frequencies which are sensitive to emission from dust heated by starlight (see the comparison with the Spitzer number counts and redshift distributions in Lacey et al. 2008 and predictions for Herschel in Lacey et al. 2010)

In this chapter we look at the fluctuating foreground produced by undeteced dusty galaxies at the Planck frequencies. We briefly review the model which used to predict angular intensity fluctuation of undetected galaxies in Section 5.2. We then look at the angular fluctuation and angular power spectrum of undetected galaxies in Section 5.3. We present a summary along with our conclusions in Section 5.4.

5.2 The model

To predict the contamination of the primordial CMB signal due to faint galaxies, we compute the angular correlation function of intensity fluctuations. In this section we introduce the theoretical concepts used in the chapter. We describe the galaxy formation model in Sec. 5.2.1 and outline how this is implemented in an N-body simulation in Sec. 5.2.2. In Sec. 5.2.3, we show how to calculate the clustering of intensity fluctuations.

Table 5.1: The main characteristics of the Planck instruments. The rows are as follows: (1) The name of the instrument. (2) The central frequency. (3) The central wavelength. (4) The flux limit for point sources (Veleva et al. 2003). (5) The angular resolution.

	LFI						HFI		
Frequency(GHz)	30	44	70	100	143	217	35	545	857
Wavelength(μm)	10000	6810	4290	3000	2100	1380	850	550	350
Flux limit(Jy)	0.23	0.25	0.24	0.16	0.13	0.12	0.18	0.49	0.48
Angular resolution(arcmin)	33	24	14	9.5	7.1	5.0	5.0	5.0	5.0

5.2.1 The hybrid galaxy formation model

We combine the GALFORM semi-analytical model of galaxy formation (Sec 5.2.1, Cole et al. 2000) with the GRASIL spectrophotometric code (Sec 5.2.1, Silva et al. 1998) which computes the stellar and dust emission from galaxies (Granato et al. 2000) along with the radio emission (Bressan et al. 2000).

The GALFORM galaxy formation model

We predict the formation and evolution of galaxy properties within the ΛCDM structure formation framework using semi-analytical model GALFORM. The GALFORM model computes the evolution of galaxies in the framework of the ΛCDM cosmology. The GALFORM model is described in Cole et al. (2000). The main processes treated include : (1) the formation of dark matter haloes by accretion of smaller haloes and mergers; (2) the formation of galactic disks following the shock-heating and radiative cooling of gas inside dark matter haloes; (3) quiescent star formation in galactic disks and bursts of star formation driven by galaxy mergers; (4) feedback both from supernovae and photoionization of the inter galactic medium; (5) chemical evolution of the stars and gas; (6) galaxy mergers. GALFORM predicts the number and properties of galaxies in dark matter haloes of different masses. The key predicted galaxy properties for this study are the stellar and cold gas masses, the star formation and merger histories, the scale size of the disk and bulge components, and the metallicity.

The model we use in this chapter is that of Baugh et al. (2005) (see also Lacey et al. 2010). The background cosmology is a spatially flat ΛCDM model with parameters Ω_m=0.3, Ω_Λ=0.7, Ω_b=0.04, $h = H_0/100\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$=0.7, and an amplitude of the initial spectrum of density fluctuations given by σ_8=0.93. The merger histories of the dark matter haloes are computed using a Monte Carlo merger technique based on the extended Press-Schechter theory (Parkinson, Helly & Cole 2008).

The model includes two modes of star formation, a "quiescent" mode in galactic disks, and a "burst" mode triggered by major and minor galaxy mergers[1] with gas rich primaries. The two modes of star formation are assumed to have different stellar Initial Mass Functions (IMFs). Quiescent star formation is assumed to have a solar neighbourhood IMF (Kennicutt 1983). Bursts of star formation are assumed to form stars with a top heavy IMF. Baugh et al. (2005) found that the top heavy IMF in bursts is necessary to reproduce the observed number counts and redshift distributions of the faint sub-mm galaxies, whilst at the same time reproducing the local galaxy population. The frequencies covered by the Planck instruments are listed in Table 5.1. To make predictions at these wavelengths we need to use the GRASIL code, which can follow the emission by dust heated by starlight and also emission in the radio.

The GRASIL spectrophotometric code

The GRASIL computes the emission from the stellar population, the absorption and emission of radiation by dust, and also radio emission by thermal and synchrotron processes powered by massive stars (Bressan et al. 2002). The main features of the GRASIL model are described in Lacey et al. 2010). The output from GRASIL is the complete SED of a galaxy from the far-UV to the radio (wavelengths 0.01 μm $\leq \lambda \leq$ 1m) based on theoretical models of stellar evolution and stellar atmospheres, radiative transfer through a two-phase dust medium to calculate both the dust extinction and dust-heating, and a distribution of dust temperatures in each galaxy

[1]The type of merger depends on the ratio of the mass of the merging satellite, $\mathrm{M_{sat}}$, to that of the central galaxy, $\mathrm{M_{cen}}$. If $\mathrm{M_{sat}/M_{cen}} \geq 0.3$ than the merger is defined as "major merger". "Minor merger" are those for which $\mathrm{M_{sat}/M_{cen}} < 0.3$

calculated from a detailed grain model (Silva et al. 1998; Bressan et al. 2000).

5.2.2 Populating on N-body simulation with galaxies

Even though analytic calculations gives reasonably accurated clustering predictions on large scale, this approach is not adequate on small scales, corresponding to galaxies in the same halo. To produce a clustering prediction that is accurate over a wide range of scales, we combine our galaxy formation model with the Millennium N-body simulation (Springel et al. 2005). To implant galaxies into the N-body simulation, there are several steps we have to go through which are set out below.

Our staring point is the hybrid GALFORM plus GRASIL model set in the concordance ΛCDM cosmology, which we refer to as MCGAL. The end point is a hybrid model implanted in the Millennium Simulation, which has a different cosmology. We denote this model as MILLGAL.

The MCGAL catalogue is set up by sampling galaxies according to their stellar mass. For each galaxy we have the stellar mass, host halo mass, a weight based on abundances and the SED.To populate the simulation with galaxies,

1. Estimate the HOD (Halo Occupation Distribution) of galaxies from MCGAL catalogue. We first construct the halo occupation distribution of galaxies, i.e. the mean number of galaxies in dark matter haloes. (Benson et al. 2000; Peacock & Smith 2000; Berlind & Weinberg 2002). The HOD quantifies the mean number of galaxies per halo as a function of halo mass. Fig. 5.1 shows the HOD of galaxies in the MCGAL catalogue at three different redshifts. We plot the HOD for central (blue line) and satellite (red line) galaxies separately. The HODs of central galaxies in the MCGAL catalogue at z=0.1 and z=1 is approximately a step functions. This is not a case at z=5 where the HOD of central galaxies is not a step function due to the absence of bursting central galaxies in massive dark matter haloes. The shape of the satellite galaxy HOD in the MCGAL catalogue is close to a power law in halo mass. The lowest mass dark matter halo in MCGAL catalogue is varies with redshift as the halo mass grid is defined at each redshift to sample haloes with a representative range of abundances. At z=0.1, the lowest halo mass considered is nearly the same as the lowest mass halo resolved in the Millennium simulation

($\sim 10^{10.5} M_{\odot}/h$, shown by the vertical dotted line in Fig. 5.1). However, the lowest dark matter halo mass in the MCGAL catalogue decreases as redshift increases. In contrast to the MCGAL catalogue, the lowest dark matter halo mass of Millennium simulation is the same at all redshifts. Therefore, we can not transplant all of the galaxies contained in the MCGAL catalogue into the Millennium simulation.

2. Matching halo mass function. The MCGAL catalogue, for historical reasons, assumes a different cosmology to that adopted in the Millennium simulation, which is based on the WMAP1 cosmology [2]. We therefore have to match the dark matter halo mass function in the MCGAL cosmology and in the N-body simulation in order to obtain the same luminosity density in MCGAL catalogue and the N-body catalogues.We apply a mass scaling to force the MCGAL and N-body halo mass function to match, which means that we relabel the masses of the MCGAL haloes. Fig. 5.2 shows the halo mass functions for the MCGAL, the N-body simulation, and the rescaled N-body mass function at z=0.1. To match the halo mass function of MCGAL catalogue at z=0.1 (red dashed line) with the N-body results(black dashed line) at same redshift, we need a mass scaling $\times 1.2$. The Black solid line shows the halo mass function obtained after this scaling. The halo mass function of the MCGAL catalogue and scaled halo mass function of Millennium simulation are nearly the same. The discrepancy between the MCGAL mass function and the rescaled N-body mass function is less than 5% over four decades in halo mass. Given this small difference, there is no need to rerun the GALFORM calculation(e.g. Neistein et al 2010). We apply the same scheme to different redshifts and find that it works equally well but with different scaling factors.

3. Placing galaxies in the N-body simulation. We denote galaxy properties in MCGAL catalogue to MILLGAL catalogue for each galaxies which is generated by

[2]The cosmological parameters used in the Millennium simulation are a matter density $\Omega_0 = 0.25$, a cosmological constant $\Lambda_0 = 0.75$, a Hubble constant $H_0 = 73\,\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, a primordial scalar spectral index $n_\mathrm{s} = 1$, baryon density $\Omega_\mathrm{b} = 0.045$ and fluctuation amplitude $\sigma_8 = 0.9$. The cosmological parameters used in the MCGAL catalogue are $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$, $H_0 = 73\,\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, $n_\mathrm{s} = 1$, $\Omega_\mathrm{b} = 0.04$, and $\sigma_8 = 0.93$.

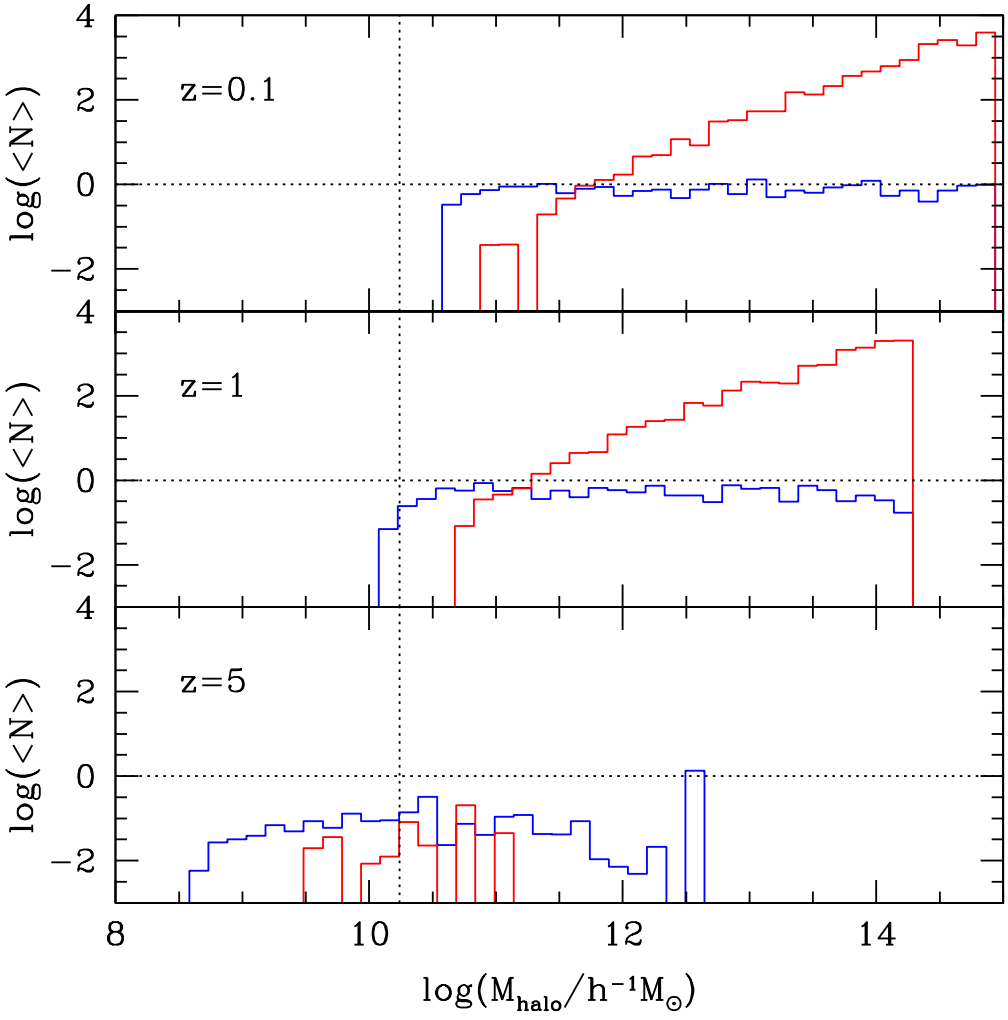

Figure 5.1: The halo occupation distribution (HOD) of galaxies in MCGAL catalogue at different redshifts as indicated by the key. The blue line shows the HOD of central galaxies and red line shows for satellite galaxies. The vertical line indicates the halo mass resolution of the Millennium and the horizontal line shows $< N >$=1.

