



Durham E-Theses

Galaxy formation and clustering in a hierarchical universe

Benson, Andrew John

How to cite:

Benson, Andrew John (2000) *Galaxy formation and clustering in a hierarchical universe*, Durham theses, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/4288/>

Use policy

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

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

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

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

Galaxy Formation and Clustering in a Hierarchical Universe

by Andrew John Benson

June 2000

Abstract

In this Thesis we describe a semi-analytic model of galaxy formation. We apply the model to the problem of galaxy clustering and show that the dependence of galaxy formation efficiency on halo mass leads to a scale-dependent bias in the distribution of galaxies relative to the distribution of mass. Remarkably, this results in a correlation function in a flat, $\Omega_0 = 0.3$, CDM model that is close to a power-law over four orders of magnitude in amplitude and which agrees well with the correlation function of galaxies measured in the APM survey. The galaxy velocity dispersion is $\sim 40\%$ lower than that of the dark matter. Biases cause the redshift space correlation functions of model galaxies and dark matter to be remarkably similar to each other. A dependence of clustering strength on galaxy luminosity exists for extremely bright galaxies and for galaxies selected either by morphology or by colour.

We present predictions for the reionization of the intergalactic medium by stars in high-redshift galaxies, including the effects of absorption by interstellar gas and dust. We combine our model with an N-body simulation to calculate the temperature anisotropies induced in the cosmic microwave background by reionization.

Finally, we test key aspects of the model. We use ROSAT PSPC data to search for extended X-ray emission from the halos of three nearby, massive, late-type galaxies. The luminosity lies well below the model prediction. We discuss this discrepancy and consider a number of possible explanations.

By comparing the statistical properties of galaxies in our model with those of galaxies formed in cosmological hydrodynamics simulations we show that the two techniques produce broadly consistent predictions. However, individual statistics, such as the galaxy mass function, may differ by factors of 2–4. We identify possible reasons for these discrepancies, thereby highlighting avenues for future work to explore.

Galaxy Formation and Clustering in a Hierarchical Universe

by Andrew John Benson

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

Department of Physics

University of Durham

June 2000

The copyright of this thesis rests
with the author. No quotation
from it should be published
without the written consent of the
author and information derived
from it should be acknowledged.



14 NOV 2000

Contents

1	Introduction	1
1.1	Historical Background	1
1.1.1	Early Ideas	1
1.1.2	The Origin of Galaxies	2
1.1.3	Further Observational Developments	6
1.2	The State of the Art	7
1.3	Motivation for this Work	9
2	A Semi-Analytic Model of Galaxy Formation	11
2.1	Dark Matter Halos	11
2.1.1	Merger Trees	11
2.1.2	Density Profile	18
2.1.3	Spin Distribution	19
2.2	Gas Dynamics	20
2.3	Disk and Star Formation	22
2.3.1	Disk Formation	22
2.3.2	Star Formation	23
2.3.3	Supernovae Feedback	24
2.3.4	Chemical Enrichment	24
2.3.5	Gas Equations	25
2.4	Galaxy Merging	26
2.4.1	Dynamical Friction	26
2.4.2	Spheroid Formation	28
2.4.3	Sizes	28
2.4.4	The Merger Rule	31
2.5	Spectral Synthesis	33

2.5.1	Stellar Spectra	33
2.5.2	Emission Line Modelling	34
2.5.3	Dust Extinction	34
2.6	Using the Model	35
2.6.1	Parameters and Outputs	36
2.6.2	Observational Constraints	38
2.7	Fiducial Model Results	39
2.7.1	Model Parameters	39
2.7.2	Basic Results	39
2.8	Discussion and Conclusions	52
3	The Nature of Galaxy Bias and Clustering	53
3.1	Introduction	53
3.2	Description of the Model	55
3.2.1	Semi-Analytic Models	55
3.2.2	Incorporation into N-body Simulations	56
3.2.3	Reference Models	57
3.3	Clustering of Galaxies	59
3.3.1	The Galaxy Two-Point Correlation Function	59
3.3.2	Systematic Effects	66
3.4	The Nature of Bias	70
3.5	Testing the Robustness of the Predictions	78
3.5.1	Models Constrained by the Luminosity Function	80
3.5.2	Models Constrained by the Tully-Fisher Relation	86
3.6	Discussion and Conclusions	88
4	The Dependence of Velocity and Clustering Statistics on Galaxy Prop-	
	erties	93
4.1	Introduction	93
4.2	Description of the Numerical Technique	96
4.3	Redshift Space Distortions	97
4.3.1	Galaxy Peculiar Motions	100
4.3.2	Comparison to Previous Work	104
4.4	The Dependence of Clustering on Galaxy Properties	110
4.4.1	Dependence of Clustering on Luminosity	110

4.4.2	Dependence of Clustering on Morphology	117
4.4.3	Dependence of Clustering on Colour	120
4.5	Discussion and Conclusions	120
5	Non-Uniform Reionization by Galaxies and its Effect on the CMB	125
5.1	Introduction	125
5.2	The Semi-Analytic Model of Galaxy Formation	127
5.2.1	Gas Cooling	129
5.2.2	Fraction of Gas in the IGM	131
5.2.3	Observational Constraints & Modelling Considerations	131
5.3	The Escape Fraction of Ionizing Photons	137
5.3.1	Global Constraints at Low Redshift	137
5.3.2	The Dependence of $f_{\text{esc,gas}}$ on Redshift and on Halo Mass	139
5.4	The Filling Factor and the Evolution of the Ionization Fronts	145
5.4.1	Model Results	150
5.5	Sensitivity of Results to Model Parameters	155
5.6	Spatial Distribution of Ionizing Sources and CMB Fluctuations	161
5.6.1	Spatial Distribution	161
5.6.2	CMB Fluctuations	165
5.7	Discussion and Conclusions	173
6	Diffuse X-ray Emission from Late-Type Galaxy Halos	175
6.1	X-rays from Galaxy Formation	175
6.2	Observations	179
6.2.1	Target Selection	179
6.2.2	Data Reduction	180
6.2.3	Uncertainties in the Observations	183
6.3	Discussion	186
6.4	Conclusions	191
7	A Comparison of Semi-Analytic and Smoothed Particle Hydrodynamics	
	Galaxy Formation	195
7.1	Introduction	195
7.2	The SPH Simulations and Semi-Analytic Models	198
7.2.1	SPH Simulations	198

7.2.2	Semi-Analytic Models of Galaxy Formation	201
7.3	Comparison of the Two Models	206
7.3.1	Properties of Halo Gas	206
7.3.2	Properties of Galaxies	215
7.3.3	Spatial Distribution of Galaxies	224
7.4	Discussion and Conclusions	225
8	Conclusions	233
8.1	Applications of the Model	233
8.1.1	Galaxy Clustering	233
8.1.2	Reionization	234
8.2	Tests of the Model Ingredients	234
8.2.1	The Growth of Galactic Disks	234
8.2.2	Comparison to Simulations of Gas Hydrodynamics	235
8.3	Future Prospects	235
A	A Semi-Analytic Model of Galaxy Formation: Supporting Theory	239
A.1	Fundamental Cosmology	239
A.1.1	The Evolution of a Homogeneous and Isotropic Universe	239
A.1.2	Linear Growth of Perturbations	241
A.1.3	Spherical Top-Hat Collapse	243
A.2	Halo Angular Momentum	247
A.3	Gas Equations	249
A.4	Dynamical Friction Timescale	251
B	Non-Uniform Reionization by Galaxies: Supporting Theory	255
B.1	Calculation of the Escaping Fraction	255
B.1.1	Escaping Fraction in the DS94 Model: Stars in Mid-Plane	255
B.1.2	Escaping Fraction in DS94 Model: Stars Tracing Gas	257
B.1.3	Escaping Fraction in the DSGN98 Model	260
B.1.4	Escaping Fraction in Starbursts	261
B.2	Calculation of Clumping Factor	263
B.3	The Spectrum of CMB Secondary Anisotropies	266

C	Models for the X-ray Emission from Cooling Gas	269
C.1	Calculation of X-ray Luminosity in Specific Halo Potentials	271
C.1.1	Cooling in an Isothermal Halo	272
C.1.2	Gas with a Constant Density Core	273
C.1.3	Cooling in an NFW Halo	274
C.1.4	The Influence of the Central Galaxy	276

List of Figures

2.1	Example overdensity trajectory	13
2.2	Flow of gas between three phases in the semi-analytic model	25
2.3	The b_J and K-band luminosity functions in the fiducial models	41
2.4	The I-band Tully-Fisher relation in the fiducial models	43
2.5	The sizes of galaxy disks predicted by the fiducial models	45
2.6	The cold gas content of spiral and irregular galaxies as a function of their luminosity	48
2.7	The variation in metallicity with luminosity in the fiducial models	50
2.8	The luminous star formation rate per unit comoving volume as a function of redshift	51
3.1	Mass function of halos containing galaxies with $M_B - 5 \log h \leq -19.5$ in our τ CDM and Λ CDM models	56
3.2	B and K band luminosity functions for the τ CDM and Λ CDM reference models	57
3.3	Two-point correlation functions for galaxies brighter than $M_B - 5 \log h =$ -19.5 in a Λ CDM cosmology	60
3.4	Tully-Fisher relations in the τ CDM and Λ CDM reference models	62
3.5	The locations of galaxies brighter than $M_B - 5 \log h = -19.5$ in our two reference models	64
3.6	The two-point correlation functions of galaxies brighter than $M_B - 5 \log h =$ -19.5 in our two reference models	65
3.7	The galaxy correlation function in the Λ CDM reference model, with and without the effects of dynamical friction on satellite galaxy positions	68

3.8	A comparison of correlation functions for galaxies brighter than $M_B - 5 \log h = -19.5$ in the Λ CDM reference model with the results of a model with an artificial mass resolution designed to mimic the models of Kauffmann et al. (1999a)	69
3.9	The B-band mass-to-light ratio of halos in the reference models	71
3.10	Mean number of galaxies and galaxy pairs brighter than $M_B - 5 \log h = -19.5$ per halo as a function of halo mass for the two reference models	72
3.11	Correlation functions constructed from different samples of dark matter particles compared to the observed and model galaxy correlation functions in the two reference models	74
3.12	The probability, $P(N)$, of occupation by N galaxies (multiplied by $N(N - 1)$ for clarity) for halos in the two reference models	76
3.13	The luminosity function of all galactic systems shown for the two reference models	79
3.14	Galaxy correlation functions for variant models in the τ CDM and Λ CDM cosmologies	83
3.15	B-band luminosity functions from the variant models in the τ CDM and Λ CDM cosmologies	84
3.16	The average number of galaxies brighter than $M_B - 5 \log h = -19.5$ per halo as a function of halo mass for variant models in the τ CDM and Λ CDM cosmologies	84
3.17	The B-band luminosity function and galaxy correlation function for two τ CDM models constrained to match the I-band Tully-Fisher relation	87
4.1	Real and redshift-space correlation functions in the Λ CDM model	98
4.2	Pairwise velocities of dark matter particles and galaxies measured in the Λ CDM simulation	101
4.3	The three-dimensional velocity dispersion of individual galaxies and dark matter in spheres of radius r measured relative to the mean velocity within the sphere	103
4.4	Line-of-sight velocity dispersions, σ_{los} , for dark matter and galaxies from our own model and from that of Kauffmann et al. (1999a)	105
4.5	The number of galaxies per halo brighter than $M_B - 5 \log h = -18.5$ as a function of halo mass in our model and in that of Kauffmann et al. (1999a)	106

4.6	The line-of-sight pairwise velocity dispersion, σ_{los} , for galaxies with $M_{\text{B}} - 5 \log h \leq -19.5$ from the variant models	109
4.7	Redshift-space correlation functions of galaxies selected by b_{J} -band absolute magnitude	111
4.8	Redshift-space correlation functions of galaxies brighter than $M_{b_{\text{J}}} - 5 \log h = -20.5$ and -21.5 measured in the 512 ³ simulation	112
4.9	Real-space correlation functions of galaxies selected by b_{J} -band absolute magnitude	113
4.10	Real-space correlation functions of galaxies selected by b_{J} -band absolute magnitude from the variant models	114
4.11	The local b_{J} -band luminosity function for the variant models	115
4.12	Real-space two-point correlation function for galaxy samples selected by morphological type	118
4.13	The mass function of host dark matter halos, weighted by the number of galaxies with $-20.0 < M_{b_{\text{J}}} - 5 \log h < -19.0$ per halo, for late type galaxies and early type galaxies in our model	119
4.14	The real-space correlation function of galaxies brighter than $M_{\text{B}} - 5 \log h = -19.5$ in our model split by B-R colour	121
5.1	The fraction of baryons remaining in the IGM as a function of redshift	132
5.2	The comoving number density of galaxies brighter than a given ionizing luminosity as a function of redshift	135
5.3	The distribution of DLAS at $z = 1, 2$ and 3 respectively as a function of their HI column density and the HI mass function of galaxies at $z = 0.0$	136
5.4	The $\text{H}\alpha$ luminosity function at $z = 0$	140
5.5	The $\text{H}\alpha$ luminosity density of the Universe as a function of redshift	141
5.6	The escape fraction, f_{esc} , at $z = 0$ as a function of halo mass	143
5.7	The variation in (ionizing luminosity-weighted) mean disk scale length and gas mass for galaxies in our model as a function of halo mass	146
5.8	The (ionizing luminosity-weighted) mean escape fraction for all galaxies as a function of redshift	147
5.9	The gas clumping factor, f_{clump} , as a function of redshift	151
5.10	The filling factor, F_{fill} , as a function of redshift determined using three different models for the absorption of ionizing photons by gas inside galaxies	152