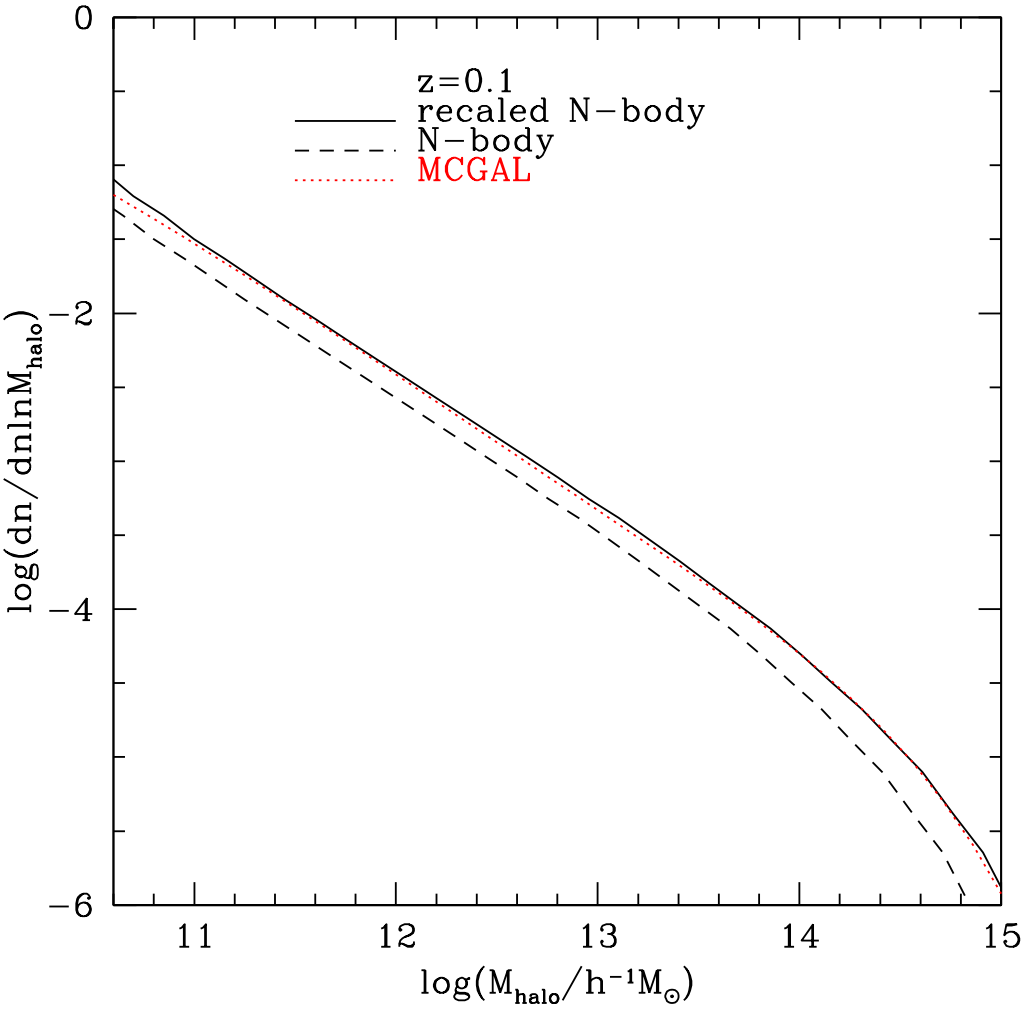

Figure 5.2: The halo mass function at redshift 0.1. The red dotted line shows the halo mass function used in MCGAL catalogue. The black dashed line is the halo mass function z∼0.1 in the Millennium simulation. The black solid line shows the scaled halo mass function for matching from Millennium to MCGAL.

using HOD. The number of galaxies as a function of halo mass is assumed to have a Poisson distribution for $N > 1$. For halo masses for which the HOD predicts $N < 1$, a fraction of haloes is populated with a satellite galaxy at random: i.e. if the random number chosen from a uniform distribution between zero and one, $x < N$, then the halo is assigned a satellite, otherwise it has no satellite. The comparison of two catalogues is given in Fig. 5.3. The luminosity functions of the two catalogues are in very good agreement, both for central galaxies and satellites at low redshift (e.g. z=0.1 in the left panel of Fig. 5.3). The luminosity function of MILLGAL catalogue and MCGAL catalogues differ at high redshift because the former does not include galaxies which are predicted to be in lower mass dark matter haloes than those resolved by the Millennium simulation. However, the MILLGAL catalogue reproduces well the luminosity function of MCGAL catalogue for higher luminosity galaxies which are hosted by the dark matter haloes resolved in Millennium simulation (see the right panel in Fig. 5.3).

5.2.3 Measurement of angular correlation function

Luminosity and Flux Correlations Functions

We can define a spatial luminosity density correlation function, $\xi_L(\vec{x})$, as

$$< \rho_L(\vec{x_1})\rho_L(\vec{x_2}) >=< \rho_L >^2 \ (1 + \xi_L(\vec{x_1} - \vec{x_2}))\,, \tag{5.1}$$

where $\rho_L(\vec{x_1})$ is the luminosity density at position $\vec{x_1}$ and $< \rho_L >$ is the mean luminosity density. Similarly, we can define an angular flux or surface brightness correlation function, $w_{\rm f}(\vec{\theta})$, as

$$< f(\vec{\theta_1})f(\vec{\theta_2}) >=< f >^2 \ \left(1 + w_{\rm f}(\vec{\theta_1} - \vec{\theta_2})\right), \tag{5.2}$$

where the flux/surface brightness is related to the luminosity density via

$$f(\vec{\theta})\,{\rm d}^2\theta = \frac{1}{4\pi} \int \frac{x^2 \rho_L}{{\rm d}_L^2(x)}\,{\rm d}x\,{\rm d}^2\theta, \tag{5.3}$$

where $d_L(x)$ is the luminosity distance to comoving distance x. The above equation is correct if we are considering bolometric luminosity densities and fluxes. However in practice we are nearly always interested in fluxes that are measured over a limited

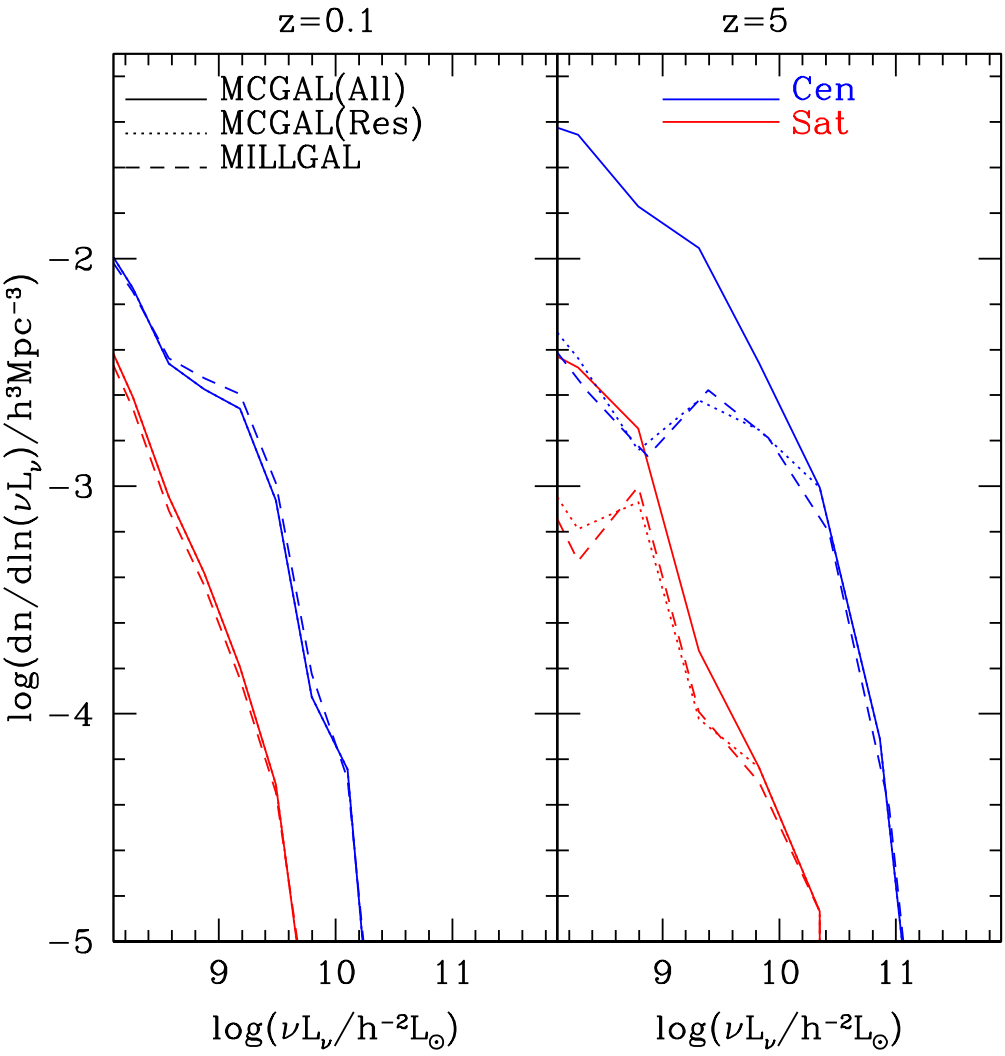

Figure 5.3: The luminosity functions in the 857GHz (350μm) waveband at z=0.1 (left) and z=5 (right). Solid line shows the full MCGAL catalogue and dotted line shows from this catalogue restricted to haloes above the mass of dark matter haloes in Millennium simulation. Dashed line comes from the MILLGAL catalogue. The Blue line is for the luminosity function contributed by central galaxies. The Red line is for satellite galaxies.

frequency band. This introduces an extra $(1+z)$ factor to account for the change in the band width with redshift:

$$f_\nu(\vec{\theta})\,\mathrm{d}^2\theta = \frac{1}{4\pi}\int \frac{(1+z)x^2\rho_L}{d_L^2(x)}\,\mathrm{d}x\,\mathrm{d}^2\theta, \tag{5.4}$$

where ρ_L is now the luminosity density in a band centred on rest frame frequency $\nu(1+z)$.

We use Limber's approximation which depends on two related approximations to obtain the angular flux correlation function. Firstly it assumes that the mean number density of galaxies, $< n(x) >$, varies sufficiently slowly with redshift (here labelled by the comoving radial coordinate x) that over the range of pair separations for which $\xi(\vec{x_1}-\vec{x_2}) \neq 0$, $< n(x_1) > \approx < n(x_2) >$. Secondly, we assume the small angle approximation, i.e. the angular separation of pairs of galaxies for which $\xi(\vec{x_1}-\vec{x_2}) \neq 0$ is small.

Using the above approximations, we can relate the spatial correlation function $\xi(r)$ to the angular correlation of the flux $w_f(\theta)$, through Limber's eqn :

$$w_\mathrm{f}(\theta) = \left(\frac{1}{4\pi}\right)^2 \frac{1}{< f_\nu >^2} \tag{5.5}$$
$$\int_{-\infty}^{\infty}\int_0^{\infty} \frac{(1+z)^2\,x^4 < \rho_L(x) >^2}{d_L^4(x)}\,\xi_L\left((u^2+x^2\theta^2)^{1/2}\right)\mathrm{d}x\,\mathrm{d}u,$$

In a flat universe $(\Omega+\Lambda=1)$ the luminosity distance $d_L=(1+z)x$ and so the above reduces to

$$w_\mathrm{f}(\theta) = \left(\frac{1}{4\pi}\right)^2 \frac{1}{< f_\nu >^2} \tag{5.6}$$
$$\int_{-\infty}^{\infty}\int_0^{\infty} \frac{< \rho_L(x) >^2}{(1+z)^2}\,\xi_L\left((u^2+x^2\theta^2)^{1/2}\right)\mathrm{d}x\,\mathrm{d}u,$$

where

$$< f_\nu >= \frac{1}{4\pi}\int_0^{\infty} \frac{< \rho_L(x) >}{(1+z)}\,\mathrm{d}x. \tag{5.7}$$

Assuming a linear clustering bias factor that depends only on halo mass, i.e galaxy i with luminosity L_i located in a halo of mass M_i^H has bias factor $b_i = b(M_i^\mathrm{H})$ and that we have a complete catalogue of galaxies from a simulation of volume L_box^3 then

$$< \rho_L >= \sum_i L_i / L_\mathrm{box}^3 \tag{5.8}$$

and

$$\xi_L(r) = \bar{b}_L^2\, \xi_{\rm M}(r), \qquad \text{where} \qquad \bar{b}_L = \sum_i L_i b_i / \sum_i L_i. \tag{5.9}$$

These quantities can be evaluated at a discrete set of redshifts to give $< \rho_L(x) >$ and $\xi_L(r, x)$ (where redshift is labelled here by radial comoving distance x) which can be input into Eq. 5.5 to compute the angular clustering of flux. The quantities we need to calculate are $< f_\nu >$ at the frequencies corresponding to the Planck bands and the fluctuations in this background which are given by $< f_\nu >^2\ w_{\rm f}(\theta)$. These quantities are predicted by the galaxy formation model described in the previous section.

The angular power spectrum of the intensity fluctuations can be obtained from the angular correlation function

$$C_l(\theta) = \int_0^{2\pi} \int_0^{\pi} < f_\nu >^2\, w_f(\theta) P_l(\cos\theta) \sin\theta {\rm d}\theta {\rm d}\phi. \tag{5.10}$$

5.3 Results

5.3.1 Basic predictions from the MCGAL catalogue

We study the clustering of undetected galaxies in the different frequency channels of the Planck satellite. First, we use the galaxy catalogue constructed using hybrid GRASIL+GALFORM model with MC trees, MCGAL. Fig. 5.4 shows the luminosity densities predicted using all of the galaxies in the MCGAL catalogue (solid lines) and the galaxies in the MCGAL catalogue which are hosted by dark matter haloes which can resolved by the Millennium simulation (dotted lines) at different redshift. In all nine Planck wavebands, the luminosity density increases up to redshift around 5 and decreases between z$\sim$5 and z=10. The amplitude of the luminosity density increases as the value of L_ν increases at each redshift due to the shape of the SED. The luminosity density of the galaxies in MCGAL catalogue which are located in dark matter haloes resolvable by the Millennium simulation is lower than that of all the galaxies in the MCGAL catalogue for z $>$ 2. The galaxies in low mass dark matter haloes contribute significantly to the luminosity density at high z. As a

result, the predictions with and without the Millennium halo mass limit become progressively more different with increasing redshift.

From the properties of galaxies in the MCGAL catalogue, we calculate the luminosity weighted effective bias of galaxies using the halo bias prescription of Sheth et al. (2001) (see Eq. 5.9). The effective bias presented in Fig. 5.5. are computed using all the galaxies in MCGAL catalogue (solid lines) and only those galaxies which could be resolved in the Millennium simulation (dotted lines). The effective biases predicted at $z < 0.5$ are slightly below unity. These galaxies are slightly less clustered than the dark matter. In contrast to the luminosity density, the overall trend is that the effective bias monotonically increases with redshift. The predicted effective bias at redshift 4 is ~ 4 for the all of galaxies in MCGAL catalogue, i.e. the clustering of undetected galaxies in the Planck wavebands is 15$\sim$20 times higher than that of dark matter at z=4. The effective biases for the galaxies in MCGAL catalogue which are located in dark matter haloes which could be resolved in the Millennium simulation is higher than that predicted all galaxies beyond $z \sim 2$.