5.11	The ratio of the total number of ionizing photons which have escaped into the IGM per comoving volume by redshift z to the comoving number density of hydrogen nuclei in the IGM	154
5.12	The $z = 0$ $H\alpha$ luminosity functions of our variant models	157
5.13	The projected density of ionized gas in a slice through the Λ CDM N-body simulation at $z = 3$ shown as a greyscale image, with the densest regions being black	164
5.14	Filling factors as a function of redshift in the N-body simulation	166
5.15	The secondary CMB anisotropies measured from the simulation	167
5.16	The effect of varying the escape fraction, $f_{\text{esc,gas}}$, on the secondary CMB anisotropies	168
5.17	The effect of varying Ω_b on the secondary CMB anisotropies	172
6.1	Source images for each galaxy in the sample	182
7.1	The mass function of dark matter halos at $z = 2$ which are progenitors of present-day halos of mass 10^{12} – $10^{13}h^{-1}M_\odot$ and 10^{13} – $10^{14}h^{-1}M_\odot$ in the SCDM and Λ CDM cosmologies	202
7.2	The global fraction of gas in each of three phases: uncollapsed gas, hot halo gas and galaxy gas	207
7.3	The global fraction of gas in each of three phases: uncollapsed gas, hot halo gas and galaxy gas for the Λ CDM cosmology with $N'_{\text{SPH}} = 32$	209
7.4	The mass of gas relative to the total (i.e. dark matter plus gas) halo mass at $z = 0$ in progenitor galaxies, as a function of present-day halo mass	210
7.5	The mean mass of gas in the galaxy phase relative to the total (i.e. dark matter plus gas) mass of the $z = 0$ halo, scaled by the universal baryon fraction, as a function of redshift in progenitors of present-day halos of mass 10^{13} – $10^{14}h^{-1}M_\odot$ and 10^{14} – $10^{15}h^{-1}M_\odot$	212
7.6	The mass of gas relative to the total (i.e. dark matter plus gas) halo mass at $z = 0$ in progenitor galaxies, as a function of present-day halo mass in the Λ CDM cosmology	213
7.7	The differential galaxy mass function in the SCDM and Λ CDM cosmologies at various epochs	216
7.8	The differential galaxy mass function at $z = 0$ in the Λ CDM cosmology shown with and without the fix to the extended Press-Schechter theory	217

7.9	The mass function of galaxies which are progenitors of present-day galaxies with (cold gas plus stellar) masses $10^{11.5}-10^{12}h^{-1}M_{\odot}$	218
7.10	The differential galaxy mass function at selected redshifts in the Λ CDM cosmology predicted by the SDSA model when (i) using a faster merging rate and (ii) keeping the core radius of the hot gas density profile fixed . .	222
7.11	The mean number of progenitor galaxies with mass greater than 64 SPH particles and 128 particles in the Λ CDM cosmology, as a function of redshift, for halos of present-day mass $10^{13}-10^{14}h^{-1}M_{\odot}$ and $10^{14}-10^{15}h^{-1}M_{\odot}$	226
7.12	The mean number of progenitor galaxies with mass greater than 64 SPH particles and 128 particles, as a function of redshift in halos of present-day mass $10^{13}-10^{14}h^{-1}M_{\odot}$ and $10^{14}-10^{15}h^{-1}M_{\odot}$ in the Λ CDM cosmology shown in the cases where (i) a faster merger rate is used and (ii) the cooling rate is enhanced by keeping the core radius fixed in the hot gas density profile	227
7.13	Two-point galaxy correlation functions from the SPH, FSA and SDSA models	230
7.14	The Λ CDM galaxy two-point correlation function in the SPH and SDSA models at $z = 0$ shown for a model with an enhanced merger rate, and for a model with a fixed gas core radius	231
B.1	The average escape fractions for a galactic disk in the DS94 model with OB associations in the disk mid-plane and distributed as the cold gas . .	258
B.2	Mean escape fraction for a galaxy disk in the DSGN98 model as a function of τ_0	262

List of Tables

2.1	Parameters of the fiducial models	40
2.2	The percentage of galaxies of each morphological type at $z = 0$ in the two fiducial models	46
3.1	The parameters of our two reference models	61
3.2	The mean number of galaxy pairs per halo, N_P , calculated for three differ- ent distributions of halo occupancy, all with the same mean	77
3.3	Variant models in the τ CDM cosmology	81
3.4	Variant models in the Λ CDM cosmology	82
5.1	The parameters of our standard model	128
5.2	Results of the variant models at $z = 5$	158
6.1	Observed properties of the targeted galaxies	177
6.2	Measured rotation velocity and predicted gas temperature for the galaxies in our sample	180
6.3	X-ray fluxes and bolometric luminosities for the halos of three spiral galaxies	184
6.4	The total counts per second in six blank fields	185
6.5	Predictions from our models, assuming the standard cosmology and halo gas with $Z = 0.3Z_\odot$	192

Declaration

The work described in this thesis was undertaken between 1997 and 2000 whilst the author was a research student under the supervision of Prof. Carlos Frenk and Dr. Shaun Cole in the Department of Physics at the University of Durham. This work has not been submitted for any other degree at the University of Durham or at any other University.

Portions of this work have appeared in the following papers:

- Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, **311**, 793 (Chapter 3).
- Benson A. J., Bower R. G., Frenk C. S., White S. D. M., 2000, MNRAS, **314**, 557 (Chapter 6).

The copyright of this thesis rests with the author. No quotation from it should be published without his prior written consent and information derived from it should be acknowledged.

This thesis could not have been completed without the guidance and expertise of my supervisors Carlos Frenk and Shaun Cole. Their enthusiasm for the subject matter is infectious, and makes astrophysics seem like a worthwhile pursuit even on those days when everything falls to pieces. I'd also like to thank them, along with Carlton Baugh and Cedric Lacey, for developing the model of galaxy formation that I make extensive use of in this thesis.

In no particular order I'd like to acknowledge the help I've received from all the others that I've worked with over the past three years: Richard Bower, Simon White, Istvan Szapudi, Frazer Pearce, Adrian Jenkins, Ray Sharples, Adi Nusser and Naoshi Sugiyama.

Finally, of course, there's all those Young (some of them) Trendy (few of them) Astronomers (most of them) who've made being at Durham more fun than it would have been if they hadn't been here. The Room 6 boys, past and present, for their unique attributes: Ale Terlevich (for his ability to fix my computer even when in Birmingham), Steve Hatton (just for being unique), Scott Kay (for his semi-paralytic theories of galaxy formation), Mikey Beasley (for putting up with it all), Peder Norberg (for approximately 12 kilos — that's 26.4 pounds Peder — of Swiss cheese), Graham Smith (for telling people off on the phone in an entertaining manner) and Colin Wilson (for doing something other than extra-galactic astronomy). And all the others: Russel Smith, Simon Shaw, Katherine Gunn, Eric Bell, James Turner, Dave Gilbank, Kevin Pimblett, Harold Dyson, Mike Balogh, Chris Power, Amaya Gaztelu and Richard McDermid.

Special thanks to my mother, brother and sisters for always being there.

Finally, despite that unfortunate incident on the train after interview, things worked out with Fiona in the end. So "Thank you" Fiona — you make the finest soups known to humanity.

This thesis is dedicated to Oadby railway bridge. Nice try, but you lost.

"That is the truth of our world, Max. It can't be easily summed up with math..."

– Sol in the film " π "

Chapter 1

Introduction

We* briefly review the development of extragalactic astronomy and the associated history of galaxy formation modeling.

1.1 Historical Background

1.1.1 Early Ideas

The first detailed observations of galaxies external to our own (neglecting the Magellanic Clouds, which have undoubtedly been viewed by human eyes for many millenia) were made by the Earl of Rosse, using a 72 inch telescope which he had constructed in Birr, Ireland. Lord Rosse made sketches showing the spiral structure of galaxies such as M51. Galaxies, or simply spiral nebulae as they were then known, have been the subject of much debate ever since.

Their true nature remained uncertain for a long time. For example, at the start of the 20th century many believed that they were proto-planetary systems, with Sir Robert Ball of Cambridge University writing in his book “In The High Heavens” (Ball 1901)

Probably this nebula [M51] will in some remote age gradually condense down into more solid substances. It contains, no doubt, enough material to make many globes as big as our Earth.

In fact, both Thomas Wright and Immanuel Kant had suggested that the spiral nebulae were actually “island-universes”, i.e. stellar systems comparable to our own Milky Way, rather than small gaseous nebula within the Milky Way. The debate as to whether spiral nebulae were Galactic or extra-Galactic was famously taken up by Shapley (who favoured a Galactic explanation) and Curtis (who took the opposite view) (Shapley & Curtis 1921). The debate ran for nearly five years in the early 1920’s, and wasn’t fully resolved until 1924 when Edwin Hubble measured the distance to the Andromeda galaxy (M31) by observing a Cepheid variable star in the outskirts of that galaxy (Hubble 1929).

*Throughout this Thesis I will use the conventional ‘we’ to refer to the first person.



The distance to M31, which is now known to be around 1 Mpc, proved conclusively that it lay outside of the Milky Way, which was already known to be much smaller than this (Shapley 1919), although its true size is still somewhat uncertain (Kerr & Lynden-Bell 1986).

Hubble's work was important for a second reason. Having measured the distances to nearby galaxies, he found that their velocity of recession was well correlated with their distance. More distant galaxies were seen to be moving away from us more rapidly (Hubble & Humason 1931). This was the first observational evidence for the expansion of the Universe.

1.1.2 The Origin of Galaxies

For many years after its formulation, Einstein's General Theory of Relativity as applied to cosmology was only used to consider the idealized case of an isotropic and homogeneous Universe (an assumption which seems reasonable at early time or on large scales as discussed below). However, the Universe is certainly not homogeneous on small scales as we observe galaxies which are strongly clustered. Some means of generating such inhomogeneities was therefore needed. It had been known for some time that small density perturbations in an otherwise isotropic and homogeneous universe would begin to grow through the action of gravitational instability. The growth of perturbations in a cosmological background was first determined by Lifshitz (1946), and extensively developed by Peebles (1967). The growing perturbations would eventually decouple from the expanding universe and collapse. After a process of violent relaxation the perturbation would become a virialized, self-gravitating object. The number of such objects expected for a very general class of initial perturbations was first estimated by Press & Schechter (1974).

The origin of the required density fluctuations remained uncertain until, in the early 1980's, Alan Guth proposed a period of rapid inflationary expansion in the very early Universe in order to solve the horizon and flatness problems (Guth 1981).

The flatness problem relates to the fact that in the evolution of a standard big-bang universe (dominated either by matter or radiation) the mean density of the Universe is always driven away from the critical density required for collapse (see Appendix A.1.1). Observations indicate that the present day density is within an order of magnitude or so of the critical density, which therefore implies that just after the Big Bang the Universe must have been extremely close to critical density, leading to a problem of fine tuning in the initial conditions.

The horizon problem relates to the observed fact that the temperature of the microwave background radiation is extremely isotropic (fractional fluctuations being of the order 10^{-5} , Bennett et al. 1996). Unless we postulate that the Universe was set up in such a perfectly isotropic manner this implies that, at some early time, the Universe was able to reach thermal equilibrium. In the standard Big Bang theory this is not possible as the finite speed of light means that the distance over which thermal equilibrium could have been reached at the epoch of decoupling (when the cosmic microwave background last interacted with matter) is much smaller than the present horizon size. Thus in the standard Big Bang theory there is no way to explain why the Universe is so homogeneous on large scales (apart from the unpleasant assumption that it was simply made that way).

Inflation solves these problems by postulating a period of rapid, exponential growth to occur shortly after the Big Bang and which was driven by a scalar field. This rapid expansion naturally drives the Universe extremely close to critical density and furthermore will expand a small region of the Universe within which thermal equilibrium has been obtained to a size much greater than the present horizon. A side effect of this process would be the creation of tiny density perturbations in the material of the Universe arising from quantum fluctuations in the scalar field driving inflation. These small-scale fluctuations would then be expanded to cosmological sizes by the process of inflation.

One of the most important discoveries of recent times lent convincing evidence to the idea that galaxies grew from such primordial fluctuations. This discovery was made by the COBE satellite which detected anisotropies at the level of 10^{-5} in the temperature of the cosmic microwave background (CMB) (Bennett et al. 1996). These anisotropies had been predicted to occur in any cosmological world-model in which density inhomogeneities existed. Their direct observation, and the measurement of their amplitude, is both an important confirmation and a strong constraint on models of structure formation through the gravitational instability of primordial fluctuations. How these perturbations will grow depends upon the material content of the Universe. It seems that not all of this matter is easily detected.

Observations of clusters of galaxies seem to indicate that these were relaxed, virialized systems, and so an application of the virial theorem ought to allow an estimate of their mass to be made. Such estimates were originally carried out by Zwicky (1933) and much later by Oemler (1973), Rood (1974) and Bahcall (1974), who found that the mass of the cluster was typically ten times greater than the mass in the luminous galaxies (see the review by Bahcall 1977). Later X-ray observations of clusters revealed the presence of

hot, diffuse gas in clusters, with a mass comparable to that of the luminous galaxies, but nevertheless it still appears that 80% of the mass of clusters is in some non-luminous form. Further evidence for the existence of some type of “dark matter” came from observations of spiral galaxy rotation curves, which indicate that the rotation speed of such galaxies is almost constant with radius over a range of scales where the enclosed mass of the luminous galaxy is increasing only very slowly (see, for example, Faber & Gallagher 1979). Such a result suggests the existence of a significant quantity of non-luminous matter within these galaxies.