5.3.2 Predictions for the clustering of faint extragalactic sources.

The predictions for the angular correlation function of intensity fluctuations using the MCGAL catalogue combined with an analytic calculation of the effective bias and a simple prescription for the dark matter correlation function is expected to be accurate on large scales. We use the galaxy catalogue, MILLGAL, generated using the method described in Sec. 5.2.2 to predict the small scale clustering more accurately. The main difference in the two predictions is the auto correlation function of galaxies. Fig. 5.6 shows the auto correlation functions at different redshifts from the analytic calculation (dashed lines and dotted lines) using the ansatz of Smith et al. (2003) compared with the direct pair count estimates from the MILLGAL catalogue (solid lines). In Fig. 5.6, the difference in clustering strength between the two correlation functions calculated using the Smith et al. (2003) prediction at z=5 comes from the difference in the effective bias as shown in Fig. 5.5. The shape of the correlation function on small scales, ($< \sim$5Mpc/h) is different between the

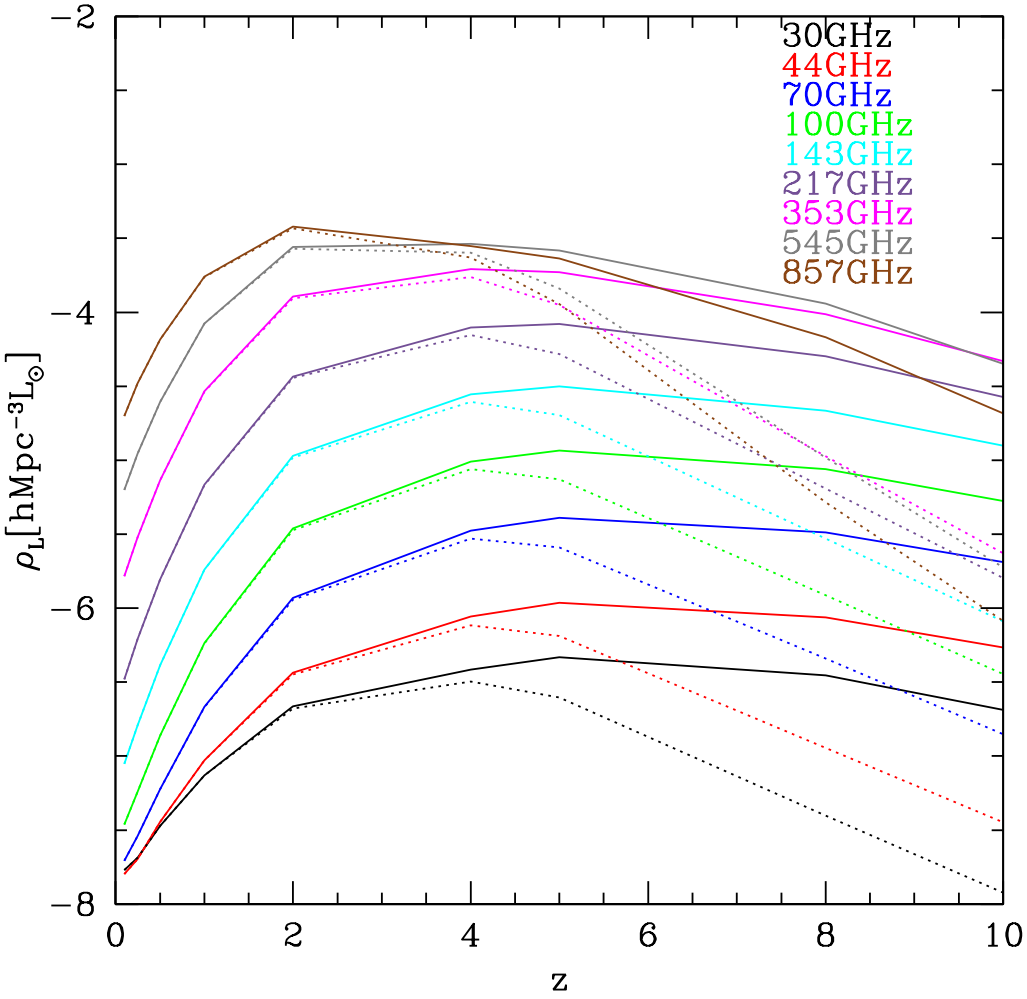

Figure 5.4: The luminosity density in the Planck wavebands as a function of redshift predicted by MCGAL catalogue. The solid lines show the predictions using all galaxies in the MCGAL catalogue and the dotted lies show the results using only those galaxies hosted by haloes which could be resolved in the Millennium.

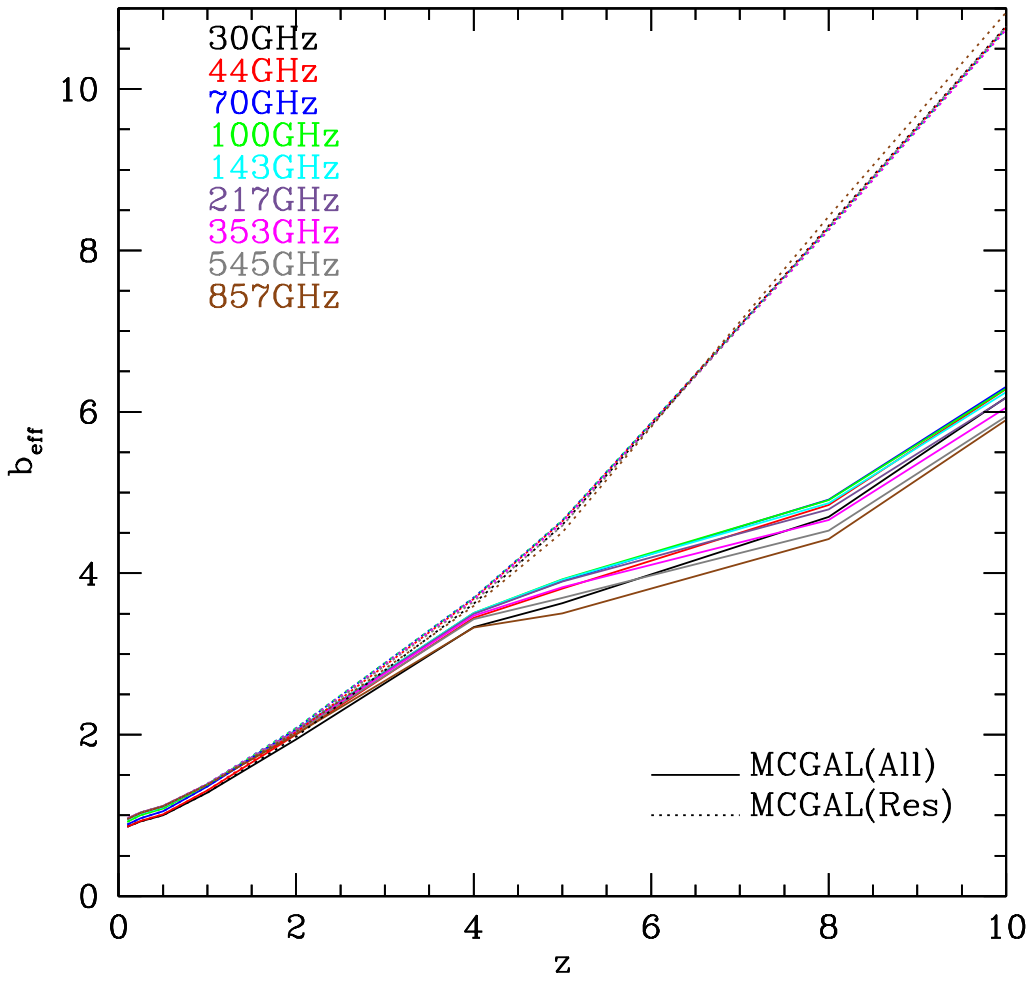

Figure 5.5: The luminosity-weighted effective bias for each wavebands calculated using the analytic approximation of Sheth et al. (2001) as a function of redshift. The effective bias increases as redshift increases.

N-body and analytic methods, especially at high redshift. We note that the correlation functions measured in the Millennium simulation are reliable up to $50h^{-1}$Mpc. We extend the predictions to larger scales using an matched version of the $P(k)$ predicted using the Smith et al. formula. The luminosity-weighted auto correlation of MILLGAL catalogue galaxies has different amplitude at each Planck waveband, however the shape of auto correlation function same as in Fig. 5.6.

Fig. 5.7 shows the cumulation contribution to the mean flux density from different redshifts at 30GHz. The mean flux density integrated up to z=5 contributes over 90% of the value integrated to z=10. The contribution of the mean flux density from z=8 to 10 is less than 5%. The most important redshifts are around z$\sim$2 where the finite resolution of the Millennium has little impact.

Fig. 5.8 shows the angular flux correlation function for 30GHz showing the contribution from different redshifts. The angular flux correlation function is largely determined by galaxies with z<4.

As we mentioned in Sec. 5.2.2, the MILLGAL catalogue cannot contain all of the galaxies in the MCGAL catalogue because of the finite mass resolution of the Millennium. We have checked the impact of the limited halo mass resolution on the angular flux correlation function and angular correlation function of intensity fluctuation. Fig. 5.9 shows the predicted angular flux correlation functions using all galaxies in the MCGAL catalogue (solid lines) and galaxies which are located in dark matter halo which could be resolved in the Millennium (dotted lines). The difference in the predicted angular flux correlation functions in these two cases can be traced to the difference in the effective bias as shown in Fig. 5.5.

The angular correlation function of intensity fluctuations for undetected galaxies are shown in Fig. 5.10. Solid lines are for all galaxies in the MCGAL catalogue and dashed lines are for galaxies in haloes which could be resolved in the Millennium simulation. The angular correlation function of intensity fluctuations are nearly identical for the two different dark matter halo resolutions. Even though the angular flux correlation function is not same (see Fig. 5.9), the angular correlation functions of intensity fluctuations give the same results, because the contribution from the high luminosity galaxies to the angular correlation function of intensity fluctuation

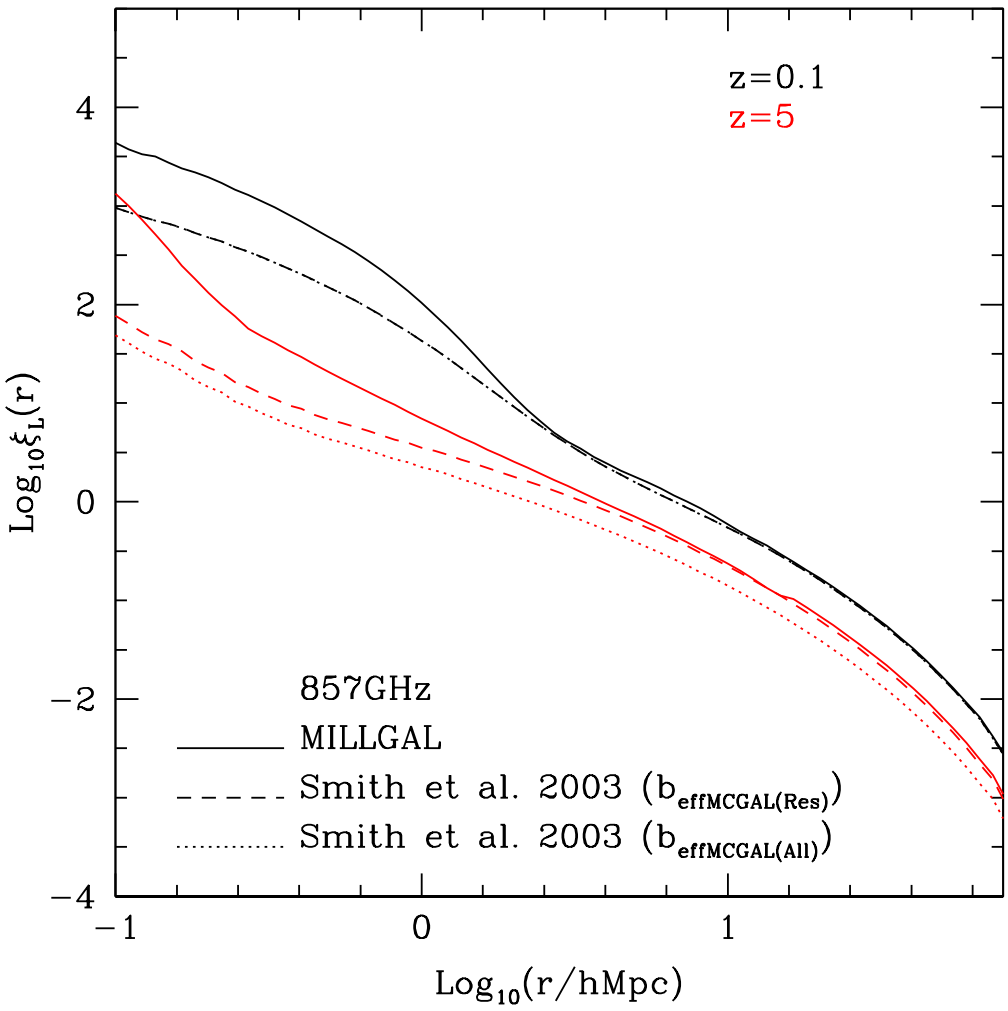

Figure 5.6: The two point flux correlation function at different redshifts. Dashed lines and dotted lines are calculated analytically using Smith et al. (2003) for the non-linear dark matter correlation function, applying the effective bias ($b_{eff_{MCGAL(All)}}$) using all of the galaxies in the MCGAL catalogue and the effective bias ($b_{eff_{MCGAL(Res)}}$) using galaxies located in dark matter haloes resolvable in the Millennium simulation in the MCGAL catalogue, respectively. Solid lines are measured using MILLGAL catalogue. The different colours show different redshifts as indicated by the key.

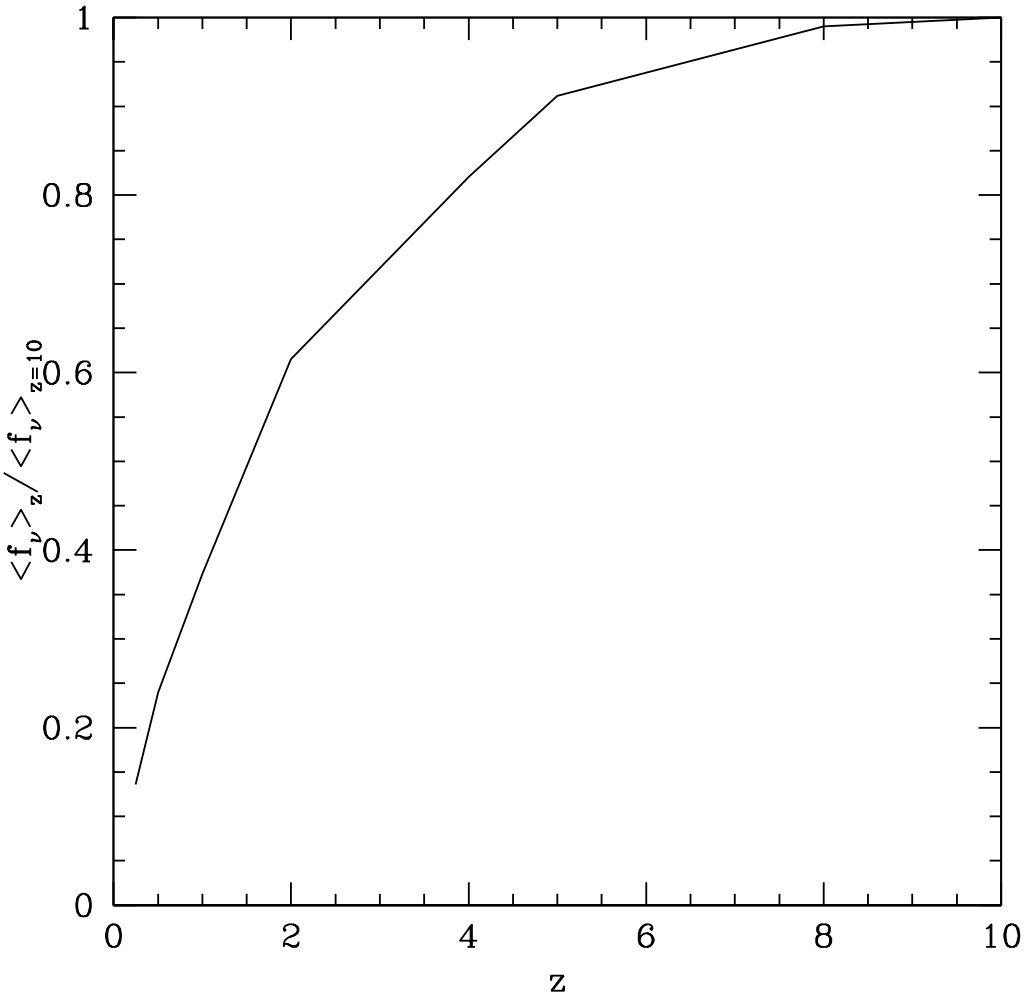

Figure 5.7: The cumulative fractional mean flux density contributed by different redshifts in the MCGAL catalogue (see Eq. 5.2.3) for 30GHz band. The mean flux density has been normalized by the mean flux density integrated up to z=10.

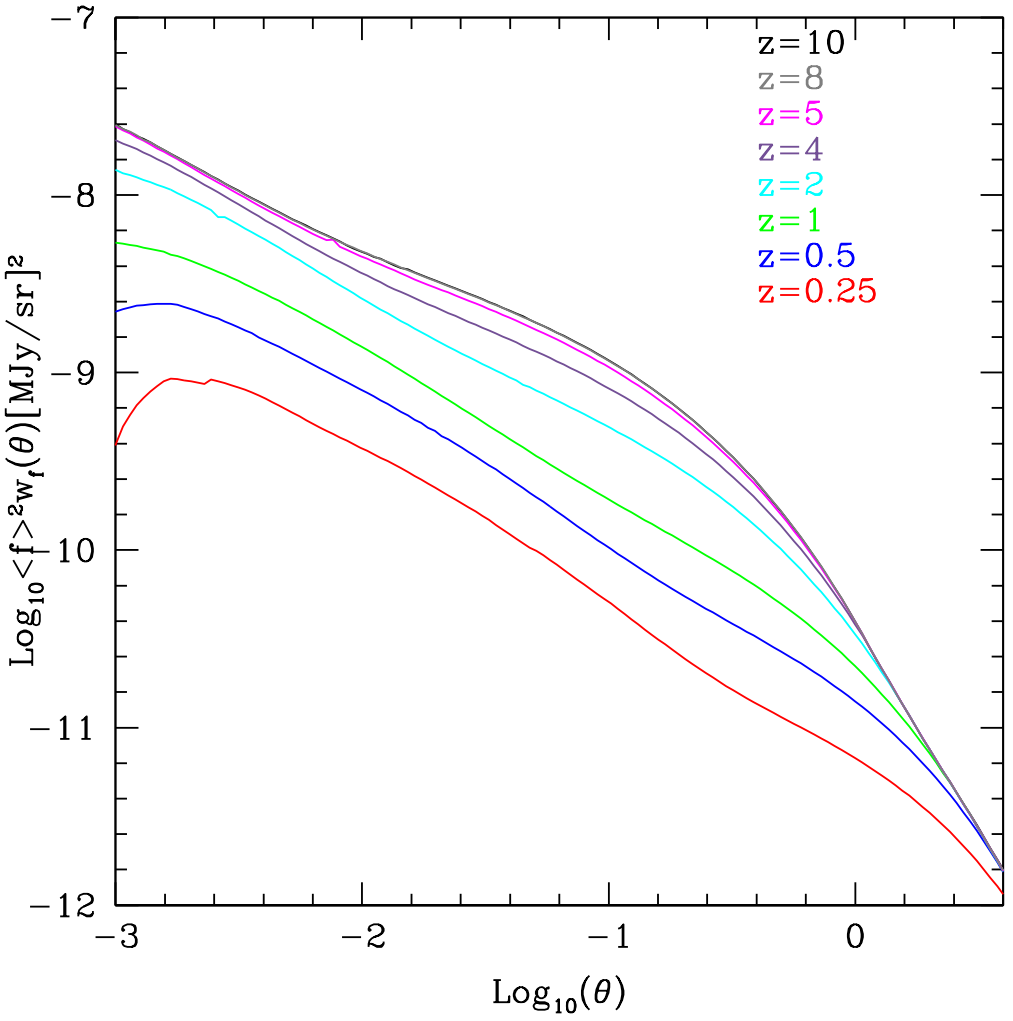

Figure 5.8: The predicted angular flux correlation function of undetected galaxies at 30GHz band adopting different upper limits for the redshift integration, as indicated by the key.

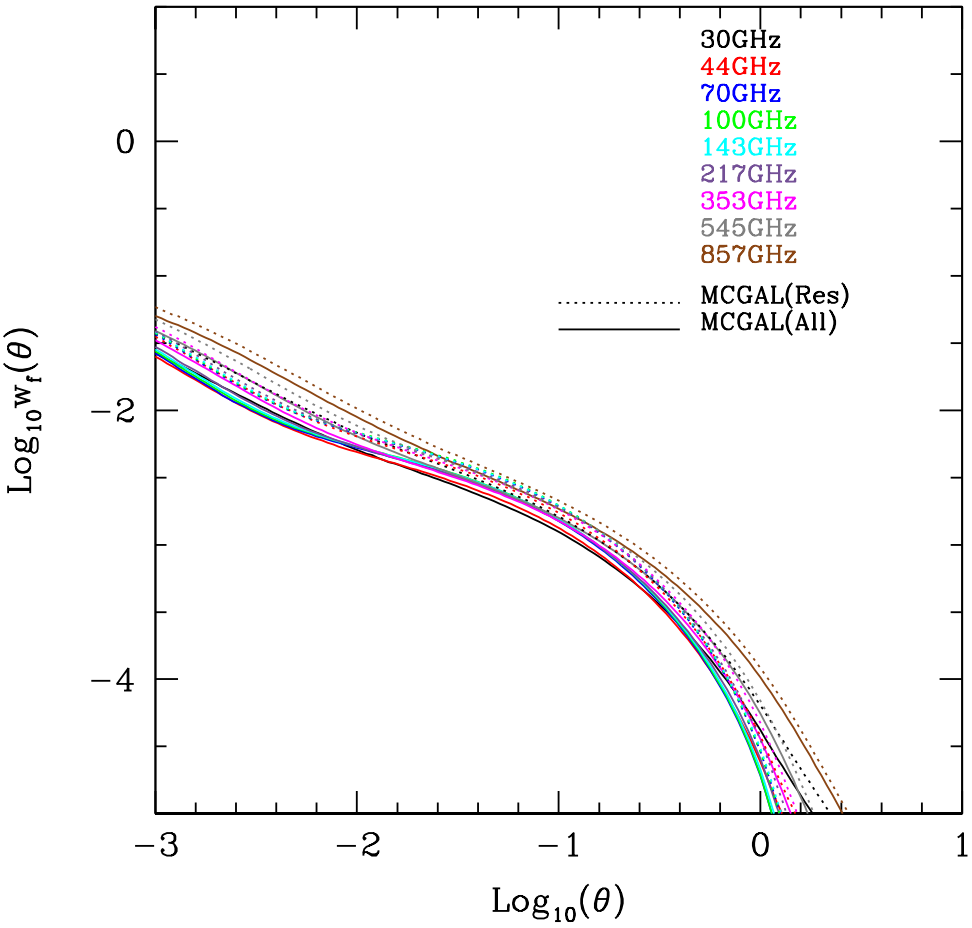

Figure 5.9: The predicted angular flux correlation function of undetected galaxies for nine Planck wavebands as indicated by the key using analytic calculations for the bias factor and correlation function. Solid lines are for all galaxies in the MCGAL catalogue and dashed lines are for galaxies in haloes which could be resolved in the Millennium simulation.

is much greater than that of the low luminosity galaxies.

Fig. 5.11 shows the predicted angular flux correlation function of undetected galaxies for the nine Planck satellite frequencies as indicated by the key. Dashed lines are for all galaxies in the MCGAL catalogue in haloes which could be resolved in the Millennium simulation. Solid lines are for galaxies in the MILLGAL catalogue. The shape of correlation function in small angular scales and the amplitude difference between the predicted angular flux correlation functions using the two different catalogues are caused by the differences of auto correlation function between using analytic calculation and direct calculation using MILLGAL catalogue, see Fig. 5.6.

Fig. 5.12 shows the angular correlation function of intensity fluctuations of undetected galaxies for Planck. Dotted lines use analytic calculation and solid lines are using direct calculation in the MILLGAL catalogue. The amplitude and shape at large angular separations is nearly the same. The difference at small angular separations is due to the more accurate calculation of the one halo term in the N-body calculation.

Fig. 5.13 shows the angular power spectrum of the intensity fluctuations of undetected galaxies in the Planck wavebands. Dotted lines use analytic calculations and the solid lines use the direct calculation in the MILLGAL catalogue. The deviation of the angular power spectrum of intensity fluctuation calculated in the two methods starts from $l{\sim}1000$ which is caused by the difference of auto correlation function on small scales. The contaminations of undetected faint extragalactic sources for the three frequencies in the LFI instrument and the two lowest frequencies in HFI instrument are negligible across the whole range of scales above the angular resolutions of Planck. The angular power spectrum due to the undetected faint extragalactic sources is larger than the CMB power spectrum from $l \sim 1000$ for 353GHz, $l \sim 100$ for 545GHz and for all at 857GHz.

5.4 Summary and conclusions

In the Planck frequency range, there are several important emission foregrounds to the cosmological signal due to Galactic emission and extra-galactic sources.

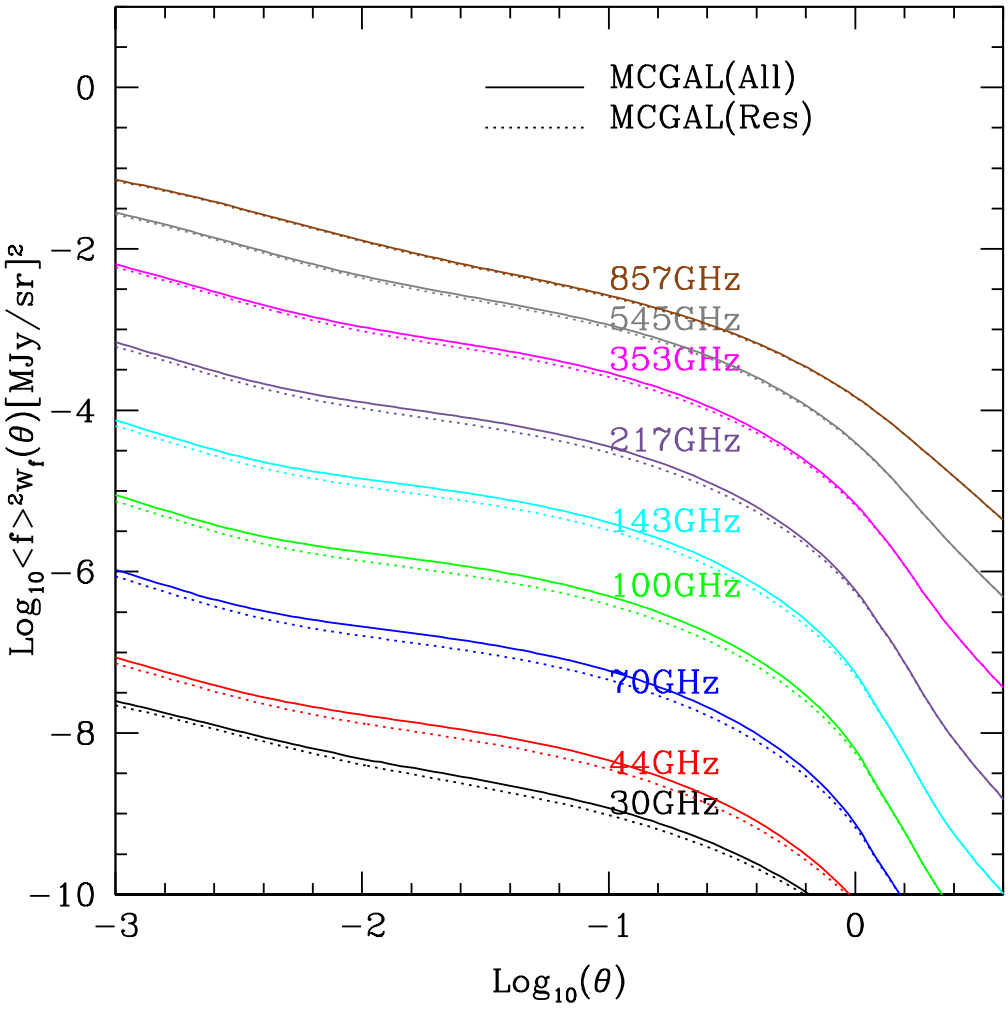

Figure 5.10: The angular correlation function of intensity fluctuations of undetected galaxies for the nine Planck satellite wavebands using analytical calculations for the bias factor and correlation function. Solid lines are for all galaxies in MCGAL catalogue and dashed lines are for galaxies in haloes which could be resolved in the Millennium simulation.