So what is the nature of this dark matter? Perhaps the simplest possibility is that it is normal, baryonic matter which for some reason does not emit any light. For example, the dark matter could consist of brown dwarfs, planets or even black holes. However, estimates of the abundance of the light elements from models of primordial nucleosynthesis indicate that only a small fraction of the required dark mass could be baryonic (Walker et al. 1991). This result does depend upon the value of Hubble’s constant. For example, if $H_0 = 25$ km/s/Mpc and present evidence for a low density Universe (e.g. Perlmutter et al. 1999) is correct the entire mass of the Universe could be baryonic. Note however the present estimates of H_0 favour a value closer to 70 km/s/Mpc (Mould et al. 2000). Barring the possibility of a low H_0 , the dark matter is most likely to be non-baryonic material.

Suggestions for suitable dark matter candidates have come from particle physics, and can be conveniently split into two categories: hot dark matter and cold dark matter. Hot dark matter, such as a massive (around 30 eV) neutrino, moves at relativistic velocities at the epoch of matter-radiation equality. Cold dark matter, on the other hand, consists of much more massive particles which therefore becomes sub-relativistic much earlier in the Universe’s life.

These two flavours of dark matter behave very differently. As hot dark matter moves relativistically at high redshift it is able to free-stream out of a growing density perturbation into surrounding underdense regions. As such it tends to smooth away any density perturbation smaller than the free-streaming scale, which is typically around $10^{16} M_\odot$ (Bond, Efstathiou & Silk 1980). In numerical simulations of universes dominated by hot dark matter, White, Frenk & Davis (1983) demonstrated that the first objects to form were pancake-like structures (i.e. they collapsed along one axis first) with masses comparable to the free-streaming scale. In order for these very massive objects to form galaxies they would have to fragment in some way. White, Frenk & Davis (1983) showed that the characteristic clustering scale of hot dark matter is much greater than that observed for

galaxies and so concluded that hot dark matter seemed incompatible with observations under reasonable assumptions about the background cosmology (e.g. that the Universe was not at a super-critical density).

Cold dark matter (CDM) behaves very differently. As it moves at sub-relativistic speeds small density perturbations are not smoothed away, and in fact are the first perturbations to collapse (since the fluctuations laid down by inflation and modified by the subsequent evolution of the universe have greatest amplitude on small scales). Larger objects are then built up by the merging of smaller objects. N-body simulations of CDM dominated universes carried out by Davis et al. (1985) demonstrated that dark matter halos with a wide range of masses are present at any one time in such universes. They found also that the clustering of CDM in a low-density universe had about the same amplitude as that measured for galaxies, although it did have a very different shape. Furthermore the dispersion in dark matter particle velocities was found to be much higher than that inferred from observations of galaxies. Davis et al. (1985) concluded that none of these CDM models was fully consistent with the observations, but noted that galaxies may not represent a random sampling of the dark matter distribution, i.e. they may be “biased”, and made a simple model of such biased galaxy formation by identifying galaxies with the highest peaks of the initial density field. The resulting galaxy catalogues were found to show strong differences from the dark matter, for example in a critical density model the ‘galaxies’ were more clustered than the dark matter, and in better agreement with the observational data. Realising the importance of such biases, Davis et al. (1985) stated:

Clearly, what is now required is a proper physical model for galaxy formation which can be grafted onto simulations to see if the distribution of our “galaxies” is indeed realistic.

Although the dark matter seems to be the gravitationally dominant component of our Universe it is not visible since, by definition, it emits no light. Dark matter may, however, be detectable in other ways: for example by direct detection of WIMPs (Gaitskell 1998) or by microlensing if the dark matter consists of primordial black holes (Hawkins 1998). The luminous galaxies that we can actually see are made of the small quantity of baryonic material present in the Universe. This gas has an important property, the ability to cool radiatively, that makes its dynamical behaviour very different from that of the dark matter. Once a virialized object has formed and the gas within it has been heated, through the actions of shocks during violent relaxation (in which the energy and angular momenta

of the baryons are redistributed by strong fluctuations in the gravitational potential, Lynden-Bell 1967), the gas will begin to cool, through Bremsstrahlung, recombination radiation and other atomic processes.

The first to examine in detail the significance of this cooling process for galaxy formation were White & Rees (1978) who extended earlier work by Binney (1977), Rees (1977) and Silk (1977) and noted that without significant dissipation the sizes of galaxies and their survival through the continuous merging processes in a hierarchical universe seemed inexplicable. Dissipation allowed the galaxies to become much more concentrated, and sufficiently bound to survive for long periods of time. Since this cooling gas would have been imparted with some angular momentum through the actions of tidal torques, as first described by Hoyle (1949) and developed by Peebles (1969), Doroskevich (1970) and White (1984), it may be expected to form into a centrifugally supported disk, perhaps similar to a spiral galaxy.

By calculating the rate at which a gaseous plasma would cool in a dark matter halo White & Rees (1978) were able to predict the characteristic luminosity of bright galaxies, and the shape of the observed luminosity function (although with a somewhat steeper faint end slope than had been observed, indicating the need for processes such as supernovae feedback and galaxy merging).

1.1.3 Further Observational Developments

At the same time as significant progress was being made in theoretical studies of the Universe new observational instruments and techniques provided ever more detailed and complete information on the galaxy population, both at low and high redshifts.

Much of this observational work was originally aimed at establishing the cosmological world-model (i.e. the parameters which describe the evolution of the Universe). Although early workers in this field recognised the possibility of evolution of the galaxy population, which would make a determination of cosmological parameters much more difficult, even in the 1970's much work was still being carried out to measure the cosmic deceleration from observations of galaxies. The evolutionary changes in galaxies were viewed as a small correction that had to be made before cosmological parameters could be recovered (the term 'evolutionary correction' is still widely used today). With the arrival of quantitative predictions for stellar evolution (Sandage 1961, Tinsley 1972) (and also the growing understanding of how physical structures form in an expanding universe) the evolution of galaxies became the main goal of this type of observational study.

Throughout the 1980's and early 1990's observations focussed on measuring the number of galaxies brighter than a given magnitude, the correlations between the angular positions of galaxies, the distribution of galaxy colours, and, with spectroscopic surveys, the distribution of galaxy redshifts. Koo & Kron (1992) give a comprehensive review of work carried out in this period.

In the last decade a wealth of observational data has become available, much of which probes important epochs in the evolution of the Universe and its constituent galaxies. A large fraction of this data has been obtained by the Hubble Space Telescope and ground based 8 metre and larger telescopes which have looked deeper into the Universe than has previously been possible. One of the most remarkable results from the ground based telescopes has been the detection of galaxies at $z \sim 3$ by use of the Lyman-break technique (Steidel et al. 1996). At such high redshifts the Universe is only around one eighth of its present age, being between one and two billion years old. That bright galaxies have already formed in this short (by cosmological standards) time is a remarkable fact, and a very strong test for models of galaxy formation.

The early formation of such galaxies has also been measured in another way, through determinations of the universal rate of star formation as a function of redshift. Estimates of this quantity at low redshift have now been supplemented by determinations of the star formation rate at higher redshifts (Madau et al. 1996). Although these measurements are still somewhat uncertain (due mostly to the uncertainty in the degree of extinction of starlight by dust in distant galaxies) it is clear that the process of galaxy (and star) formation began very early in the history of the Universe.

1.2 The State of the Art

The creation of a realistic theory of galaxy formation proceeded through several intermediate steps, which have culminated in models of the type we will describe later in this Thesis (Chapter 2).

Following on from the work of White & Rees (1978), Cole (1991) implemented calculations of gas cooling within an evolving hierarchy of dark matter halos, using a Monte-Carlo approach based upon the recently developed extended Press-Schechter theory (Bond et al. 1991, Bower 1991), and demonstrated that such a model could produce a galaxy luminosity function which bore a good resemblance to the observed function, if, and only if, the effects of feedback from supernovae were included in the model. The Monte-Carlo

approach employed by Cole (1991) enabled a more detailed study of the effects of galaxy merging than had previously been possible within earlier models (e.g. White 1989).

White & Frenk (1991) considered a similar model to Cole (1991) but also included simple models of chemical enrichment and used synthetic stellar spectra to turn the star formation histories of their galaxies into luminosities. White & Frenk (1991) found similar results to Cole (1991) and noted that without merging of galaxies these models produced too many faint galaxies, resulting in a luminosity function which was much steeper at its faint end compared to that observed, as had been seen by White & Rees (1978).

At around the same time a model of galaxy formation was developed by Lacey & Silk (1991), which was similar in spirit, but different in detail, from that of White & Frenk (1991). Lacey & Silk (1991), and also Lacey et al. (1993), used the statistics of peaks in a Gaussian random field to model the formation and merging of halos and assumed that star formation in galaxies was triggered by tidal disturbances from nearby systems. The results from this study were qualitatively similar to those of White & Frenk (1991).

More elaborate versions of these models were developed in the following few years, resulting in the models of Kauffmann, White & Guiderdoni (1993), Cole et al. (1994) and Baugh, Cole & Frenk (1996a), which included in their calculations full stellar population synthesis models, estimates of the frequency of galaxy-galaxy mergers and, for the first time, a model for the formation of elliptical galaxies and spiral bulges (previous models having only considered galaxy disks). Later additions to these models have included more elaborate models of chemical enrichment (Kauffmann 1996a, Cole et al. 2000), determinations of the sizes of galaxies (Cole et al. 2000), detailed models of the extinction of starlight by dust (Kauffmann et al. 1999a, Cole et al. 2000), estimates of the frequency of satellite-satellite mergers (Somerville & Primack 1999) and calculations of the growth of black holes in the centres of galaxies and the associated quasar activity (Kauffmann & Haehnelt 2000).

An alternative approach to the problem of galaxy formation also made large advances around this same time. The addition of a baryonic component (and the relevant physics of hydrodynamics and dissipation) into N-body simulations allowed the formation of galaxies to be studied with these techniques (e.g. Katz, Hernquist & Weinberg 1992, Evrard, Summers & Davis 1994, Frenk et al. 1996, Katz, Weinberg & Hernquist 1996, Navarro, Frenk & White 1997, Pearce et al. 1999, Thacker et al. 1999, Blanton et al. 2000). Whilst this approach allows the physics of gas dynamics and cooling to be followed accurately, without the need for simplifying assumptions, the large computational demand made by

such simulations limits their effectiveness, making it very difficult to include physical processes, such as star formation and the associated supernovae feedback, which are known to be of importance for galaxy formation (Navarro & White 1993).

1.3 Motivation for this Work

A number of questions concerning the formation and evolution of galaxies remain unanswered. The physical model of galaxy formation described in Chapter 2, coupled with improved observations, can now begin to investigate these questions directly. In this Thesis we will study in detail the origins of the spatial and velocity distribution of galaxies by taking up the suggestion of Davis et al. (1985) and incorporating the galaxy formation model into N-body simulations of dark matter.

Furthermore, since any model must be based on certain fundamental assumptions we will investigate the validity of certain assumptions underlying the galaxy formation model that we employ.

Finally, we will explore one aspect of the connection between the galaxy population and the physics of the early Universe by studying the reionization of the intergalactic medium (IGM).

Here we aim to address the following specific problems:

- Do galaxy disks really grow by accretion of gas from a surrounding halo?
- How well do semi-analytic and numerical simulation models of galaxy formation agree and where might they be improved?
- How does the spatial distribution of galaxies differ from that of the dark matter? How does this depend upon the internal properties of galaxies?
- Can stellar sources alone account of the reionization of the Universe, and will the signature of this process be detectable in the cosmic microwave background?

First of all we turn our attention to describing the details and practical implementation of such a model.

Chapter 2

A Semi-Analytic Model of Galaxy Formation

Much of the work presented in this Thesis makes extensive use of the semi-analytic model of galaxy formation developed by Cole et al. (2000). This model combines several algorithms and procedures to allow the behaviour of dark matter, gas and stars to be followed. In this Chapter we will describe in detail the methods of this model. A large proportion of the work in this Chapter is based on Cole et al. (2000), who also describe in detail how the model parameters are constrained and demonstrate how the model predictions respond to changes in those parameters.

Some of the theoretical techniques upon which the model is built (specifically the linear and non-linear growth of perturbations in an expanding universe, together with derivations of various equations) are consigned to Appendix A.

2.1 Dark Matter Halos

In CDM cosmologies where the baryonic component of the Universe accounts for only a small fraction of the total mass, the dynamics of the dark matter play a crucial role in the process of galaxy formation. For this reason, the first step in the semi-analytic model is to determine various properties of dark matter halos in the universe: their formation histories, angular momenta and density profiles.

2.1.1 Merger Trees

In CDM cosmologies, the power spectrum of density fluctuations has most variance on small scales. For this reason small dark matter halos form early (as fluctuations on the smallest scales become non-linear earlier). Larger halos form later and are created by the merging of smaller halos. This merging process can be described by the extended Press-Schechter theory. Here we derive the dark matter halo mass function and the distribution of progenitor halo masses and describe how these are used to construct dark matter merger trees in the semi-analytic model.