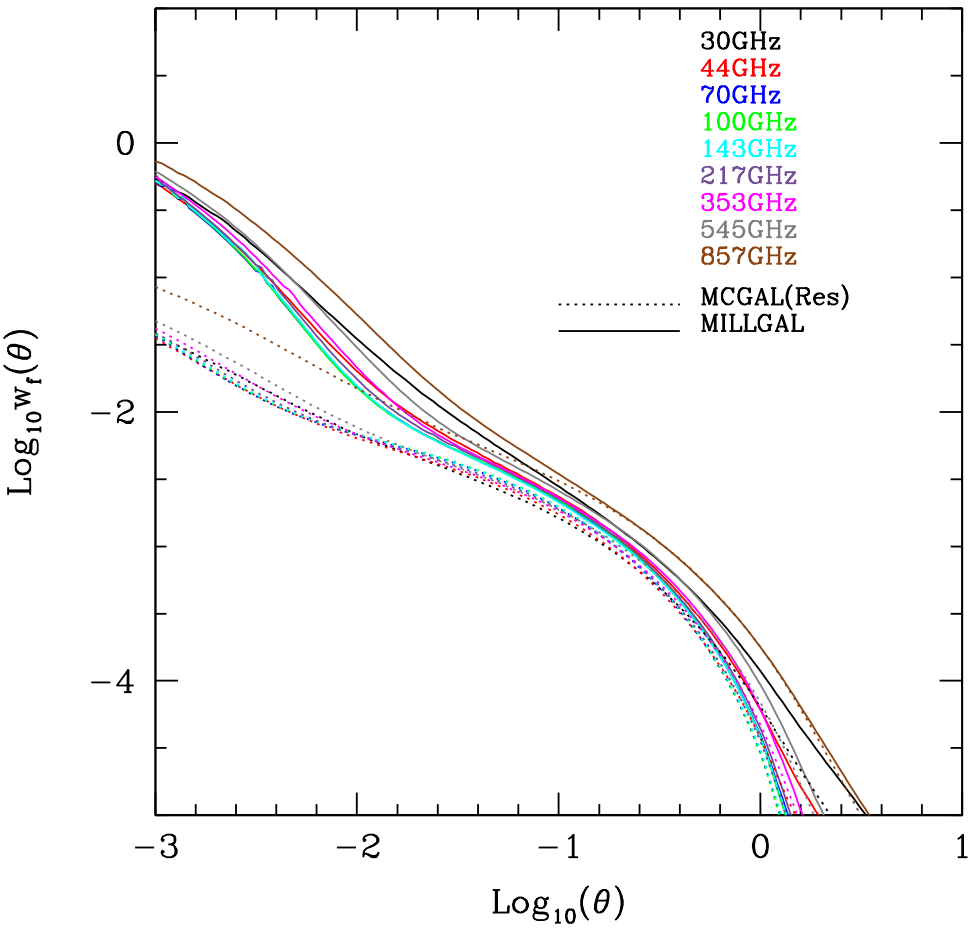

Figure 5.11: The predicted angular flux correlation function of undetected galaxies in the nine Planck satellite wavebands as indicated by the key. Dashed lines are for all galaxies in MCGAL catalogue in haloes which could be resolved in the Millennium simulation. Solid lines are for galaxies in MILLGAL catalogue.

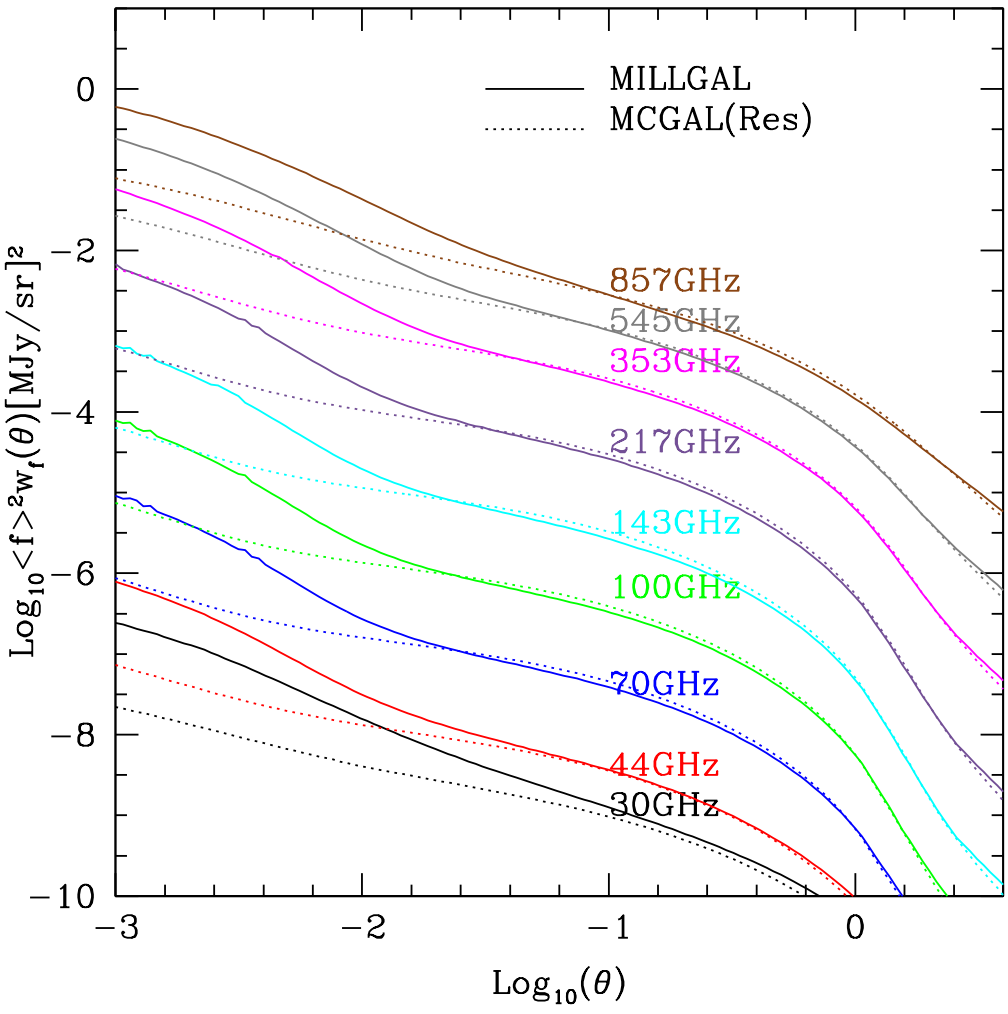

Figure 5.12: The angular correlation function of intensity fluctuations of undetected galaxies in the nine Planck satellite wavebands. Solid lines are for all galaxies in MCGAL catalogue truncated at the haloes resolution of the Millennium simulation. Solid lines are for galaxies in MILLGAL catalogue.

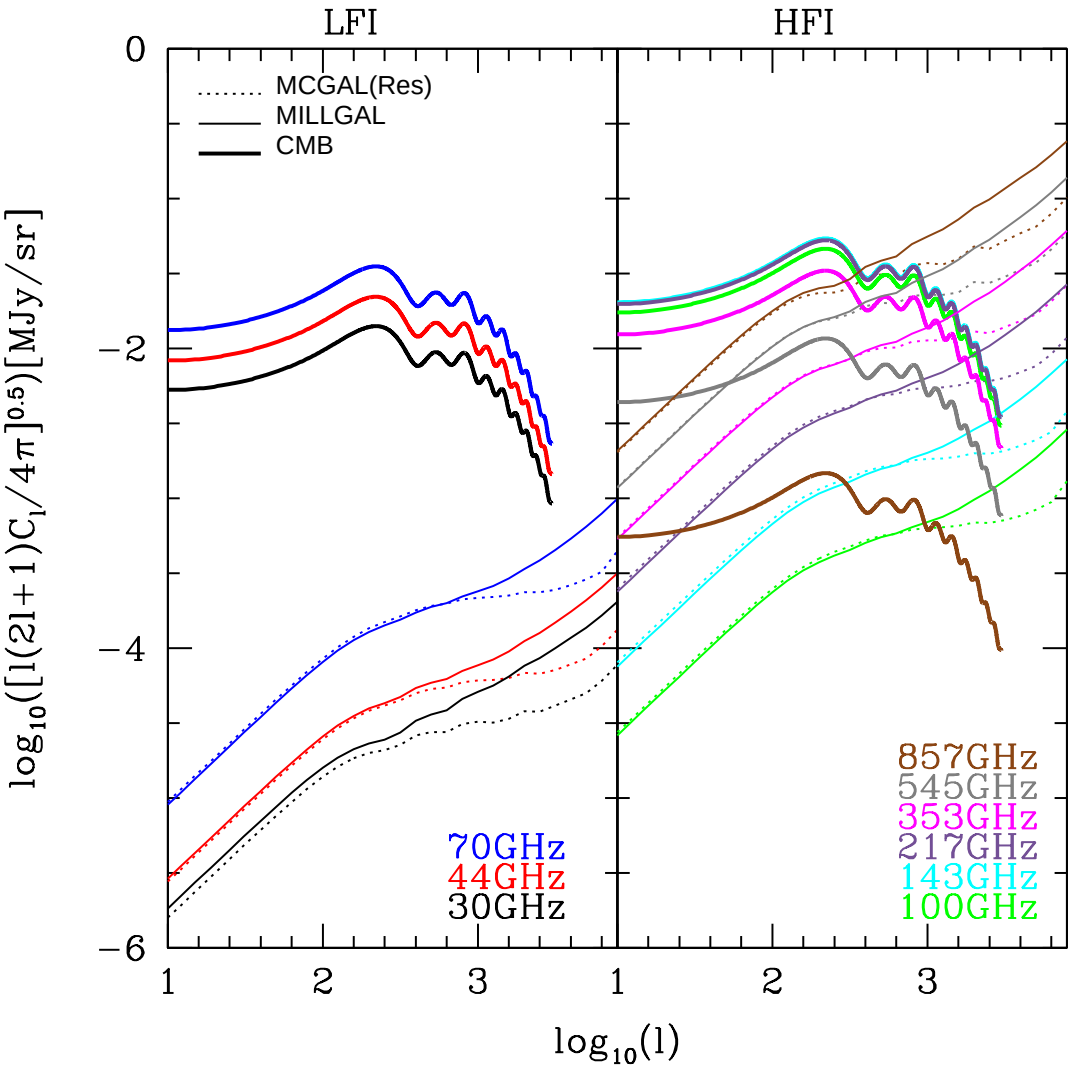

Figure 5.13: The angular power spectrum of the intensity fluctuations of undetected galaxies in the nine Planck wavebands. Solid lines are for all galaxies in the MCGAL catalogue in haloes which could be resolved in the Millennium simulation. Solid lines are for galaxies in the MILLGAL catalogue.The thick lines show the CMB power spectrums. The colours show the different frequency bands.

We have carried out a detailed study of the contamination of the Planck observations due to undetected galaxies. This is a key component of CMB data analysis. Negrello et al. (2007) predicted the angular power spectrum of intensity fluctuation for the HFI instrument of Planck (353GHz, 527GHz, and 857GHz) using the model for the formation and evolution of galaxies of Granato et al. (2004) and simple assumptions for the two-point correlation function of the undetected galaxies. We use the predicted galaxies from combining the GALFORM semi-analytic model with the GRASIL dust reprocessing model. To study the effect of clustering at small separations, (large l), we use the galaxies transplanted into the Millennium dark matter simulation to directly calculate the two-point correlation function instead of adopting an analytic calculation.

We derive the flux intensity fluctuation $< f_\nu >^2 w_f(\theta)$ using Limber's small angle approximation to predict the contribution of undetected galaxies contribution as a CMB foreground. The galaxies predicted by combining GALFORM with GRASIL reproduce the observed number counts and redshift distributions (Baugh et al. 2005; Lacey et al. 2008; Lacey et al. 2010). We use a Monte-Carlo method to populate the dark matter haloes of the Millennium simulation with our model galaxies. Using these catalogues we calculate the flux weighted correlation function of the dust emission at each redshift. We then integrate over redshift to predict the angular flux correlation function and power spectrum. We have performed an analytic calculation using the Smith et al. (2003) dark matter correlation function combined with bias factors from Sheth et al. (2001) for the halos that host our dusty galaxies.

The deviation of the two point correlation functions between using the analytic calculation and using direct calculation are different below a comoving separation, less than $r \sim 5Mpc/h$, corresponds to the 1 halo term, at all redshifts. The predicted luminosity densities of the undetected galaxies in the Planck frequency bands increase up to redshift around 5 and decreases from z$\sim$5 to z=10. The amplitudes of the luminosity density increases as a function of frequency at each redshifts. The predicted value of the effective bias at redshift below 0.5 is less than 1. The overall trend of effective bias for undetected galaxies at Planck frequencies is that the effective bias linearly increases as a function of redshift. The predicted effective bias

at redshift 4 is ~ 4, that is the clustering of undetected galaxies in Planck satellite wavebands is 15$\sim$20 times higher than that of the dark matter at z=4. We found that the identical results for the angular correlation function of intensity fluctuations at two different halo mass resolutions demonstrate that the finite resolution of the N-body simulation is sufficent for our purpose. The amplitude of the angular correlation function of intensity fluctuation in small angular scale and at high multipole number in the angular power spectrum from the MILLGAL catalogue is significantly higher than the analytic calculation. The angular power spectrum caused by the undetected faint extragalactic sources exceeds the CMB power spectrum from $l \sim 1000$ for 353GHz, $l \sim 100$ for 545GHz and all range for 857GHz.

The main results of this chapter is the predicted contribution of undetected galaxies to the CMB signal. We calculate much more accurately at small separations using MILLGAL catalogue than any others using analytic calculation. This prediction can be used as a template of the foreground fluctuation for the Planck team to include in their modelling in order to statistically separate this foreground from the primordial CMB.

Chapter 6

Summary and Future work

In this Thesis we have studied the clustering of galaxies in the hierarchical galaxy formation framework. The physically motivated GALFORM semi-analytic model successfully explains many observational results and give us the chance to make predictions for future surveys.

In Chapter 3, we have studied the luminosity dependence of clustering. The dependence of galaxy clustering on luminosity has been measured with high accuracy in the local Universe by the 2dFGRS and SDSS (Norberg et al. 2001, 2002; Zehavi et al. 2002, 2005, 2010; Jing & Borner 2004; Li et al. 2006). We have shown that the current "best bet" publicly available galaxy formation models only match the observational results in a qualitative sense. These models fail to match the trend of clustering strength with luminosity. We have demonstrated that the reason for the discrepancy is that the models predict too many satellites in massive haloes. Li et al. (2007) reached a similar conclusion comparing the clustering of galaxies in the red selected SDSS with the semi-analytical models of Kang et al. (2005) and Croton et al. (2006).

To fix this problem, we considered applying two processes which are not currently included in most galaxy formation models: mergers between satellite galaxies and the tidal disruption of satellites. The first of these processes is motivated by recent high resolution simulations of the formation of dark matter haloes which show that hierarchies of substructures persist (Diemand et al. 2008; Springel et al. 2008). Mergers which started in a progenitor halo can run to completion in the descendant

halo (Angulo et al. 2009). The disruption of satellites has been modelled analytically in the Durham model in a study of the heating of the Milky Way's disk (Benson et al. 2004). Here, we applied a simple prescription to remove luminosity from satellites based on the ratio of the host halo mass to the mass of the halo in which the satellite formed, which is related to the timescale for the satellite's orbit to decay through dynamical friction. Applying the model for the disruption of satellites changes the overall amplitude of clustering without improving the trend of clustering strength with luminosity. Including mergers between satellites, on the other hand, does alter the predictions for the luminosity dependence of clustering. By applying both extensions together, we are able to obtain a significantly improved match to the 2dFGRS measurements (Norberg et al. 2009). The hybrid model matches the observational constraints on the amount of intracluster light.

In Chapter 4, we compared the predictions for the distribution of cold gas in dark matter haloes in four versions of the Durham semi-analytical galaxy formation model, `GALFORM`. The Bower et al. (2006) and Font et al. (2008) models are publicly available from the Millennium Archive. These models overpredict the local abundance of galaxies as a function of their cold gas mass. This excess was straightforward to fix, with the primary adjustment being to reduce the model star formation timescale. This modified model, based on Bower et al. (2006) was still able to reproduce the optical luminosity function, with the same level of success as Bower et al. We also considered a galaxy formation model set in a different cosmology, to take advantage of a N-body simulation with a large enough box size to accurately model baryonic acoustic oscillations. This model also adopted a modified star formation timescale to better match the local HI mass function.

The relative importance of central and satellite galaxies has an impact on the form of the predicted correlation function. The correlation function of a galaxy sample selected by cold gas mass is remarkably similar on small scales in real and redshift space contrast with that of an optically selected galaxy sample. The clustering predictions for the HI samples are in reasonable agreement with the observational measurements by Meyer et al. (2007). The clustering in the modified version of the Bower et al. model (MHIBow06) agrees best with the HIPASS results.

One of the primary science goals of the Square Kilometre Array (SKA) is to make a high precision measurement of large-scale structure in the galaxy distribution. By measuring the apparent scale of baryonic acoustic oscillations (BAO) at a particular redshift, the cosmological distance to that redshift can be derived, thereby constraining the equation of state of the dark energy. By combining the galaxy formation model with a very large volume N-body simulation ($1h^{-3}\mathrm{Gpc}^3$), we have been able to demonstrate that galaxy samples constructed on the basis of cold gas mass can trace the BAO with the same fidelity as a near-infrared selected sample with the same number density of galaxies.

The key remaining question is how effectively do HI and optical redshift surveys sample the available geometrical volume and how does this translate into an error on the dark energy equation of state parameter? The effective survey volume varies substantially between HI surveys of different duration and for different assumptions about the split between atomic and molecular hydrogen. However, at least for the case of a cosmological constant, these differences occur in a redshift range which has little impact on the derived error on the equation of state. We find that HI surveys are comparable to the most ambitious near-infrared spectroscopic surveys currently under discussion, and will give a factor of two smaller error on w than a slitless H-α redshift survey; all are bone fide Stage IV experiments in the Dark Energy Task Force nomenclature (Albrecht et al. 2006). The uncertainty in the ratio of molecular to atomic hydrogen is one of the major unknowns at present, and leads to larger differences in the predicted counts of HI emitters than the choice of galaxy formation model. The fraction of molecular hydrogen is thought to depend upon the local conditions in the interstellar medium. This question requires further modelling (e.g. Krumholz, McKee & Tumlinson 2009), augmented by observations of the HI and CO distribution in nearby galaxies, for example by HI surveys on the SKA pathfinder MeerKAT and CO measurements using the Atacama Large Millimeter/submillimeter Array (Wootten 2008).

In Chapter 5, We studied the contamination due to undetected galaxies expected in CMB maps from Planck satellite. One of the most important process in the CMB data analysis that the cleaning up of the microwave maps in order to make the CMB

maps as accurate as possible to permit robust constraints on the cosmological parameters. Negrello et al. (2007) predicted the angular power spectrum of intensity fluctuation for the HFI instrument of Planck (353GHz, 527GHz, and 857GHz) using the physical model for the formation and evolution of galaxies of Granato et al. (2004) and an assumption for two-point correlation function of the unresolved galaxies. We use the predictions from combining GALFORM semi-analytic model with GRASIL dust reprocessing model. To study the effect of clustering at small separation scale, large l, we transplanted galaxies into the Millennium dark matter simulation to directly calculate the two-point correlation function.

We derive flux intensity fluctuation $< f_\nu >^2 w_f(\theta)$ using Limber's small angle approximation to predict the unresolved galaxies contribution to foreground contamination of the CMB map. The galaxy dust emission properties are predicted by combining GALFORM semi-analytic model with the dust reprocessing model GRASIL. The galaxies predicted by combining GALFORM with GRASIL model explain many observational results very well (Granato et al. 2000; Baugh et al. 2005; Le Delliou et al. 2005 and 2006; Lacey et al. 2008; Orsi et al. 2008; Lacey et al. 2010). We use a Monte-Carlo method to populate the dark matter haloes of the Millennium simulation with our model galaxies. Using these catalogues we calculate the flux weighted correlation function of the dust emission at each redshift. We then integrate over redshift to predict the angular flux correlation function and power spectrum.

The two point correlation functions predicted from an analytic calculation and our direct calculation using an N-body catalogue differ at comoving separations of less than $r \sim 5$Mpc/h, corresponding to the 1 halo term, at all redshifts. The predicted luminosity density of the unresolved galaxies in the Planck frequency bands increases up to redshift around 5 and decreases from z$\sim$5 to z=10. The amplitudes of the luminosity densities increases as a function of frequency at each redshifts. The predicted value of effective biases at redshift lower than 0.5 are less than 1. The overall trend for undetected galaxies is for the effective bias to monotonically increase as a function of redshift. The predicted effective bias at redshift 4 is $\sim$ 4, that is the clustering of undetected galaxies in the Planck wavebands is 16$\sim$20

times higher than that of the dark matter. The amplitudes of the angular correlation function of intensity fluctuation at small angular scale (large multipole number) in the angular power spectrum from the N-body catalogue is significantly higher than in the analytic calculation.

The main results of Chapter 5 is the prediction that undetected galaxies contribute significantly to the CMB spectrum in some bands. This depends critically on the use of an N-body simulation to predict the small scale clustering, which is more accurate than previous calculations in the literature. This prediction will be used as a template of the foreground fluctuations by the Planck team to include in their modelling in order to statistically separate this foreground from the primordial CMB.

The GALFMORM semi-analytical galaxy formation mocks are a powerful tool with which to interpret future galaxy surveys. The models predict the star formation and galaxy merger history of each galaxy, allowing us to calculate the spectral energy distribution from the galaxy's composite stellar population. Further work includes the contribution of mock catalogues for the ASKAP (Australian Square Kilometre Array Pathfinder) survey WALLABY (Widefield ASKAP L-band Legacy All-Sky Blind surveY) including H_2/HI ratio prescriptions as described in Chapter 4 and CO component prescription for each model galaxy. This will allow us to explore the connection between galaxies sampled by surveys conducted at different wavelengths. We will be able to explore the environmental dependence of gas content and star formation rate, providing important new constraints on galaxy formation models. Currently, satellite-satellite mergers are implemented in a crude way by post processing the GALFORM output. The model discussed in Chapter 3 uses the ratio M_H/M_{sat} to set the probability of a merger and tidal disruption, but does not take into account when the satellite fell into the main halo, or its orbit. Therefore, I will extend the GALFORM code for the satellite-satellite galaxy mergers which discussed in Chapter 3.

Bibliography

[1] Abdalla F. B., Blake C., Rawlings S., 2010, MNRAS, 401, 743 *"Forecasts for dark energy measurements with future HI surveys"*

[2] Abdalla F. B., Rawlings S., 2005, MNRAS, 360, 27 *"Probing dark energy with baryonic oscillations and future radio surveys of neutral hydrogen"*

[3] Albrecht A., Bernstein G., Cahn R., Freedman W. L., Hewitt J., Hu W., Huth J., Kamionkowski M., Kolb E. W., et al., 2006, Astro-ph (arXiv:0609591) *"Report of the Dark Energy Task Force"*

[4] Angulo R., Baugh C. M., Frenk C. S., Lacey C. G., 2008a, MNRAS, 383, 755 *"The detectability of baryonic acoustic oscillations in future galaxy surveys"*

[5] Angulo R., Baugh C. M., Lacey C. G., 2008b, MNRAS, 387, 921 *"The assembly bias of dark matter haloes to higher orders"*

[6] Angulo R.E., Lacey C.G., Baugh C.M., Frenk C.S., 2009, MNRAS, 399, 983 *"The fate of substructures in cold dark matter haloes"*

[7] Aguilar L.A., White S.D.M., 1985, ApJ, 295, 374 *"Tidal interactions between spherical galaxies"*

[8] Almeida C., Baugh C. M., Wake D. A., Lacey C. G., Benson A. J., Bower R. G., Pimbblet K., 2008, MNRAS, 386, 2145 *"Luminous red galaxies in hierarchical cosmologies"*

[9] Baugh C. M., Lacey C. G., Frenk C. S., Benson A. J., Cole S., Granato G. L., Silva L., Bressan A., 2004, NewAR, 48, 1239 *"Predictions for the SKA from hierarchical galaxy formation models"*

[10] Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191 *"Can the faint submillimetre galaxies be explained in the Λ cold dark matter model?"*

[11] Baugh C. M., 2006, RPPh, 69, 310 *"A primer on hierarchical galaxy formation: the semi-analytical approach"*

[12] Benson A. J., Cole, S., Frenk C. S., Baugh C. M., Lacey C. G. 2000, MNRAS, 311, 739 *"The nature of galaxy bias and clustering"*

[13] Benson A. J., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2001, MNRAS, 327, 1041. *"The clustering evolution of the galaxy distribution"*

[14] Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38 *"What Shapes the Luminosity Function of Galaxies?"*

[15] Benson A. J., Lacey C. G., Frenk C. S., Baugh C. M., Cole S., 2004, MNRAS, 351, 1215 *"Heating of galactic discs by infalling satellites"*

[16] Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587 *"The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass"*

[17] Berlind A. A., Weinberg D. H., Benson A. J., Baugh C. M., Cole S., Dav R., Frenk C. S., Jenkins A., et al., 2003, ApJ, 539, 1 *"The Halo Occupation Distribution and the Physics of Galaxy Formation"*

[18] Bertoldi F. et al., 2000, A&A, 360, 92 *"Three high-redshift millimeter sources and their radio and near-infrared identifications"*

[19] Bressan A., Silva L., Granato G. L., 2002, A&A, 392, 377 *"Far infrared and radio emission in dusty starburst galaxies"*

[20] Barnes D. G., Staveley-Smith L., de Blok W. J. G., Oosterloo T., Stewart I. M., Wright A. E., Banks G. D., Bhathal R. et al., 2001, MNRAS, 322, 486 *"The Hi Parkes All Sky Survey: southern observations, calibration and robust imaging"*

[21] Blake C. A., Glazebrook K., 2003, ApJ, 594, 665 *"Probing Dark Energy Using Baryonic Oscillations in the Galaxy Power Spectrum as a Cosmological Ruler"*

[22] Blitz L., Rosolowsky E., 2006, ApJ, 650, 933 *"The Role of Pressure in GMC Formation II: The H_2-Pressure Relation"*

[23] Booth R. S., de Blok W. J. G., Jonas J. L., Fanafoff B., 2009, Astro-ph (arXiv:0910.2935) *"MeerKAT Key Project Science, Specifications, and Proposals"*

[24] Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440 *"Excursion set mass functions for hierarchical Gaussian fluctuations"*