Dark matter halo mass function

The excursion set derivation of the Press & Schechter (1974) mass function proceeds as follows (Bond et al. 1991, Lacey & Cole 1993, see Bower 1991 for an alternative derivation). For a Gaussian random field of density fluctuations, we define the variance on mass scale M to be

$$S(M) = \sigma^2(M) = \langle |\delta(M, \mathbf{x})|^2 \rangle, \quad (2.1)$$

where $\delta(M, \mathbf{x}) = (\rho(\mathbf{x}) - \bar{\rho}) / \bar{\rho}$ is the overdensity at position \mathbf{x} when the density field is smoothed on mass scale M (i.e. with a filter enclosing, on average, a mass M). Here $\rho(\mathbf{x})$ and $\bar{\rho}$ are the density at position \mathbf{x} and the mean density of the universe respectively. The average is taken over all space. The smoothed field can be written as a convolution of the unsmoothed field with a window function $W_M(r)$

$$\delta(M, \mathbf{x}) = \int \delta(\mathbf{x}) W_M(\mathbf{x} - \mathbf{y}) d^3\mathbf{y}. \quad (2.2)$$

Taking the Fourier transform of this we find

$$\delta_{\mathbf{k}}(M) = \delta_{\mathbf{k}} \hat{W}_M(k), \quad (2.3)$$

where $\hat{W}_M(k)$ is the Fourier transform of $W_M(r)$, $\delta_{\mathbf{k}}$ are the coefficients of the Fourier transform of $\delta(\mathbf{x})$, $\delta_{\mathbf{k}}(M)$ are the coefficients of the Fourier transform of $\delta(M, \mathbf{x})$ and $k = |\mathbf{k}|$. Hence the mass variance can be written (using Parseval's theorem) as

$$S(M) = \int \langle |\delta_{\mathbf{k}}|^2 \rangle \hat{W}_M^2(k) d^3\mathbf{k}. \quad (2.4)$$

To derive the Press-Schechter mass function use is made of sharp k -space filtering, for which

$$\hat{W}_M(k) = \begin{cases} 1 & \text{if } k < k_s(M) \\ 0 & \text{if } k > k_s(M), \end{cases} \quad (2.5)$$

where $k_s(M)$ is chosen such that $S(M)$ has the form that would be obtained if a window function that was a top-hat filter in real-space were used instead. This choice for $k_s(M)$ ensures that the definition of $S(M)$ is compatible with the spherical top-hat model of Appendix A.1.3. We now imagine placing a sphere enclosing mean mass M at some point \mathbf{x} . For physically interesting power spectra, as $M \rightarrow \infty$ $S(M) \rightarrow 0$ and so $\delta(M, \mathbf{x})$ will be zero. As M is decreased from infinity $\delta(M, \mathbf{x})$ will begin to wander away from zero. The trajectory followed by $\delta(M, \mathbf{x})$ as M is decreased is completely determined by the $\delta_{\mathbf{k}}$ values. Furthermore, as we are using a sharp k -space filter, and since the various Fourier modes are uncorrelated (as we have assumed a Gaussian random field) as we make a

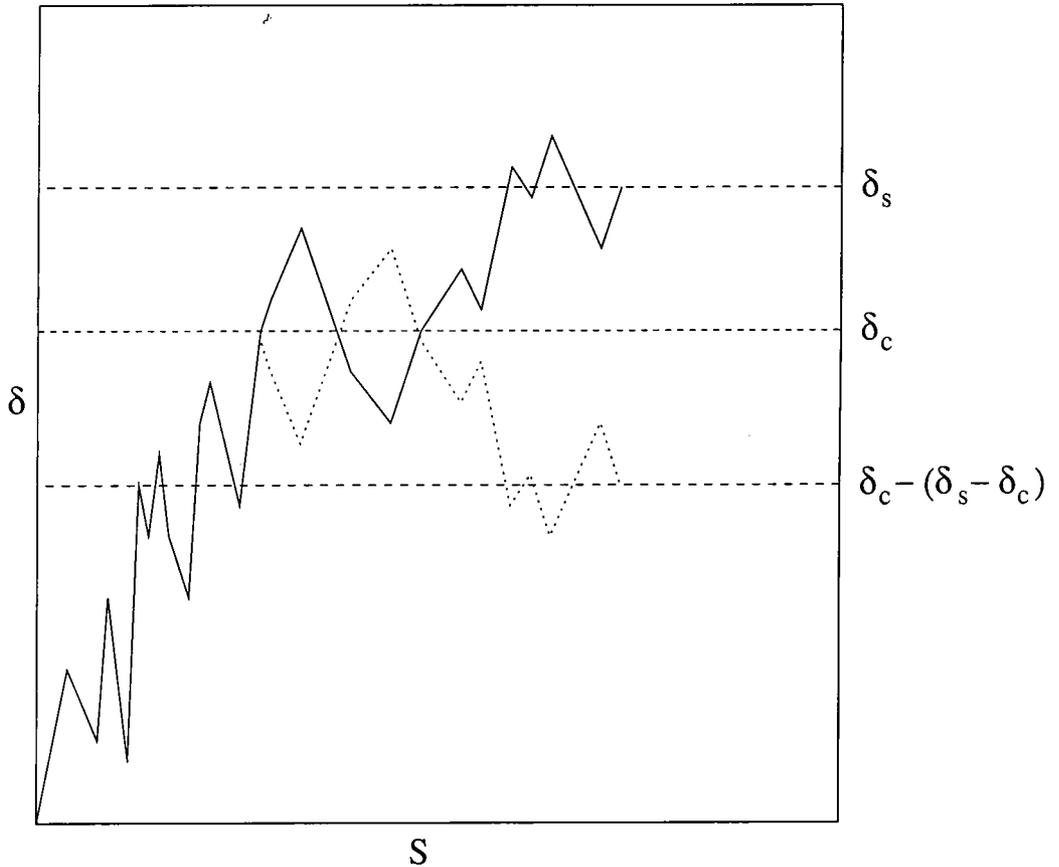


Figure 2.1: An example of a trajectory followed by $\delta(M, \mathbf{x})$ is shown as the solid line. At the end of the line the trajectory has reached a value of $\delta_s > \delta_c$. The dotted line shows an equally probable trajectory, which follows the solid line until δ_c is reached, and is then a mirror image of the solid line about $\delta = \delta_c$. The mirror image trajectory reaches a final value of $\delta_c - (\delta_s - \delta_c)$.

small decrease in M the increment in $\delta(M, \mathbf{x})$ is entirely uncorrelated with the previous trajectory of $\delta(M, \mathbf{x})$. Hence $\delta(M, \mathbf{x})$ executes a random walk.

For a Gaussian random field the distribution of overdensities is simply a Gaussian of width $\sqrt{S(M)}$, i.e.

$$P(\delta; M)d\delta = \frac{1}{\sqrt{2\pi S(M)}} \exp\left(-\frac{\delta^2}{2S(M)}\right) d\delta, \quad (2.6)$$

where $P(\delta; M)d\delta$ is the probability of a point in the universe having overdensity in the range δ to $\delta + d\delta$. Although this probability corresponds to averaging over all spatial positions, by the ergodic theorem this is equivalent to considering one spatial position and averaging over all possible realisations of the density field. Therefore, the fraction of possible trajectories which have $\delta(M) > \delta_c$ (where δ_c is some, as yet undefined, value) is

simply

$$F(S(M), \delta_c) = \frac{1}{\sqrt{2\pi S(M)}} \int_{\delta_c}^{\infty} \exp\left(-\frac{\delta^2}{2S(M)}\right) d\delta. \quad (2.7)$$

This is the fraction of trajectories above δ_c at a smoothing scale corresponding to $S(M)$. Later we will equate trajectories that have risen above δ_c with collapsed halos. A point will be part of a collapsed halo of mass greater than M if its trajectory passed above δ_c for any $S < S(M)$. The above expression accounts for all trajectories above δ_c at $S(M)$, but we also want to include trajectories which are now below δ_c but have been above it for some smaller $S(M)$. As first pointed out by Chandrasekhar (1943a), for these purely random walks for every trajectory leading to $\delta > \delta_c$ there is an equally probable walk leading to $\delta_c - (\delta - \delta_c) = 2\delta_c - \delta$ (see Fig. 2.1). Hence we can include these trajectories to give the corrected expression

$$F(S(M), \delta_c) = \frac{1}{\sqrt{2\pi S(M)}} \int_{\delta_c}^{\infty} \exp\left(-\frac{\delta^2}{2S(M)}\right) + \exp\left(-\frac{(2\delta_c - \delta)^2}{2S(M)}\right) d\delta. \quad (2.8)$$

This is the fraction of trajectories which have risen above δ_c below some value of $S(M)$. The fraction of objects making their first upcrossing through δ_c in the interval S to $S+dS$ (we will write S rather than $S(M)$ from now on where the meaning is unambiguous) must therefore be $f(S, \delta_c)dS = [dF(S, \delta_c)/dS]dS$, i.e.

$$f(S, \delta_c) = \frac{1}{2\sqrt{\pi S}} \left[\int_{\delta_c/\sqrt{2S}}^{\infty} (2x^2 - 1) \exp(-x^2) dx - \int_{-\infty}^{\delta_c/\sqrt{2S}} (2y^2 - 1) \exp(-y^2) dy \right], \quad (2.9)$$

where $x = \delta/\sqrt{2S}$ and $y = (2\delta_c - \delta)/\sqrt{2S}$. The integrals are easily solved using integration by parts, to give

$$f(S, \delta_c) = \frac{\delta_c}{\sqrt{2\pi S^{3/2}}} \exp\left(-\frac{\delta_c^2}{2S}\right). \quad (2.10)$$

If we now specify that δ_c is the extrapolated linear theory overdensity required for collapse in the top-hat model, as defined in eqn. (A.51), the above expression gives the fraction of mass in the universe which has collapsed into objects of mass M to $M+dM$ (corresponding to S to $S+dS$). The number of objects of this mass per comoving unit volume is then simply

$$\begin{aligned} \frac{dn(M, t)}{dM} dM &= \frac{\bar{\rho}_0}{M} f(S, \delta_c) \left| \frac{dS}{dM} \right| dM, \\ \frac{dn(M, t)}{dM} dM &= \left(\frac{2}{\pi}\right)^{1/2} \frac{\bar{\rho}_0}{M^2} \frac{\delta_c(t)}{\sigma(M)} \left| \frac{d \ln \sigma}{d \ln M} \right| \exp\left[-\frac{\delta_c^2(t)}{2\sigma^2(M)}\right] dM. \end{aligned} \quad (2.11)$$

Note that since δ_c and $\sigma(M)$ both scale with time in the same way (i.e. they increase in proportion to the linear growth factor) and since only their ratio enters into the above

formula we can choose to evaluate them at any epoch. Conventionally they are both evaluated at the present day ($t = t_0$).

When Press & Schechter (1974) originally derived this mass function they found that it accounted for only half of the total mass in the universe, and so simply scaled the mass function up by a factor of two (this extra mass supposedly being accreted onto halos from voids). The inclusion of trajectories that were once above δ_c but are now below it at $S(M)$ solves this problem. This derivation of this mass function makes several assumptions, such as applying the spherical top-hat model for halo collapse and filtering the density field with a sharp k -space filter (which is extended in real space). Despite these simplifications the resulting mass function agrees reasonably well with those recovered from N-body simulations of dark matter, although the discrepancies between the two are now well known (Somerville et al. 1998). These discrepancies can be removed by considering models of non-spherical collapse. In particular, Sheth, Mo & Tormen (1999) have calculated the halo mass function using an elliptical top-hat collapse model which gives a much better fit to N-body simulation results.

Progenitor halo mass distributions

It is now trivial to derive the distribution of masses for the progenitors of a halo of given mass. Firstly, we define $\delta_c(z)$ to be the extrapolated linear overdensity at the present day required for an object to collapse at redshift z . In an Einstein-de Sitter universe linear perturbations grow as $\delta \propto (1+z)^{-1}$, and so $\delta_c(z) = (1+z)\delta_c(0) = 1.686(1+z)$ as shown in Appendix A.1.2.

Suppose that a trajectory first crosses through $\delta_{c,2} = \delta_c(z[t_2])$, where $z[t]$ is the redshift corresponding to cosmic time t , at $S_2 = S(M_2)$ (i.e. it forms an object of mass M_2 at time t_2). We now continue this trajectory to larger values of S and look for where it first crosses $\delta_{c,1} = \delta_c(z[t_1]) > \delta_{c,2}$, let this occur at $S_1 = S(M_1)$. Now, since the trajectories exhibit a random walk (i.e. the continuation of the trajectory to higher S is independent of the starting position $[S_2, \delta_{c,2}]$) the fraction of trajectories beginning at $[S_2, \delta_{c,2}]$ and having their first upcrossing through $\delta_{c,1}$ at S_1 is given by eqn. (2.10) with the simple replacements $S \rightarrow S_1 - S_2$ and $\delta_c \rightarrow \delta_{c,1} - \delta_{c,2}$, i.e.

$$f(S_1, S_2)dS_1 = \frac{\delta_{c,1} - \delta_{c,2}}{\sqrt{2\pi}(S_1 - S_2)^{3/2}} \exp\left(-\frac{(\delta_{c,1} - \delta_{c,2})^2}{2(S_1 - S_2)}\right) dS_1, \quad (2.12)$$

or equivalently

$$f(M_1, M_2)dM_1 = \frac{1}{\sqrt{2\pi}} \frac{(\delta_{c1} - \delta_{c2})}{(\sigma_1^2 - \sigma_2^2)^{3/2}} \times \exp\left(-\frac{(\delta_{c1} - \delta_{c2})^2}{2(\sigma_1^2 - \sigma_2^2)}\right) \frac{d\sigma_1^2}{dM_1}dM_1, \quad (2.13)$$

where $\sigma_1 = \sigma(M_1)$ and $\sigma_2 = \sigma(M_2)$. This is the distribution of progenitor halos for the halo of mass M_2 at time t_2 .

The limiting behaviour of eqn. (2.13) as $t_1 \rightarrow t_2 - dt_1$ gives the progenitor distribution at a slightly earlier time $t_2 - dt_1$ (Lacey & Cole 1993)

$$\left. \frac{df_{12}}{dt_1} \right|_{t_1=t_2} dM_1 dt_1 = \frac{1}{\sqrt{2\pi}} \frac{1}{(\sigma_1^2 - \sigma_2^2)^{3/2}} \frac{d\delta_{c1}}{dt_1} \frac{d\sigma_1^2}{dM_1} dM_1 dt_1. \quad (2.14)$$

The number of progenitors of mass M_1 to $M_1 + dM_1$ at this slightly earlier time, dN/dM_1 , is found by multiplying the above by M_2/M_1

$$\frac{dN}{dM_1} \approx \frac{df_{12}}{dt_1} \frac{M_2}{M_1} \delta t_1, \quad (2.15)$$

for finite δt_1 .

Merger trees

A Monte-Carlo scheme for generating realisations of halo merger histories can now be developed. Since the number of progenitors diverges as $M_1 \rightarrow 0$ (although the total mass in such objects does not diverge) it is necessary to impose some minimum progenitor mass, M_{res} , for the practical purpose of constructing merger trees. Then the total number of progenitor halos more massive than M_{res} and less massive than $M_2/2$ at the earlier time is simply,

$$P = \int_{M_{\text{res}}}^{M_2/2} \frac{dN}{dM_1} dM_1, \quad (2.16)$$

whilst the fraction of the original halo mass in progenitors below M_{res} is

$$F = \int_0^{M_{\text{res}}} \frac{dN}{dM_1} \frac{M_1}{M_2} dM_1. \quad (2.17)$$

If we choose δt_1 to be small enough such that $P \ll 1$ then only binary splits will occur (since P is the number of halos of mass less than $M_2/2$, and to conserve mass if one progenitor of mass M_1 forms there must be another of mass $M_2 - M_1$). To determine the progenitors of the halo a random number, R , is picked in the range 0 to 1. If $R > P$ then no binary split occurred. In this case the mass of the halo is simply reduced to $M_2(1 - F)$ to reflect the loss of mass to progenitors below the resolution limit. If $R < P$ then a

mass between M_{res} and $M_2/2$ is drawn from the distribution given by eqn. (2.15) and two progenitors of masses M_1 and $M_2(1 - F) - M_1$ are created. Note that to determine the progenitor distribution three inputs are required: $\sigma(M)$ and the cosmological parameters Ω_0 and Λ_0 which determine $\delta_c(t)$.

By applying this process to each progenitor halo present at a succession of timesteps a complete merger history of a dark matter halo can be created. The practical implementation of this algorithm in the semi-analytic model computes the merger history on a grid of very fine timesteps. This merger tree is then placed onto a coarser grid of N_{step} timesteps (for the purposes of the remaining galaxy formation calculations). The timesteps of this grid (which may be spaced in time in a variety of ways depending on the problem being attempted) are chosen to be much smaller than the typical dynamical times of halos. The accuracy of various algorithms for merger tree construction has been explored by Somerville et al. (1998), who find that the mean properties of progenitors are reproduced reasonably well (although the higher order moments of the progenitor distribution are often not). The accuracy of this algorithm is explored briefly in §7.2.2.

Beginning at the earliest time considered, each halo is assigned a mean density as determined from its time of collapse and eqn. (A.55) (or the equivalent for non-Einstein-de Sitter cosmologies), a density profile and an angular momentum (as described in §2.1.2 and §2.1.3). Each halo is assumed to retain these properties throughout its lifetime. A halo survives until it becomes part of a halo with mass greater than f_{form} times its original mass ($f_{\text{form}} = 2$ is usually assumed). The properties of this new halo are then computed afresh. The time at which a new halo is assumed to ‘form’ is defined in a consistent manner: a new halo life begins when mergers produce a halo whose mass exceeds f_{form} times the formation mass its largest progenitor.

Initial Power Spectrum

The key input for the Press-Schechter theory is the mass variance of the primordial density field as a function of scale, $\sigma^2(M)$. This is determined from the power spectrum of the density field, $P(k)$, through the relation

$$\sigma^2(M) = \frac{1}{2\pi^2} \int_0^\infty P(k) \hat{W}_M^2(k) k^2 dk, \quad (2.18)$$

where $\hat{W}(k)$ is the window function. As noted above (§2.1.1), we filter the density field using a top-hat in real-space such that the window function becomes

$$\hat{W}_M(k) = \frac{3(\sin kR - kR \cos kR)}{(kR)^3}, \quad (2.19)$$

where R is the radius of a sphere which contains, on average, mass M . Theories of inflation predict a primordial power spectrum of the form $P(k) \propto k$, but do not predict the amplitude. This primordial spectrum is processed by various physical effects as the universe evolves. The form of the power spectrum at a particular redshift is conventionally written as (Bardeen et al. 1986, eqn. G2)

$$P(k, z) = \left[\frac{D(z)}{D(z_p)} \right]^2 T^2(k, z) P(k, z_p), \quad (2.20)$$

where $D(z)$ is the growth factor (see Appendix A.1.2), z_p is the redshift at which the primordial power spectrum was defined and $T(k, z)$ is known as the transfer function. It is the transfer function which describes the effects of physical processes on the power spectrum. The transfer function is usually calculated numerically and described by a simple fitting formula. Throughout this work we use the fitting formula of Bardeen et al. (1986) to the CDM transfer function unless otherwise stated. The CDM transfer function depends on the so-called ‘shape parameter’, Γ , defined by Efstathiou, Bond & White (1992). For a standard CDM cosmology $\Gamma \approx \Omega_0 h$. However, Γ may take a lower value if, for example, a population of massive τ neutrinos in the early universe decayed into lighter neutrinos thereby delaying the epoch of matter-radiation equivalence (see, for example, White, Gelmini & Silk 1995). This is the origin of the τ CDM cosmological models considered in this Thesis.

The only remaining unknown is the normalization of the power spectrum. This is usually specified by the quantity σ_8 , which is the linear theory mass variance in spheres of radius $8h^{-1}\text{Mpc}$ at the present day. This quantity can be constrained observationally from the abundances of hot X-ray clusters (Eke, Cole & Frenk 1996, Viana & Liddle 1996).

2.1.2 Density Profile

Once the dark matter halo has virialised we assume that its density structure can be described by a smooth, spherically symmetric profile. In reality any halo formed from the merging of smaller halos will be neither spherical nor smooth since progenitor halos can survive inside a larger halo for some time as shown by Moore et al. (1998). In high resolution N-body simulations of CDM universes Navarro, Frenk & White (1995) have shown that dark matter halos have a generic density profile within the virial radius, r_H , of the form

$$\rho(r) = \frac{\Delta_{\text{vir}} \rho_{\text{crit}}}{f(a_{\text{NFW}})} \frac{1}{r/r_H (r/r_H + a_{\text{NFW}})^2} \quad (2.21)$$

where Δ_{vir} is the mean density, in units of the critical density ρ_{crit} , inside the virial radius as defined by the spherical top hat model and, $f(a_{\text{NFW}}) = \ln(1 + 1/a_{\text{NFW}}) - 1/(1 + a_{\text{NFW}})$. This form is found to be independent of halo mass and the values of cosmological parameters. The single free parameter a_{NFW} (the inverse of the concentration parameter of Navarro et al.) is strongly correlated with the halo mass (Navarro, Frenk & White 1997) and so halos in this model are assigned a value of a_{NFW} using a procedure equivalent to that described by Navarro, Frenk & White (1997). Although this profile has been found to be an accurate representation of halo profiles found in N-body simulations (Navarro, Frenk & White 1996, Navarro, Frenk & White 1997, Eke, Navarro & Frenk 1998) very high resolution simulations of dark matter halos Moore et al. (1998, see also Kravtsov et al. 1998) find a significantly steeper slope to the density profile in the inner regions of the halo. Since this steepening of the profile occurs only within a radius of approximately 10% of the virial radius it is unlikely to cause major differences to the properties of galaxies in our models. Perhaps a more serious discrepancy is that the NFW profile seems unable to match the rotation curves of dark matter dominated dwarf galaxies (Moore et al. 1999, Navarro & Steinmetz 1999).

2.1.3 Spin Distribution

As a dark matter halo collapses it will experience tidal torques from the gravitational field caused by other structures nearby. These torques will cause the halo to spin (i.e. it will gain some angular momentum). The amount of angular momentum gained by a halo in this way is conventionally described by the dimensionless spin parameter λ_{H} , defined as

$$\lambda_{\text{H}} = \frac{J_{\text{H}}|E_{\text{H}}|^{1/2}}{GM_{\text{H}}^{5/2}}, \quad (2.22)$$

where J_{H} is the magnitude of the halo's angular momentum, E_{H} is the total energy of the halo (gravitational plus kinetic) and M_{H} is the halo's mass. The distribution of spin parameters has been measured in N-body simulations of dark matter (see, for example, Barnes & Efstathiou 1987, Efstathiou et al. 1988, Warren et al. 1992, Cole & Lacey 1996) and seems to be only very weakly dependent on halo mass and the power spectrum. Furthermore, Lemson & Kauffmann (1999) find that this distribution is independent of the environment in which the halo lives. The distribution of spin parameters is well fit by a log-normal distribution of the form (Cole & Lacey 1996)

$$P(\lambda_{\text{H}})d\lambda_{\text{H}} = \frac{1}{\sqrt{2\pi}\sigma_{\lambda}} \exp\left(-\frac{(\ln \lambda - \ln \lambda_{\text{med}})^2}{2\sigma_{\lambda}^2}\right) \frac{d\lambda_{\text{H}}}{\lambda_{\text{H}}}, \quad (2.23)$$

where the parameters σ_λ and λ_{med} have values of 0.53 and 0.039 respectively. Each newly formed halo in the semi-analytic model is assigned a spin parameter drawn at random from this distribution.

To determine the angular momentum of the gaseous material that will eventually form a galaxy it is necessary to know the mean rotational velocity of the halo, V_{rot} , as a function of radius. A simple model in which V_{rot} is constant with radius is compatible with the results of Warren et al. (1992) and Cole & Lacey (1996). To express V_{rot} in terms of known halo parameters (such as λ_{H} , and the virial radius and mass) it is necessary to evaluate the angular momentum and the total energy of the halo. A detailed calculation of V_{rot} for both isothermal and NFW halo profiles is given in Appendix A.2.

2.2 Gas Dynamics

Although the dark matter is thought to be the gravitationally dominant component of the Universe it is the baryonic gas which is of greatest importance for studies of galaxy formation (since it is this gas which forms the visible parts of galaxies). Until a growing perturbation has become strongly nonlinear, the gaseous component behaves almost identically to the dark matter, since it is too diffuse and cold for hydrodynamic or radiative processes to be of great significance, although at very low dark matter overdensities ($\lesssim 0.5$) finite gas pressure prevents the gas from become similarly rarefied (Gnedin & Hui 1998).

After virialisation, however, the gas and dark matter are expected to behave very differently. As the halo collapses the gas will shock and so be heated to approximately the virial temperature of the halo. Thereafter the hot, ionized gas can begin to cool. Above 10^4K gas begins to be collisionally ionized and so it will cool efficiently via atomic processes. As the gas loses energy in this way it is forced to sink ever lower in the gravitational potential of the halo. In this way a cooling flow (see, for example, Fabian 1994) forms, supplying gas to the central regions of the halo. In Chapter 6 we look for observational evidence that such cooling flows exist around present day galaxies.

In the semi-analytic model of Cole et al. (2000) the gas in a dark matter halo is assumed to be shock heated to the virial temperature after which it adopts an initial density profile in the spherically symmetric halo. Hydrodynamic simulations of clusters (Navarro, Frenk & White 1995, Eke, Navarro & Frenk 1998) show that, if cooling is ignored, the cluster gas adopts a profile which is more extended than that of the dark matter and which is well fit by a β -model (Cavaliere & Fusco-Femiano 1976) in which $\rho_{\text{gas}}(r) \propto (r^2 + r_{\text{core}}^2)^{-3\beta/2}$,

with $\beta \approx 2/3$ and $r_{\text{core}} \approx r_{\text{NFW}}/3 \approx r_{\text{H}}/20$. The gas temperature in these simulations is found to be close to the virial temperature, defined as $T_{\text{H}} = \frac{1}{2}\mu m_{\text{H}}V_{\text{H}}^2$ where m_{H} is the mass of the hydrogen atom and μ is the mean molecular mass of the ionized plasma, of the halo, and varies only very slowly with radius.

In the semi-analytic model gas is therefore initially assumed to adopt a β -model density profile with a constant temperature equal to the virial temperature. A value of $\beta = 2/3$ is chosen. The initial value of r_{core} is taken to be $r_{\text{core}} = r_{\text{NFW}}/3$. Cooling in progenitor halos may affect this value of r_{core} since this cooling will have removed the lowest entropy gas from the halo. As it is this low entropy gas which would normally sink to the central regions of a new halo it seems likely that the cores of halos in which significant cooling occurred in progenitors will be relatively larger. For this reason a simple model is adopted in which any halo in the model that contains less than the universal baryon fraction ($\Omega_{\text{b}}/\Omega_0$) of its mass in the form of hot gas (indicating that some cooling has already occurred) has its value of r_{core} adjusted so as to give the same gas density at r_{H} as would have been found if no gas had cooled. This of course ceases to be possible if the target density could not be reached even with gas uniformly spread throughout the halo. For this reason a hard upper limit of $r_{\text{core}} = 10r_{\text{H}}$ is imposed. In practice the exact value of this upper limit is irrelevant, since if it is reached there is so little gas remaining that cooling (and hence the density profile) is unimportant. As an alternative choice the semi-analytic model allows for the gas to be initially distributed as the dark matter.

Once the halo has formed the hot gas will begin to cool. A cooling time is defined for the gas as $\tau_{\text{cool}} = E/\dot{E}$ where E is the energy per unit mass of gas, and \dot{E} is the rate at which a unit mass of gas is radiating energy, i.e.

$$\tau_{\text{cool}} = \frac{3}{2} \frac{\rho_{\text{gas}}(r)}{\mu m_{\text{H}}} \frac{k_{\text{B}} T_{\text{gas}}}{n_{\text{e}}^2(r) \Lambda(T_{\text{gas}}, Z_{\text{gas}})}, \quad (2.24)$$

where k_{B} is Boltzmann's constant, n_{e} is the number density of free electrons in the gas and $\Lambda(T_{\text{gas}}, Z_{\text{gas}})$ is the cooling function for gas at temperature T_{gas} and with metallicity Z_{gas} .