[25] Bower R. G., 1991, MNRAS, 248, 332 *"The evolution of groups of galaxies in the Press-Schechter formalism"*

[26] Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645 *"Breaking the hierarchy of galaxy formation"*

[27] Birnboim Y., Dekel A., 2003, MNRAS, 345, 349 *"Virial shocks in galactic haloes?*

[28] Chandrasekhar S., 1943, ApJ, 97, 255 *"Dynamical Friction. I. General Considerations: the Coefficient of Dynamical Friction."*

[29] Cole S. 1991, ApJ, 367, 45 `Modeling galaxy formation in evolving dark matter halos"

[30] Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781 *"A Recipe for Galaxy Formation"*

[31] Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168 *"Hierarchical galaxy formation"*

[32] Cole S., Norberg P., Baugh C. M., Frenk C. S., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., et al., 2001, MNRAS, 326, 255 *"The 2dF galaxy redshift survey: near-infrared galaxy luminosity functions"*

[33] Cole S., Percival W. J., Peacock J. A., Norberg P., Baugh C. M., Frenk C. S., Baldry I., Bland-Hawthorn J., et al., 2005, MNRAS, 362, 505 *"The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications"*

[34] Cole S., Kaiser N., 1989, MNRAS, 237, 1127 *"Biased clustering in the cold dark matter cosmogony"*

[35] Coles P., 1993, MNRAS, 262, 1065 *"Galaxy formation with a local bias"*

[36] Colless M., Dalton G., Maddox S., Sutherland W., Norberg P., Cole S., Bland-Hawthorn J., Bridges T., et al., 2001, MNRAS, 328, 1039 *"The 2dF Galaxy Redshift Survey: spectra and redshifts"*

[37] Cooray A., Sheth R., 2002, Physics Reports, 372, 1 *"Halo models of large scale structure"*

[38] Croton D. J., Springel V., White S. D. M., De Lucia G., Frenk C. S., Gao L., Jenkins A., Kauffmann G., et al., 2006, MNRAS, 365, 11 *"The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies"*

[39] Cattaneo A., Blaizot J., Weinberg D. H., Keres D., Colombi S., Dav R., Devriendt J., Guiderdoni B., Katz N., 2007, MNRAS, 377, 63 *"Accretion, feedback and galaxy bimodality: a comparison of the GalICS semi-analytic model and cosmological SPH simulations"*

[40] De Lucia G., Kauffmann G., White S. D. M., 2004, MNRAS, 349, 1101 *"Chemical enrichment of the intracluster and intergalactic medium in a hierarchical galaxy formation model"*

[41] De Lucia G., Springel V., White S. D. M., Croton D. J., Kauffmann G, 2006, MNRAS, 366, 499 *"The formation history of elliptical galaxies"*

[42] De Lucia G., Blaizot J., 2007, MNRAS, 375, 2 *"The hierarchical formation of the brightest cluster galaxies"*

[43] Diemand J., Kuhlen M., Madau P., 2008, ApJ, 679, 1680 *"Erratum: "Formation and Evolution of Galaxy Dark Matter Halos and Their Substructure" "*

[44] Dole H. et al., A&A, 372, 364 *"FIRBACK: III. Catalog, source counts, and cosmological implications of the 170 mu m ISO"*

[45] Dole H. et al., 2004, ApJS, 154, 93 *"Confusion of Extragalactic Sources in the Mid- and Far-Infrared: Spitzer and Beyond"*

[46] Eisenstein D. J., Hu W., 1998, ApJ, 496, 605 *"Baryonic Features in the Matter Transfer Function"*

[47] Eisenstein D. J. et al., 2005, ApJ, 633, 560 *"Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies"*

[48] Eke V. R., Frenk C.S., Baugh C.M., Cole S., Norberg P., Peacock J.A., Baldry I., Bland-Hawthorn J., et al., 2004, MNRAS, 355, 769 *"Galaxy groups in the Two-degree Field Galaxy Redshift Survey: the luminous content of the groups"*

[49] Elbaz D., Cesarsky C. J., Chanial P., Aussel H., Franceschini A., Fadda D., Chary R. R., 2002, A&A, 384, 848 *"The bulk of the cosmic infrared background resolved by ISOCAM"*

[50] Fanidakis N., Baugh C.M., Benson A.J., Bower R.G., Cole S., Done C., Frenk C.S., 2009, Astro-ph (arXiv:0911.1128) *"Grand unification of AGN activity in the LambdaCDM cosmology"*

[51] Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23 *"Power-spectrum analysis of three-dimensional redshift surveys"*

[52] Fernandez-Conde N. Lagache G., Puget J.-L., Dole H., 2008, A&A, 481, 885 *"Simulations of the cosmic infrared and submillimeter background for future large surveys. I. Presentation and first application to Herschel/SPIRE and Planck/HFI"*

[53] Font A. S., Bower R. G., McCarthy I. G., Benson A. J., Frenk C. S., Helly J. C., Lacey C. G., Baugh C. M., et al., 2008, MNRAS, 389, 1619 *"The colours of satellite galaxies in groups and clusters"*

[54] Granato G. L., Lacey C. G., Silva L., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, ApJ, 542, 710 *"The Infrared Side of Galaxy Formation. I. The Local Universe in the Semianalytical Framework"*

[55] Granato G. L., De Zotti G., Silva L., Bressan A., Danese L., 2004, ApJ, 600, 580 *"A Physical Model for the Coevolution of QSOs and Their Spheroidal Hosts"*

[56] Genzel R., Cesarsky C. J., 2000, ARA&A, 38, 761 *"Extragalactic Results from the Infrared Space Observatory"*

[57] Giovanelli R., Haynes M. P., Kent B. R., Saintonge A., Stierwalt S., Altaf A., Balonek T., Brosch N., et al., 2007, AJ, 133, 2569 *"The Arecibo Legacy Fast ALFA Survey. III. H I Source Catalog of the Northern Virgo Cluster Region"*

[58] Gnedin O. Y., 2003, ApJ, 589, 752 *"Dynamical Evolution of Galaxies in Clusters"*

[59] Gonzalez J. E., Lacey C. G., Baugh C. M., Frenk C. S., Benson A. J., 2009, MNRAS, 397, 1254 *"Testing model predictions of the cold dark matter cosmology for the sizes, colours, morphologies and luminosities of galaxies with the SDSS"*

[60] Gonzalez-Nuevo J., Toffolatti L., 2005, ApJ, 621, 1 *"Predictions of the Angular Power Spectrum of Clustered Extragalactic Point Sources at Cosmic Microwave Background Frequencies from Flat and All-Sky Two-dimensional Simulations"*

[61] Haiman Z., Knox L., 2000, ApJ, 530, 124 *"Correlations in the Far-Infrared Background"*

[62] Hamilton, A. J. S., 1993, ApJ, 417, 19 *"Better Ways to Measure the Galaxy Correlation Function*

[63] Harker G., Cole S., Helly J., Frenk C. S., Jenkins A., 2006, MNRAS, 367, 1039 *"A marked correlation function analysis of halo formation times in the Millennium Simulation"*

[64] Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75 *''GALICS- I. A hybrid N-body/semi-analytic model of hierarchical galaxy formation"*

[65] Henriques B. M., Bertone S., Thomas P. A., 2008, MNRAS, 383, 1649 *"The effect of dwarf galaxy disruption in semi-analytic models"*

[66] Hnriques B. M. B., Thomas P. A., 2010, MNRAS, 403, 768 *"Tidal disruption of satellite galaxies in a semi-analytic model of galaxy formation*

[67] Huchtmeier, W. K., Richter, O.-G., 1988, A & A, 203, 237 *"H I observations of galaxies in the Kraan-Korteweg-Tammann catalogue of nearby galaxies. III - Global parameters of the galaxies"*

[68] Holland W. S., Greaves J. S., Zuckerman B., Webb R. A., McCarthy C., Coulson I. M., Walther D. M., Dent W. R. F. et al., 1998, Nature, 392, 788 *"Submillimetre images of dusty debris around nearby stars"*

[69] Jenkins A., Frenk C.S., White S.D.M., Colberg J.M., Cole S., Evrard A.E., Couchman H.M.P., Yoshida N., 2001, MNRAS, 321, 372 *"The mass function of dark matter haloes"*

[70] Jiang C. Y., Jing Y. P., Faltenbacher A., Lin W. P., Li C., 2008, ApJ, 675, 1095 *"A Fitting Formula for the Merger Timescale of Galaxies in Hierarchical Clustering"*

[71] Jiang C. Y., Jing Y.P., Lin W.P., 2010, A&A, 510, 60 *"Influence of baryonic physics on the merger timescale of galaxies in N-body/hydrodynamical simulations"*

[72] Jing Y. P., Borner G., 2004, ApJ, 617, 782 *"The Pairwise Velocity Dispersion of Galaxies: Luminosity Dependence and a New Test of Galaxy Formation Models"*

[73] Jennings E., Baugh C. M., Angulo R. E., Pascoli, S., 2010, MNRAS, 401, 2181 *"Simulations of quintessential cold dark matter: beyond the cosmological constant"*

[74] Johnston et al., 2008, ExA, 22, 151 *"Science with ASKAP. The Australian square-kilometre-array pathfinder"*

[75] Kang X., Jing, Y. P., Mo H. J., Borner G., 2005, ApJ, 631, 21 *"Semianalytical Model of Galaxy Formation with High-Resolution N-Body Simulations"*

[76] Kaiser N., 1987, MNRAS, 227, 1 *"Clustering in real space and in redshift space"*

[77] Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201 *"The Formation and Evolution of Galaxies Within Merging Dark Matter Haloes"*

[78] Kauffmann G., Nusser A., Steinmetz M., 1997, MNRAS, 286, 795 *"Galaxy formation and large-scale bias"*

[79] Kauffmann G., Colberg J.M., Diaferio A., White S.D.M., 1999, MNRAS, 303, 188 *"Clustering of galaxies in a hierarchical universe - I. Methods and results at z=0"*

[80] Keres D., Katz N., Weinberg D. H, Dave R., 2005, MNRAS, 363, 2 *"How do galaxies get their gas?*

[81] Kennicutt R. C. Jr., 1983, ApJ, 272, 54 *"The rate of star formation in normal disk galaxies"*

[82] Kennicutt R. C. Jr., 1998, ARA&A, 36, 189 *"Star Formation in Galaxies Along the Hubble Sequence"*

[83] Kim H. S., Baugh C. M., Cole S., Frenk C. S., 2009, MNRAS, 400, 1527 *"Modelling galaxy clustering: is new physics needed in galaxy formation models?"*

[84] Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlober S., Allgood B. Primack J. R., 2004, ApJ, 609, 35 *"The Dark Side of the Halo Occupation Distribution*

[85] Krick J. E., Bernstein R. A., 2007, AJ, 134, 466 *"Diffuse Optical Light in Galaxy Clusters. II. Correlations with Cluster Properties"*

[86] Krumholz M. R., McKee C. F., Tumlinson J., 2009, ApJ, 693, 216 *"The Atomic-to-Molecular Transition in Galaxies. II: H I and H_2 Column Densities"*

[87] Kroupa P., Aarseth S., Hurley J., 2001, MNRAS, 321, 699 *"The formation of a bound star cluster: from the Orion nebula cluster to the Pleiades"*

[88] Lacey C., Cole S., 1993, MNRAS, 262, 627 *"Merger rates in hierarchical models of galaxy formation"*

[89] Lacey C. G., Baugh C. M., Frenk C. S., Silva L., Granato G. L., Bressan A., 2008, MNRAS, 385, 1155 *"Galaxy evolution in the infrared: comparison of a hierarchical galaxy formation model with Spitzer data"*

[90] Lacey C. G., Baugh C. M., Frenk C. S., Benson A. J., Orsi A., Silva L., Granato G. L., Bressan A., 2010, MNRAS, 405, 2 *"Predictions for Herschel from Λ-cold dark matter: unveiling the cosmic star formation history"*

[91] Lagache G., Dole H., Puget J.-L., 2003, MNRAS, 338, 555 *"Predictions for Cosmological Infrared Surveys from Space with the Multiband Imaging Photometer for SIRTF"*

[92] Lah P., et al. 2009, MNRAS, 399, 1447 *"The HI gas content of galaxies around Abell 370, a galaxy cluster at z = 0.37"*

[93] Lagos C. D. P., Cora S.A., Padilla N.D., 2008, MNRAS, 388, 587 *"Effects of AGN feedback on ΛCDM galaxies"*

[94] Landy S.D., Szalay A. S., 1993, ApJ, 412, 64 *"Bias and variance of angular correlation functions*

[95] Laureijs R. et al., 2009, arXiv : 0912.0914 *"Euclid Assessment Study Report for the ESA Cosmic Visions"*

[96] Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473 *"Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models"*

[97] Li C., Kauffmann G., Jing Y.P., White S.D.M., Borner G., Cheng F.Z., 2006, MNRAS, 368, 21 *"The dependence of clustering on galaxy properties"*

[98] Li G. L., Mao S., Jing Y. P., Lin W. P., Oguri M., 2007, MNRAS, 378, 496 *"Properties of wide-separation lensed quasars by clusters of galaxies in the Sloan Digital Sky Survey"*

[99] Li C., Jing, Y. P., Kauffmann G., Borner G., Kang X., Wang L., 2007, MNRAS, 376, 984 *"Luminosity dependence of the spatial and velocity distributions of galaxies: semi-analytic models versus the Sloan Digital Sky Survey"*

[100] Li C., White S. M. D., 2009, MNRAS, 398, 2177 *"The distribution of stellar mass in the low-redshift Universe"*

[101] Le Delliou M., Lacey C., Baugh C. M., Guiderdoni B., Bacon R., Courtois H., Sousbie T., Morris S. L., 2005, MNRAS, 357, 11 *"The abundance of Lyα emitters in hierarchical models"*

[102] Le Delliou M., Lacey C., Baugh C. M., Morris S. L., 2006, MNRAS, 365, 712 *"The properties of Lyα emitting galaxies in hierarchical galaxy formation models"*

[103] Madgwick D. S., Hawkins E., Lahav O., Maddox S., Norberg P., Reacock J. A., Baldry I. K., Baugh C. M., et al., 2003, MNRAS, 344, 847 *"The 2dF Galaxy Redshift Survey: galaxy clustering per spectral type"*