The cooling function in a collisionally ionized gas has been tabulated by Sutherland & Dopita (1993). Since cooling is dominated by two-body processes, the cooling time is inversely proportional to the gas density and so is shorter in the centre of dark matter halos where the density is highest. This results in a cooling wave which propagates out from the centre of the halo behind which gas has cooled significantly. The radius of this cooling wave is approximately the radius at which the cooling time of the gas equals the

age of the halo. In the semi-analytic model all gas within this radius is assumed to have cooled and sunk to the centre of the halo, providing that the free-fall time for the gas at the cooling radius is shorter than the halo lifetime. If this is not the case only gas for which the free-fall time, given by

$$t_{\text{ff}}(r) = \int_0^r \left[\int_r^{r''} -\frac{2GM(r')}{r'^2} dr' \right]^{-1/2} dr'', \quad (2.25)$$

is less than the halo lifetime is allowed to reach the halo centre. The accuracy of this simple cooling model is tested against numerical simulations of galaxy formation in Chapter 7.

In Chapter 5 we consider the effects of Compton cooling, in which the ionized halo gas is cooled as cosmic microwave background photons scatter off electrons in the halo plasma.

2.3 Disk and Star Formation

Once gas has cooled and sunk to the centre of the dark matter halo the semi-analytic model assumes that a rotationally supported gas disk forms as will be described in §2.3.1. In §2.3.2 we describe how, having settled into this disk, the gas begins to transform into stars. The most massive stars can return a significant amount of mass and energy into the inter-stellar medium (ISM) during their lifetimes and in the supernova event which signals the end of their life. As will be described in §2.3.3, the semi-analytic model assumes that this energy input can reheat some of the gas in the galactic disk and cause it to be expelled from the disk out into the surrounding halo. The mass that high mass stars return to the ISM is also enriched in metals (which are formed within these stars). In §2.3.4 we describe the model of chemical evolution used to follow this process. Finally, in §2.3.5 we list the complete set of equations which govern the flow of gas and metals between the hot, cold and stellar gas reservoirs.

2.3.1 Disk Formation

As gas cools and flows to the centre of the halo it will settle into a rotationally supported disk providing that it does not transfer all of its angular momentum to the surrounding dark matter through tidal torques. The semi-analytic model assumes that no angular momentum is transferred from gas to halo during cooling and furthermore assumes that the specific angular momentum of gas is equal to that of the dark matter at the radius from

which it originated*. Hydrodynamic simulations have repeatedly failed to create realistic galaxy disks, since gas typically cools very early before tidal torques have imparted sufficient angular momentum to it, and, furthermore, the lumpy nature of the halo merging process tends to transfer angular momentum outwards as gas cools, making the problem worse (see, for example, Navarro & Steinmetz 1997, Weil, Eke & Efstathiou 1998). However, as noted by Cole et al. (2000) the semi-analytic model assumptions seem to be in line with the simulated halos of Eke, Navarro & Frenk (1998), who include a crude ‘feedback’ scheme to prevent gas cooling too early. Knowing the angular momentum of the disk material it is possible to calculate its size, as will be described in §2.4.3. The model assumes that the loss of any material from the disk (through supernovae feedback for example) does not affect the specific angular momentum of the remaining material.

2.3.2 Star Formation

With cold gas now settled into a rotationally supported disk, the semi-analytic model assumes that gas begins to turn into stars at a rate

$$\psi = M_{\text{cold}}/\tau_{\star}, \quad (2.26)$$

where M_{cold} is the mass of cold gas in the disk and τ_{\star} is a characteristic timescale for star formation. The physics which determine the global star formation rate for a galaxy are still poorly understood (see Kennicutt (1998b) for a recent review). For this reason the semi-analytic model adopts the following flexible and generic form for τ_{\star} which is broadly compatible with the data of Kennicutt (1998a),

$$\tau_{\star} = \epsilon_{\star}^{-1} (V_{\text{disk}}/200 \text{ km s}^{-1})^{\alpha_{\star}} \tau_{\text{disk}}, \quad (2.27)$$

where V_{disk} is the disk circular velocity and τ_{disk} is the dynamical timescale of the disk ($\tau_{\text{disk}} = \pi r_{\text{disk}}/V_{\text{disk}}$), both evaluated at the half mass radius of the disk. The first term,

*Since the gas and dark matter density profiles have different shapes in our model, whilst the rotation speed for each component at any given radius is assumed to be the same the specific angular momentum of gas and dark matter averaged over the entire halo will not be the same. Cole et al. (2000) have chosen to force the average specific angular momentum of gas to be the same as that of the dark matter by reducing V_{rot} for the gas by a constant factor throughout each halo (the value of this factor depends on the particular density profiles of gas and dark matter used). For the profiles used throughout this Thesis, Cole et al. (2000) find that this modification only changes the size of galaxy disks in the model (once other parameters have been adjusted to recover a good match to the galaxy luminosity function). Typically, disks are found to be smaller by a factor of approximately 0.75.

ϵ_* , is an efficiency which determines the fraction of gas that is turned into stars per dynamical time of the disk for a Milky Way like galaxy (i.e. one with a rotation speed of around 200km/s). The second term controls how this efficiency depends on the disk circular velocity. The two parameters ϵ_* and α_* are constrained by requiring the model to reproduce a range of observational data as described by Cole et al. (2000) (see also §2.6.2).

2.3.3 Supernovae Feedback

Stellar winds and supernovae originating from high mass stars inject energy (both thermal and kinetic) into the ISM (see, for example, Martin 1999). The semi-analytic model assumes that this energy deposition causes some of the cold gas in the ISM to be reheated and ejected from the disk back into the halo. The efficiency of this process is controlled by a parameter, β , which is defined as the mass of cold gas ejected into the halo per unit mass of stars formed. Since gas should be able to escape more easily from galaxies with smaller gravitational potentials, β is allowed to depend on the disk circular velocity. The parameterization chosen for β is

$$\beta = (V_{\text{disk}}/V_{\text{hot}})^{-\alpha_{\text{hot}}}, \quad (2.28)$$

where V_{hot} and α_{hot} are free parameters which can be fixed by constraining the semi-analytic model to match certain observations of the local Universe.

2.3.4 Chemical Enrichment

Some fraction of the metals produced in high mass stars are returned to the ISM by stellar winds and supernovae. These metals become incorporated into later generations of stars and may also be returned to the hot halo gas by the effects of supernovae feedback. An instantaneous recycling model is adopted to follow the chemical enrichment of gas in the semi-analytic model (Searle & Sargeant 1972, Pagel & Patchett 1975).

For a given stellar initial mass function (IMF) the yield, p , (which is the mass of metals returned to the ISM per unit mass of stars formed) and a recycled fraction, R , (defined as the mass of gas returned to the ISM per unit mass of stars formed) are chosen. An allowance is made for the possibility that some fraction, e , of the metals produced are ejected directly into the hot halo gas surrounding the galaxy. Typically this fraction is assumed to be zero, however.

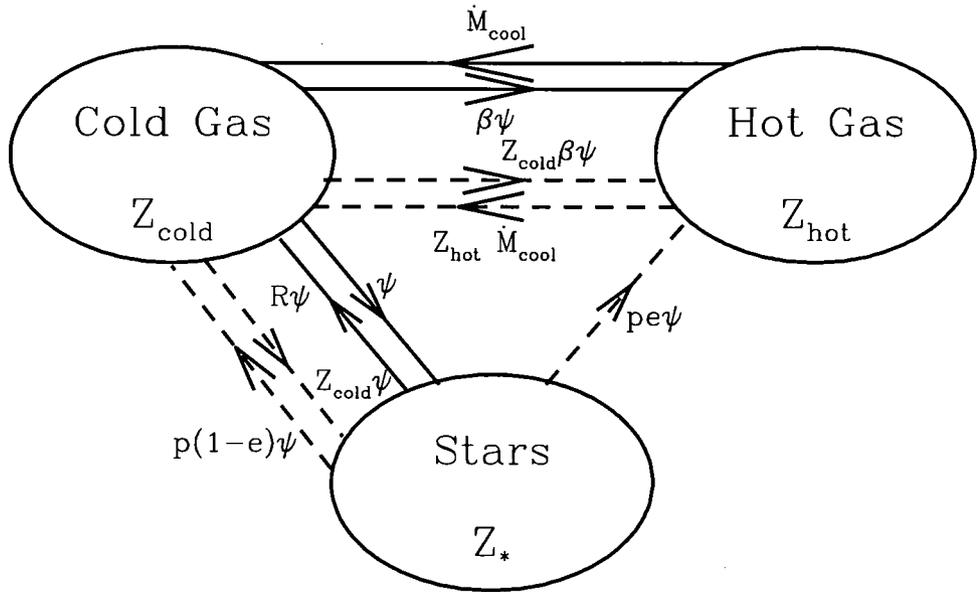


Figure 2.2: Schematic showing the flow of mass and metals between three phases (hot halo gas, cold gas in the galactic disk and stars in the disk). Solid lines show routes and rates at which mass is transferred between the three phases, whilst dashed lines show routes and rates for the transfer of metals. ψ is the instantaneous star formation rate, \dot{M}_{cool} is the rate of accretion of gas from the hot halo, p and R are the yield and recycled fraction of the IMF respectively, β and e are the feedback and ejection parameters described in the text and Z_{hot} , Z_{cold} and $Z_{\text{*}}$ are the metallicities of hot halo gas, cold disk gas and stars respectively.

Unlike the “closed box” model of chemical enrichment the metallicity of gas and stars in this model are affected by accretion of cold gas from the surrounding halo (and the metallicity of that gas). The effective yield, p_{eff} , also differs from the “closed box” model due to the effects of feedback and the ejection term (e) such that we find $p_{\text{eff}} = p(1 - e)/(1 - R + \beta)$, as will be shown in §2.3.5.

2.3.5 Gas Equations

The processes described above determine the rates of change of the masses of gas and metals in three phases: hot gas in the dark matter halo, cold gas in the galaxy disk, and gas locked up in stars. The flow of gas and metals between these phases is shown in Fig. 2.2.

The equations governing the flow of mass and metals are as follows:

$$\dot{M}_* = (1 - R)\psi \quad (2.29)$$

$$\dot{M}_{\text{hot}} = -\dot{M}_{\text{cool}} + \beta\psi \quad (2.30)$$

$$\dot{M}_{\text{cold}} = \dot{M}_{\text{cool}} - (1 - R + \beta)\psi \quad (2.31)$$

$$\dot{M}_*^Z = (1 - R)Z_{\text{cold}}\psi \quad (2.32)$$

$$\dot{M}_{\text{hot}}^Z = -\dot{M}_{\text{cool}}Z_{\text{hot}} + (pe + \beta Z_{\text{cold}})\psi \quad (2.33)$$

$$\begin{aligned} \dot{M}_{\text{cold}}^Z &= \dot{M}_{\text{cool}}Z_{\text{hot}} \\ &+ (p(1 - e) - (1 + \beta - R)Z_{\text{cold}})\psi, \end{aligned} \quad (2.34)$$

where M_{hot} , M_{cold} and M_* are the masses of baryons in the hot, cold and stellar phases respectively (note that the mass of stars is the mass which is permanently locked up in stars, i.e. it excludes mass which is recycled and returned to the ISM). A dot indicates a derivative with respect to time and a superscript Z indicates a mass of metals. The metallicities Z are defined as $Z_{\text{hot}} = M_{\text{hot}}^Z/M_{\text{hot}}$ and $Z_{\text{cold}} = M_{\text{cold}}^Z/M_{\text{cold}}$.

These equations can be solved to give the change in mass and metallicity in each gas reservoir over a short timestep. A full derivation of these solutions is given in Appendix A.3.

2.4 Galaxy Merging

The process of gas cooling within a dark matter halo can only create galaxy disks in this model. Elliptical galaxies, and the spheroidal components of spiral galaxies are assumed to be made via the destruction of two pre-existing galaxies in a violent merger event, or by gradual accretion of smaller satellites (i.e. minor mergers) as any stars from such galaxies are assumed to end up in the spheroidal component of the central galaxy. In this section we describe the details of the dynamical friction which drives galaxy merging and also explain how the sizes of galaxy disks and spheroids are determined.

2.4.1 Dynamical Friction

In a hierarchical universe a galaxy which formed in a small halo of mass M_s may, at some later time, find itself falling into a larger halo of total mass M_H . Numerical simulations (see, for example, Moore et al. 1998) of this process indicate that the core of the dark matter halo in which the galaxy formed survives for several dynamical times inside the larger halo.

Once the galaxy, which will now be known as the satellite galaxy, enters the larger halo it will begin to experience the effects of dynamical friction, whereby its orbital energy is transferred to the dark matter of the larger halo through gravitational interaction. This process will eventually force the satellite galaxy to spiral to the centre of the halo, where it will merge with any galaxy which happens to be there.

The merger timescale for galaxies is derived in detail in Appendix A.4 and can be conveniently written as

$$\tau_{\text{mrg}} = f_{\text{df}} \Theta_{\text{orbit}} \tau_{\text{dyn}} \frac{0.3722 M_{\text{H}}}{\ln(\Lambda_{\text{C}}) M_{\text{S}}}, \quad (2.35)$$

where f_{df} is a dimensionless parameter describe below, M_{S} is the mass of the satellite galaxy plus its original dark matter halo, $\tau_{\text{dyn}} = \pi r_{\text{h}}/V_{\text{C}}$ is the dynamical time of the larger halo, and the Coulomb logarithm is taken to be $\ln(\Lambda_{\text{C}}) = \ln(M_{\text{H}}/M_{\text{S}})$. The factor Θ_{orbit} combines all dependency on the orbital parameters of the satellite and is defined as

$$\Theta_{\text{orbit}} = [J/J_{\text{c}}(E)]^{0.78} [r_{\text{c}}(E)/r_{\text{H}}]^2, \quad (2.36)$$

where J and E are the initial orbital angular momentum and energy respectively of the satellite and $J_{\text{c}}(E)$ and $r_{\text{c}}(E)$ are the angular momentum and radius respectively of a circular orbit with energy E . For the initial orbital parameters of the satellite Cole et al. (2000) pick a value of $\log \Theta_{\text{orbit}}$ from a Gaussian with $\langle \log \Theta_{\text{orbit}} \rangle = -0.14$ and variance $\langle (\log \Theta_{\text{orbit}})^2 \rangle - \langle \log \Theta_{\text{orbit}} \rangle^2 = 0.0676$, which they find to be a good fit to the distribution of $\log \Theta_{\text{orbit}}$ found by Tormen (1997) in cosmological N-body simulations. Since the merger timescale, τ_{mrg} , determined in this way is based on a number of assumptions (e.g. treating the satellite as a point mass) the parameter f_{df} is introduced to allow some freedom in determining merger timescales. The merger rate resulting from this expression is compared to that found in numerical simulations of galaxy formation in Chapter 7.

When a new halo is formed all satellite galaxies are assigned a value of $\log \Theta_{\text{orbit}}$ chosen from the distribution described above, and the corresponding value of τ_{mrg} computed. The satellite is then made to merge with the central galaxy of the halo after a time τ_{mrg} has elapsed. If the halo merges with another (thereby forming yet another new halo) before this time is reached each satellite is assigned a new τ_{mrg} and begins the whole process again.

2.4.2 Spheroid Formation

When a galaxy of mass M_1 merges with another galaxy of mass $M_2 > M_1$ the event is termed a ‘major merger’ if $M_1 > f_{\text{ellip}}M_2$. In this case the galaxies are assumed to be completely disrupted, leaving a single spheroidal remnant. Any gas present in either of the original galaxies is turned into stars using the normal star formation rules, but with a very short star formation timescale, in a starburst. For other (i.e. non-major) mergers the stars of the satellite galaxy are added to the bulge of the central galaxy, and any gas present is added to the disk of the central galaxy.

2.4.3 Sizes

Since the specific angular momentum and total mass of the gas cooling to form a galactic disk are known it is possible to compute the size of the disk as will be described below. With a model for the pseudo-specific angular momentum of spheroids formed in merger events (see §2.4.4) a similar calculation, also described below, can be carried out for these objects. The formation of the disk and spheroid of a galaxy will have a gravitational effect on the surrounding dark matter halo. We explore this effect first.

Halo contraction

As gas cools and condenses within a halo to form a galaxy at the centre it will exert a gravitational influence on the dark matter in the halo. The gravitational pull of the galaxy will in fact compress the halo, causing the dark matter to move towards the halo centre. This process can significantly affect the circular velocity of the galaxy itself.

For a halo particle in a periodic orbit the quantity $\oint pdq$, where p is the canonical momentum conjugate to the coordinate q and where the integral is taken along the orbit of the particle, is an adiabatic invariant. This can be seen to be true since

$$\frac{d}{dt} \oint pdq = \oint \frac{dp}{dt} dq = \oint \frac{\partial H}{\partial q} dq = 0, \quad (2.37)$$

where H is the Hamiltonian of the system and the final step follows from the fact that the Hamiltonian (i.e. the total energy of the particle) must be unchanged around the orbit. Thus for a particle on a circular orbit in a spherically symmetric potential we find that the quantity rV_H (the pseudo-specific angular momentum of the particle) is an adiabatic invariant (i.e. $p = mV_H$ for a particle of mass m in a circular orbit, and so $\oint pdq = 2\pi rmV_H$). As long as the condensation of the central galaxy occurs sufficiently slowly

this quantity will be conserved for each particle in the halo (for a thorough discussion of this result see Barnes & White 1984, Blumenthal et al. 1986, Flores et al. 1993).

Applying this result to each spherical shell of dark matter the initial, r_0 , and final, r , radii of the shell can be related by

$$r_0 V_{\text{cH0}}(r_0) = r V_{\text{cH}}(r), \quad (2.38)$$

where $V_{\text{cH0}}(r_0)$ and $V_{\text{cH}}(r)$ are the halo circular velocities before and after galaxy condensation respectively. With the assumption that the baryonic component of the halo is distributed like the dark matter, and defining $1 - f_{\text{H}}$ to be the fraction of the total mass which forms the central galaxy the final, $M_{\text{H0}}(r_0)$, and initial, $M_{\text{H}}(r)$, masses within the shell (assuming no shell crossing) are related by

$$M_{\text{H}}(r) = f_{\text{H}} M_{\text{H0}}(r_0). \quad (2.39)$$

The circular velocity of the halo can now be found by assuming that both contributors (halo material and galaxy material) are spherically symmetric. For the case of galaxy disks Cole et al. (2000) claim that this is a better estimate of the effect of the disk on the dark matter than using the circular velocity due to the disk in the disk plane, which is somewhat larger. The circular velocity is then

$$V_{\text{cH}}^2(r) = G[M_{\text{H}}(r) + M_{\text{D}}(r) + M_{\text{B}}(r)]/r, \quad (2.40)$$

where $M_{\text{D}}(r)$ and $M_{\text{B}}(r)$ are the masses of disk and bulge respectively within radius r . The consistency condition $M_{\text{H0}}(r) = f_{\text{H}} M_{\text{H0}}(r) + M_{\text{D}}(r) + M_{\text{B}}(r)$ (which ensures that the virial radius of the halo is unchanged) must be satisfied. Combining (2.38), (2.39) and (2.40) gives

$$r_0 M_{\text{H0}}(r_0) = r [f_{\text{H}} M_{\text{H0}}(r_0) + M_{\text{D}}(r) + M_{\text{B}}(r)]. \quad (2.41)$$

Equation (2.41) relates, for given disk and bulge mass profiles, the final radius of any halo shell to its initial radius, and therefore allows the rotation curve of the halo including the effects of the central galaxy to be determined.

Disk and bulge sizes

In §2.4.3 we described how the mass profile of the dark matter halo in the presence of a central galaxy can be calculated. Equation (2.41), which defines the mass profile, depends upon the mass profiles of the disk and spheroidal components of the central galaxy. Here we describe how these profiles are calculated in the semi-analytic model.

For the disk component the total mass, M_D , and the specific angular momentum of the disk material, j_D , are known (see §2.1.3 and §2.2). The disk is assumed to adopt an exponential surface density profile, $\Sigma_D(r)$, of the form

$$\Sigma_D(r) = \frac{M_D}{2\pi h_D^2} \exp\left(-\frac{r}{h_D}\right), \quad (2.42)$$

where h_D is the radial scale length of the disk, related to the disk half-mass radius, r_D , by $2(r_D/h_D + 1) \exp(-r_D/h_D) = 1$ or, approximately, $r_D = 1.68h_D$. The total disk mass within radius r is then simply

$$\begin{aligned} M_D(r) &= \int_0^r 2\pi r \Sigma_D(r) dr \\ &= M_D \int_0^r \exp\left(-\frac{r}{h_D}\right) \frac{r}{h_D} \frac{dr}{h_D} \\ &= M_D [1 - (1 + r/h_D) \exp(-r/h_D)]. \end{aligned} \quad (2.43)$$

The specific angular momentum of the disk is given by

$$\begin{aligned} j_D &= \int_0^\infty 2\pi r \Sigma_D(r) r V_{cD}(r) dr / M_D \\ &= r_D V_{cD}(r_D) \int_0^\infty \frac{r^2}{h_D^2} \exp\left(-\frac{r}{h_D}\right) \frac{V_{cD}(r)}{1.68 V_{cD}(r_D)} \frac{dr}{h_D} \\ &= k_D r_D V_{cD}(r_D), \end{aligned} \quad (2.44)$$

where

$$k_D = \int_0^\infty \frac{r^2}{h_D^2} \exp\left(-\frac{r}{h_D}\right) \frac{V_{cD}(r)}{1.68 V_{cD}(r_D)} \frac{dr}{h_D}, \quad (2.45)$$

and $V_{cD}(r)$ is the circular velocity of the disk. For a flat rotation curve eqn. (2.45) gives a value of 1.19 for k_D . Cole et al. (2000) note that the value of k_D is only weakly dependent on the form of the rotation curve used and so $k_D = 1.19$ is used for all galaxies in the semi-analytic model.

Using the relation $V_{cD}^2(r_D) = GM(r_D)/r_D$ (where $M(r_D)$ is the total mass within radius r_D) along with (2.44) we find,

$$\begin{aligned} j_D^2 &= k_D^2 r_D^2 V_{cD}^2(r_D) \\ &= k_D^2 G r_D \left[f_H M_{H0}(r_{D0}) + \frac{1}{2} k_h M_D + M_B(r_D) \right], \end{aligned} \quad (2.46)$$

where we have used the relation $M_H(r) = f_H M_{H0}(r_0)$ (eqn. 2.39), and $M_D(r_D) = M_D/2$ by definition of r_D . The constant k_h accounts for the disk geometry (since the disk is not spherically symmetric its contribution to V_{cD}^2 is not simply $GM_D(r)/r$). For an exponential disk Binney & Tremaine (1987, eqn. 2-169) find $k_h = 1.25$.

The disk size, r_D , must satisfy both eqn. (2.46) and eqn. (2.41) evaluated at r_D , i.e.

$$r_{D0}M_{H0}(r_{D0})r_D = \left[f_H M_{H0}(r_{D0}) + \frac{1}{2}M_D + M_B(r_D) \right]. \quad (2.47)$$

This is simplified slightly by using eqn. (2.46) to eliminate M_B from (2.47) giving

$$r_{D0}M_{H0}(r_{D0}) = \frac{j_D^2}{k_D^2 G} - \frac{1}{2}(k_h - 1)r_D M_D. \quad (2.48)$$

The spheroidal component of the galaxy is treated in a similar way. Firstly we assume that the projected density of the spheroid is described by a de Vaucouler's $r^{1/4}$ -law (see, for example, Binney & Tremaine 1987, eqn. 1-13). This profile is usually characterised by its effective radius, r_{eff} , which is the radius containing half the mass in projection. It is related to the three-dimensional half mass radius, r_B , by $r_B = 1.35r_{\text{eff}}$. A pseudo-specific angular momentum can be defined for the spheroid:

$$j_B = r_B V_c(r_B) \quad (2.49)$$

where $V_c(r)$ is the circular velocity profile inside the spheroid. j_B is assumed to be conserved, except during mergers of galaxies at which time a new value of j_B is calculated (see §2.4.4). Analogous to the disk case, we find two equations which must be satisfied by the spheroid radius:

$$\begin{aligned} j_B^2 &= r_B^2 V_c^2(r_B) \\ &= Gr_B \left[f_H M_{H0}(r_{B0}) + M_D(r_B) + \frac{1}{2}M_B \right] \end{aligned} \quad (2.50)$$

and

$$r_{B0}M_{H0}(r_{B0}) = \frac{j_B^2}{G}. \quad (2.51)$$

Note that in (2.50) it is assumed that the mean effect of the disk on the spheroid can be estimated by spherically averaging over the disk.

For a given initial halo mass distribution, $M_{H0}(r_0)$, disk and bulge masses, M_D and M_B , and disk and bulge specific angular momenta, j_D and j_B , (2.46), (2.48), (2.50) and (2.51) form a coupled set of equations which can be solved (with a careful numerical approach) for r_D and r_B using an iterative procedure.

2.4.4 The Merger Rule

As shown in §2.4.3, to determine the radius of the spheroidal component of a galaxy we need to determine its pseudo-specific angular momentum. To do this the semi-analytic

model assumes that spheroids form when two galaxies collide (their orbits having been degraded by dynamical friction). The galaxies are assumed to merge when their separation equals the sum of their half mass radii. If we assume that each galaxy (i.e. the two merging galaxies and the one that is formed by the merger) are in virial equilibrium, then their total energy equals one half of their gravitational self-binding energy,

$$E_{\text{bind}} = -\frac{c}{2} \frac{GM}{r}, \quad (2.52)$$

where M and r are the mass and half mass radius respectively of the galaxy and c is a constant which depends on the distribution of the mass. For an $r^{1/4}$ -law galaxy $c = 0.45$ can be found by numerical integration whilst for an exponential disk $c = 0.49$ (Binney & Tremaine 1987, eqn. 2-171). For simplicity a value of $c = 0.5$ is adopted for all galaxies in the model. The orbital energy of the two galaxies just before they merge is given by

$$E_{\text{orbit}} = -\frac{f_{\text{orbit}}}{2} \frac{GM_1M_2}{r_1 + r_2}, \quad (2.53)$$

where M_1 and M_2 are the masses of the merging galaxies and r_1 and r_2 are their half mass radii. The parameter f_{orbit} depends on the orbital parameters of the galaxy pair. A value of $f_{\text{orbit}} = 1$ is adopted, which corresponds to point mass galaxies in circular orbits about their centre of mass. Conservation of energy then requires

$$E_{\text{bind,new}} = E_{\text{bind},1} + E_{\text{bind},2} + E_{\text{orbit}}, \quad (2.54)$$

where $E_{\text{bind,new}}$ is the binding energy of the new galaxy. Upon using (2.52) and (2.53), this becomes

$$\frac{(M_1 + M_2)^2}{r_{\text{new}}} = \frac{M_1^2}{r_1} + \frac{M_2^2}{r_2} + \frac{f_{\text{orbit}}}{c} \frac{M_1M_2}{r_1 + r_2}, \quad (2.55)$$

where r_{new} is the half mass radius of the new, spheroidal galaxy.

Cole et al. (2000) note that these equations imply that in a merger of two identical galaxies the radius of the new galaxy is 4/3 times larger than that of the original galaxies, which is in reasonable agreement with the value of 1.42 found in simulations of equal mass galaxy mergers (Barnes 1992).