[104] Malbon R. K.; Baugh C. M.; Frenk C. S.; Lacey C. G., 2007, MNRAS, 382, 1394 *"Black hole growth in hierarchical galaxy formation"*

[105] Magliocchetti M., Porciani C., 2003, MNRAS, 346, 186 *"The halo distribution of 2dF galaxies"*

[106] Marin F. A., Gnedin N. Y., Seo H. J., Vallinotto A., 2010, ApJ, 718, 972 *"Modeling the Large-scale Bias of Neutral Hydrogen"*

[107] McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, MNRAS, 383, 593 *"Ram pressure stripping the hot gaseous haloes of galaxies in groups and clusters"*

[108] Meyer M. J., Zwaan M. A., Webster R. L., Staveley-Smith L., Ryan-Weber E., Drinkwater M. J., Barnes D. G., Howlett M., Kilborn V. A, Stevens J. et al., 2004, MNRAS,350,1195 *"The HIPASS catalogue - I. Data presentation"*

[109] Meyer M. J., Zwaan M. A., Webster R. L., Brown M. J. I., Staveley-Smith. L., 2007, ApJ, 654, 702 *"The Weak Clustering of Gas-rich Galaxies"*

[110] Mo H.J., White S.D.M., 1996, MNRAS, 282, 347 *"An analytic model for the spatial clustering of dark matter haloes"*

[111] Monaco P., Fontanot F., Taffoni G., 2007, MNRAS, 375, 1189 *"The MORGANA model for the rise of galaxies and active nuclei"*

[112] Moore B., Katz N., Lake G., 1996, ApJ, 457, 455 *"On the Destruction and Overmerging of Dark Halos in Dissipationless N-Body Simulations"*

[113] Negrello M., Perrotta F., Gonzalez-Nuevo J., Silva L., de Zotti G., Granato G. L., Baccigalupi C., Danese L., 2007, MNRAS, 377, 157 *""*

[114] Norberg P., Baugh C. M., Hawkins Ed., Maddox S., Peacock J. A., Cole S., Frenk C. S., Bland-Hawthorn J., et al., 2001, MNRAS, 328, 64 *"The 2dF Galaxy Redshift Survey: luminosity dependence of galaxy clustering"*

[115] Norberg P., Baugh C. M., Hawkins Ed., Maddox S., Madgwick D., Lahav O., Cole S., Frenk C. S., et al., 2002, MNRAS, 332, 827 *"The 2dF Galaxy Redshift Survey: the dependence of galaxy clustering on luminosity and spectral type"*

[116] Norberg P., Cole S., Baugh C. M., Frenk C. S., Baldry I., Bland-Hawthorn J., Bridges T., Cannon R., et al., 2002, MNRAS, 336, 907 *"The 2dF Galaxy*

Redshift Survey: the b_J-band galaxy luminosity function and survey selection function"

[117] Norberg P., Baugh C. M., Gaztanaga E., Croton D. J., 2009, MNRAS, 396, 19 *"Statistical analysis of galaxy surveys - I. Robust error estimation for two-point clustering statistics"*

[118] Obreschkow D., Rawlings S., 2009, MNRAS, 394, 1857 *"Understanding the H_2/HI ratio in galaxies"*

[119] Okamoto T., Gao Liang, Theuns T., 2008, MNRAS, 390, 920 *"Mass loss of galaxies due to an ultraviolet background"*

[120] Orsi A., Lacey C. G., Baugh C. M., Infante L., 2008, MNRAS, 391, 1589 *"The clustering of Lyα emitters in a ΛCDM Universe"*

[121] Orsi A., Baugh C. M., Lacey C. G., Cimatti A., Wang Y., Zamorani G., 2010, MNRAS, 405, 1006 *"Probing dark energy with future redshift surveys: a comparison of emission line and broad-band selection in the near-infrared"*

[122] Papovich C. et al., 2004, ApJS, 154, 70 *"The 24 Micron Source Counts in Deep Spitzer Space Telescope Surveys"*

[123] Parkinson C. D., Cole S., Helly J., 2008, MNRAS, 383, 557 *"Generating dark matter halo merger trees"*

[124] Peroux C., McMahon R. G., Storrie-Lombardi L. J., Irwin M. J., 2003, MNRAS, 346, 1103 *"The evolution of Ω_{HI} and the epoch of formation of damped Lyman α absorbers"*

[125] Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144 *"Halo occupation numbers and galaxy bias"*

[126] Popping A., Dave R., Braun R., Oppenheimer B. D., 2009, A&A, 504, 15 *"The simulated H I sky at low redshift"*

[127] Power C., Baugh C. M., Lacey C. G., 2010, MNRAS, 406, 43 *"The redshift evolution of the mass function of cold gas in hierarchical galaxy formation models"*

[128] Rawlings S., Abdalla F. B., Bridle S. L., Black C. A., Baugh C. M., Greenhill L. J., van der Hulst J. M., 2004, NewAR, 48, 1013 *"Galaxy evolution, cosmology and dark energy with the Square Kilometer Array"*

[129] Richstone D. O., 1976, ApJ, 294, 642 *"Collisions of galaxies in dense clusters. II - Dynamical evolution of cluster galaxies"*

[130] Righi M., Hernndez-Monteagudo C., Sunyaev R. A., 2008, A&A, 478, 685 *"Carbon monoxide line emission as a CMB foreground: tomography of the star-forming universe with different spectral resolutions"*

[131] Sage L. J., 1993, A & A, 272, 123 *"Molecular Gas in Nearby Galaxies - Part One - Co/ Observations of a Distance-Limited Sample"*

[132] Sanchez A. G., Baugh C. M., Angulo R., 2008, MNRAS, 390, 1470 *"What is the best way to measure baryonic acoustic oscillations?"*

[133] Sanchez A. G., Crocce M., Cabre A., Baugh C. M., Gaztanaga E., 2009, MNRAS, 400, 1643 *"Cosmological parameter constraints from SDSS luminous red galaxies: a new treatment of large-scale clustering"*

[134] Schilizzi R.T., Dewdney P. E. F., Lazio T. J. W., 2008, Proc. SPIE, 7012, 70121I *"The Square Kilometre Array"*

[135] Seljak U., 2000, MNRAS, 318, 2039 *"Analytic model for galaxy and dark matter clustering"*

[136] Seo H. J., Eisenstein D. J., 2003, ApJ, 598, 720 *"Probing Dark Energy with Baryonic Acoustic Oscillations from Future Large Galaxy Redshift Surveys"*

[137] Sheth R. K., Tormen G., 1999, MNRAS, 308, 119 *"Large-scale bias and the peak background split*

[138] Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1 *"Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes"*

[139] Silva L., Granato G. L., Bressan A., Danese L, 1998, ApJ, 509, 103 *"Modeling the Effects of Dust on Galactic Spectral Energy Distributions from the Ultraviolet to the Millimeter Band"*

[140] Smith R. E., Peacock J. A., Jenkins A., White S. D. M., Frenk C. S., Pearce F. R., Thomas P. A., Efstathiou G., Couchman H. M. P., 2003, MNRAS, 341, 1311 *"Stable clustering, the halo model and non-linear cosmological power spectra"*

[141] Song Y. -S., Cooray A., Knox L., Zaldarriaga M., 2003, ApJ, 590, 664 *"The Far-Infrared Background Correlation with Cosmic Microwave Background Lensing"*

[142] Sunyaev R. A., Zeldovich Ia. B., 1980, ARA&A, 18, 537 *"Microwave background radiation as a probe of the contemporary structure and history of the universe"*

[143] Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726 *"Populating a cluster of galaxies - I. Results at [formmu2]z=0"*

[144] Springel V., White S. D. M., Jenkins A., Frenk C. S., et al. 2005, Nat, 435, 629 *"Simulations of the formation, evolution and clustering of galaxies and quasars"*

[145] Springel V., Frenk C. S., White S. D. M., 2006, Nat, 440, 1137 *"The large-scale structure of the Universe"*

[146] Springel V., Wang J., Vogelsberger M., Ludlow A., Jenkins A., Helmi A., Navarro J. F., Frenk C. S., et al., 2008, MNRAS, 391, 1685 *"The Aquarius Project: the subhaloes of galactic haloes"*

[147] Storrie-Lombardi L. J., Irwin M. J., McMahon R. G., 1996, MNRAS, 282, 1330 *"APM z>∼4 survey: distribution and evolution of high column density HI absorbers"*

[148] Taylor J. E., Babul A., 2001, ApJ, 559, 716 *"The Dynamics of Sinking Satellites around Disk Galaxies: A Poor Man's Alternative to High-Resolution Numerical Simulations"*

[149] Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22 *"Karhunen-Loeve Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets?"*

[150] Tinker J. L., Norberg P., Weinberg D. H., Warren M. S., 2007, ApJ, 659, 877 *"On the Luminosity Dependence of the Galaxy Pairwise Velocity Dispersion"*

[151] van den Bosch F. C., Mo H. J., Yang X., 2003, MNRAS, 345, 923 *"Towards cosmological concordance on galactic scales"*

[152] Vega O., Clemens M. S., Bressan A., Granato G. L., Silva L., Panuzzo P., 2008, A&A, 484, 631 *"Modelling the spectral energy distribution of ULIRGs. II. The energetic environment and the dense interstellar medium"*

[153] Vielva P., Martnez-Gonzlez E., Gallegos J. E., Toffolatti L., Sanz J. L., 2003, MNRAS, 344, 89 *"Point source detection using the Spherical Mexican Hat Wavelet on simulated all-sky Planck maps"*

[154] Verheijen M., van Gorkom J. H., Szomoru A., Dwarakanath K. S., Poggianti B. M., Schiminovich D., et al., 2007, ApJ, 668, 9 *"WSRT Ultradeep Neutral Hydrogen Imaging of Galaxy Clusters at z ∼ 0.2: A Pilot Survey of Abell 963 and Abell 2192"*

[155] Wake D. A. et al., 2008, MNRAS, 387, 1045 *"The 2dF-SDSS LRG and QSO Survey: evolution of the clustering of luminous red galaxies since z = 0.6"*

[156] Waugh M., Drinkwater M. J., Webster R. L., Staveley-Smith L., Kilborn V. A., Barnes D. G., Bhathal R., de Blok W. J. G. et al., 2002, MNRAS, 337, 641 *"The large-scale distribution of neutral hydrogen in the Fornax region"*

[157] Welch G. A., Sastry G. N., 1971, ApJ, 169, 3 *"Photographic Detection of "intergalactic" Matter in the Coma Cluster"*

[158] Wetzel A. R., Cohn J. D., White M., 2009, MNRAS, 395, 1376 *"Simulating subhaloes at high redshift: merger rates, counts and types"*

[159] Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, MNRAS, 366, 2 *"Properties of galaxy groups in the Sloan Digital Sky Survey - I. The dependence of colour, star formation and morphology on halo mass"*

[160] White S. D. M., Rees M. J., 1978, MNRAS, 183, 341 *"Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering"*

[161] White S. D. M., Frenk C. S., 1991, ApJ, 379, 52 *"Galaxy formation through hierarchical clustering"*

[162] Wootten A., 2008, Astrophysics Space Sci, 313, 9 *"ALMA capabilities for observations of spectral line emission"*

[163] Wyithe J. S. B., Brown M. J. I., Zwaan M. A., Meyer M., 2009, Astro-ph (arXiv:0908.2854) *"The Halo Occupation Distribution of HI Galaxies"*

[164] Wyithe J.S.B., Brown M.J.I., 2010, MNRAS, 404, 876 *"The halo occupation distribution of HI from 21-cm intensity mapping at moderate redshifts"*

[165] Yang X., Mo H. J., Jing Y.P., van den Bosch F. C., Chu Y., 2004, MNRAS, 350, 1153 *"Populating dark matter haloes with galaxies: comparing the 2dF-GRS with mock galaxy redshift surveys"*

[166] Yang X., Mo H. J., van den Bosch F. C., 2009, ApJ, 693, 830 *"The Subhalo-Satellite Connection and the Fate of Disrupted Satellite Galaxies"*

[167] York D. G., Adelman J., Anderson John E. Jr., Anderson S. F., Annis J., Bahcall N. A., Bakken J. A., Barkhouser R., et al., 2000, AJ, 120, 1579 *"The Sloan Digital Sky Survey: Technical Summary"*

[168] Zehavi I., Blanton M. R., Frieman J. A., Weinberg D. H., Mo H. J., Strauss M. A., Anderson S. F., Annis J., et al., 2002, ApJ, 571, 172 *"Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data"*

[169] Zehavi I., Zheng Z., Weinberg D. H., Frieman J. A., Berlind A. A., Blanton M. R., Scoccimarro R., Sheth R. K., et al., 2005, ApJ, 630, 1 *"The Luminosity and Color Dependence of the Galaxy Correlation Function"*

[170] Zehavi et al., 2010, Astro-ph, arXiv:1005.2413 *"Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity"*

[171] Zheng Z., Berlind A. A., Weinberg D. H., Benson A. J., Baugh C. M., Cole S., Dav R., Frenk C. S., et al., 2005, ApJ, 633, 791 *""*

[172] Zheng Z., Coil A. L., Zehavi I., 2007, ApJ, 667, 760 *"Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies"*

[173] Zibetti S., 2008, IAUS, 244, 176 *"Statistical Properties of the IntraCluster Light from SDSS Image Stacking"*

[174] Zibetti S., White S. D. M., Schneider D. P., Brinkmann J., 2005, MNRAS, 358, 949 *"Intergalactic stars in $z \sim 0.25$ galaxy clusters: systematic properties from stacking of Sloan Digital Sky Survey imaging data"*

[175] Zwaan et al., 2003, AJ, 125, 2842 *"The 1000 Brightest HIPASS Galaxies: The H I Mass Function and Ω_{HI}"*

[176] Zwaan M. A., Meyer M. J., Satveley-Smith L., Webster R. L., 2005, MNRAS, 359, 30 *"The HIPASS catalogue: Ω_{HI} and environmental effects on the HI mass function of galaxies"*

[177] Zwicky F., 1951, PASP, 63, 61 *"The Coma Cluster of Galaxies"*