The computed value of r_{new} can be used in eqn. (2.50), along with $V_{\text{cB}}(r_{\text{B}}) = G(M_1 + M_2)/2r_{\text{new}}$ to find the value of j_{B} . This value of j_{B} is then used in (2.51) to determine the value of r_{B} in the contracted halo profile, using eqns. (2.50) and (2.51), which is then assumed to be the true value of r_{B} . This value of r_{B} typically differs only slightly from r_{new} , indicating that the treatment of the dark matter during the merger is approximately self-consistent.

2.5 Spectral Synthesis

Having determined the masses, sizes and star formation and chemical histories of each galaxy the semi-analytic model then turns these quantities into observable properties, i.e. the luminosity of the galaxy in any desired wave band.

The technique of stellar population synthesis, originally implemented by Tinsley (1972, 1980) and developed by Bruzual & Charlot (1993) and Bressan, Chiosi & Fagotto (1994) amongst others, is used, in which the raw galaxy spectral energy distribution (SED) is determined by summing the SEDs of all the stars which make up the galaxy taking into account the age of each episode of star formation (§2.5.1). Emission lines, which arise in HII regions near young stars are also modelled (§2.5.2). These raw SEDs are then modified to account for the presence of obscuring material (i.e. dust) within the galaxy itself (§2.5.3).

2.5.1 Stellar Spectra

In the stellar population synthesis technique the SED per unit mass of stars, $\Phi_\lambda(t, Z)$, is computed for a single stellar population (i.e. stars of a single metallicity and age), as a function of the population's age, t , and metallicity, Z . The SEDs of single stellar populations have been computed by Bruzual & Charlot (2000). These single stellar populations can be combined, suitably weighted by a star formation rate, to give the total SED of a galaxy.

To compute the $\Phi_\lambda(t, Z)$ a form must be chosen for the stellar IMF. In the semi-analytic model the IMF is assumed to be constant in space and time. The standard choice of IMF is that of Kennicutt (1983), which has been found to agree most closely with available data in the Solar neighbourhood (Scalo 1998). Other choices of IMF are available within the semi-analytic model. The possibility of brown dwarfs being formed is also included in this model. The fraction of brown dwarfs formed is determined by the parameter Υ , which is defined as (mass in visible stars + mass in brown dwarfs)/(mass in visible stars). Therefore $\Upsilon \geq 1$. The effect of Υ is simply to reduce all luminosities by a factor of Υ . The chosen IMF also determines the values of R and p (the recycled fraction and yield of the stars respectively). Values of R and p calculated from stellar evolution models (Renzini & Voli 1981, Woosley & Weaver 1995, Marigo, Bressan & Chiosi 1996, Portinari, Chiosi & Bressan 1998) and appropriate for various IMFs are given by Cole et al. (2000).

The raw SED of each galaxy is found by convolving the star formation rate, $\psi(t)$, with the single stellar population SEDs:

$$S_\lambda(t) = \int_0^t \Phi_\lambda(t-t', Z(t')) \psi(t') dt', \quad (2.56)$$

where $S_\lambda(t)$ is the resulting SED at time t and $Z(t)$ is the metallicity of stars formed at time t . In evaluating this integral the SEDs of Bruzual & Charlot (2000) are linearly interpolated in t and $\log Z$ to intermediate ages and metallicities. Finally, the broad band luminosity of the galaxy can be determined by summing the SED under the appropriate filter response function.

Cole et al. (2000) note that the resulting galaxy SEDs are quite sensitive to the choice of IMF (and its associated yield). The resulting uncertainty is greater than the differences between different stellar population synthesis models found by Charlot, Worthey & Bressan (1996) for old stars. For the case of very young stars, however, there is still a significant variation in spectra between population synthesis models.

2.5.2 Emission Line Modelling

Emission line properties are calculated in this model by first determining the luminosity in Lyman continuum photons produced by the stellar population. This is then combined with models of HII regions to determine the luminosities and equivalent widths of emission lines such as H α and OII.

For the purposes of this Thesis we will be interested only in the H α line (see Chapter 5). The luminosity of this line is found by simply assuming that all photons shortwards of 912Å (i.e. those that are energetic enough to ionize hydrogen) emitted by stars *will* ionize a hydrogen atom in the gas surrounding the star. We then assume “case B” recombination, meaning we ignore recombinations directly to the ground state ($n = 1$), since such recombinations just produce another ionizing photon. Thus, only recombinations to levels $n = 2, 3, \dots$ are counted. The fraction of such recombinations which produce an H α photon ($n = 2 \rightarrow 1$) is then taken from Osterbrock (1974).

2.5.3 Dust Extinction

The presence of dust in the ISM can significantly attenuate the optical luminosity of a galaxy as dust grains both absorb and scatter light. The effect of dust is even greater on far-ultraviolet luminosities. Within the semi-analytic model the effect of dust on each

galaxy SED is computed using the model of Ferrara et al. (1999), in which the radiative transfer of light from stars in a disk and bulge is followed through dust distributed smoothly in the galactic disk. The output of this model is the attenuation as a function of wavelength (λ), galaxy inclination angle (i), central optical depth of the galaxy in the V-band (τ_{V0}), ratio of bulge and disk scale lengths (r_e/h_D) and the ratio of dust and stellar vertical scale heights ($h_{z,dust}/h_{z,stars}$).

Stars are assumed to be distributed in a bulge and in an exponential disk with a vertical scale height equal to 0.0875 times the disk's radial scale length. This ratio was adopted by Ferrara et al. (1999) to match the observed values for the old disk population of galaxies like the Milky Way. A fixed value of $h_{z,dust}/h_{z,stars}$ is assumed ($h_{z,dust}/h_{z,stars} = 1$ being the standard choice) and galaxies are assigned random inclination angles i . The central optical depth in the V-band is computed from the known properties of the galaxy:

$$\tau_{V0} \propto \frac{M_{dust}}{\pi r_{disk}^2} \propto \frac{M_{cold} Z_{cold}}{r_{disk}^2}, \quad (2.57)$$

where Z_{cold} is the metallicity of the cold gas in the galaxy disk. Note that the mass of dust in the galaxy is assumed to be proportional to both the mass of cold gas in the disk and the metallicity of that gas. The above equation is normalized by requiring gas of metallicity $Z = 0.02$ to have the dust to gas ratio measured in the local ISM. Cole et al. (2000) show that this leads to

$$\tau_{V0} = 0.043 \left(\frac{M_{cold}/(2\pi h_R^2)}{M_{\odot} \text{pc}^{-2}} \right) \left(\frac{Z_{cold}}{0.02} \right). \quad (2.58)$$

Cole et al. (2000) find that the effects of dust are quite insensitive to the choice of $h_{z,dust}/h_{z,stars}$ and also to the choice of extinction curve (the Milky Way extinction curve is used as standard, but a Small Magellanic Cloud extinction curve may also be used). In the models used in this Thesis the reemission of the absorbed energy at longer wavelengths is not included (since we do not consider wavelengths where this emission is important). The model of Cole et al. (2000) has, however, now been extended to include reemission by dust (see Granato et al. 2000).

2.6 Using the Model

The semi-analytic model described generates several Monte-Carlo merger trees for halos in a range of masses. The galaxy formation rules described in this Chapter are then applied to each merger tree to determine the properties of galaxies residing within that halo, at any redshift(s) of interest.

In §2.6.1 we show the parameters of the model, and also describe the model outputs. In §2.6.2 we briefly describe the observational constraints which have been applied to the model. Both of these topics are covered in a much more thorough manner by Cole et al. (2000).

2.6.1 Parameters and Outputs

The parameters of the semi-analytic model can be split into four classes: numerical parameters; cosmological parameters; galaxy formation parameters and dynamical parameters. These are reviewed below.

Numerical parameters

The parameters M_{res} and N_{steps} , which determine the progenitor halo mass resolution and the number of timesteps used in constructing merger trees, should not be thought of as free parameters of the model. Instead they simply control the numerical implementation of the model and are adjusted until the model results converge.

Cosmological parameters

The cosmological parameters Ω_0 , Λ_0 , h^\dagger (as defined in Appendix A.1), Γ , σ_8 (as defined in §2.1.1) and Ω_b (the mass density of baryons in the Universe expressed in units of the critical density), specify the background cosmology in which the galaxy formation calculations take place. As such these parameters can be constrained from observations unrelated to galaxy formation (e.g. measurements of fluctuations in the cosmic microwave background such as will be made by the MAP and Planck satellites).

Galaxy formation parameters

This set of parameters determines the processes of star formation, feedback and chemical enrichment within each galaxy. Star formation is described by two parameters, ϵ_* and α_* , along with an IMF. Feedback is described by two parameters, V_{hot} and α_{hot} , along with a third, e , which determines the fraction of metals ejected into the hot halo gas by supernovae. Finally, chemical enrichment is described by a recycled fraction, R , and a yield, p , (both of which are fully specified, in principle, by the choice of IMF). A final parameter, Υ , determines the fraction of stars formed which are brown dwarfs.

[†]Throughout this Thesis we will express the Hubble constant, H_0 as $100h\text{km/s/Mpc}$.

Dynamical parameters

Finally, f_{df} and f_{ellip} control the rate of galaxy merging and the frequency of violent mergers respectively.

Model output

The semi-analytic model produces a list of properties for each galaxy residing in each dark matter halo which was modelled. The properties given for each galaxy are:

- a flag indicating whether the galaxy is a central galaxy or a satellite galaxy;
- mass of stars in disk and bulge components;
- mass of cold gas in the disk component;
- metallicities and ages of disk and bulge stars weighted by mass or luminosity in any band;
- metallicities of cold gas and hot halo gas;
- instantaneous star formation rate;
- luminosities in any desired band;
- disk and bulge half mass radii (plus the combined half mass radius);
- circular velocities at disk and bulge half mass radii (or the luminosity weighted circular velocity);
- mass and circular velocity at the virial radius of the halo in which the galaxy was last a central galaxy, and of the halo in which the galaxy resides at the output time.

The effects of dust can be computed from these properties (namely the cold gas mass and metallicity and the disk and bulge radii). The surface brightness and isophotal magnitudes can also be determined from the luminosities and disk and bulge radii.

Since the abundance of each halo modelled is known (from the Press-Schechter mass function) it is possible to construct mock galaxy catalogues for any required volume of the universe which contain any of the above properties.

For the purposes of the work carried out in this Thesis other outputs have been added to the model. These are:

- the three dimensional position and velocity of each galaxy (see Chapter 3);

- the ionising ($\lambda < 912\text{\AA}$) luminosity from each halo, plus the size of the ionization front growing around the halo (see Chapter 5).

2.6.2 Observational Constraints

In order to fix the values of the model parameters a set of observations are used to constrain the model. Here we will not describe the procedures used to constrain the various free parameters since a detailed account of this process has been given by Cole et al. (2000), who also explore in detail the effects of changing model parameters. In all of the work presented in this Thesis we use parameter values similar to those of the model presented by Cole et al. (2000)[†], unless otherwise stated.

Instead we list below the primary observational constraints and the parameters which they are used to fix. In §2.7 we will show these same observational constraints and present the results from the fiducial models in each case.

The morphological mix of bright galaxies in the local Universe is used to constrain the parameter f_{ellip} . The feedback parameters V_{hot} and α_{hot} are constrained by the local b_J -band luminosity function, the Tully-Fisher relation and the size of galaxy disks. A low value of α_{hot} is required to produce a straight Tully-Fisher relation, whilst a high value helps reduce the slope of the faint end of the luminosity function. The value of $\alpha_{\text{hot}} = 2$ that has been adopted is a compromise between these two considerations. The value of V_{hot} also affects the faint end of the luminosity function, but is further constrained by the sizes of galaxy disks. The star-formation parameters ϵ_* and α_* were constrained by the masses of cold gas in L_* spiral galaxies and faint spiral galaxies respectively. The stellar IMF is chosen to conform to observations of the Solar neighbourhood and to obtain a good match to the observed distributions of galaxy colours. Corresponding values of R and p are taken from Cole et al. (2000). The dynamical friction parameter, f_{df} , is usually left equal to 1, but may be altered if doing so significantly improves the model luminosity function. The value of Υ is chosen to force the b_J -band luminosity function to match that of Zucca et al. (1997) at around L_* . Finally, the ejection parameter e is left equal to zero for all models considered in this Thesis.

[†]Since much of the work in this Thesis was completed before the to be published version of Cole et al. (2000) was finalized it was impossible to have exactly the same parameters as their final model.

2.7 Fiducial Model Results

In this section we will list the parameters of two fiducial models and present some of the basic model results.

2.7.1 Model Parameters

Table 2.1 shows the parameters of the two fiducial models for which results are shown in this Chapter.

2.7.2 Basic Results

In this section we present some of the basic results from the model of galaxy formation for the two fiducial models that we have described. We will not explore the sensitivity of these results to changes in the model parameters as such a study was carried out in detail by Cole et al. (2000), and the effects of parameter variations are considered later (see Chapters 3, 4 and 5).

Luminosity Functions

The local galaxy luminosity function is the most important constraint placed on this model of galaxy formation. Throughout this Thesis we will consider only those models which give a good match to this function to be realistic, and will use this as a criterion for ruling out models. Strong emphasis is placed on this particular constraint as it has been well measured over a range of luminosities and because it is a fundamental property of the galaxy population. Furthermore, matching this function is of great importance if the model is to produce realistic predictions for other quantities (e.g. a model which fails to reproduce the bright end of the luminosity function may give incorrect results when tested against a magnitude limited sample).

In Fig. 2.3 we show the model luminosity functions in the blue optical b_J band and in the near infrared K band. A compilation of observational data is shown as noted in the figure. The various observational determinations all agree well at the bright ends, but show very different results at fainter magnitudes. This is particularly noticeable in the b_J band where the dispersion between surveys is greater than the statistical errors. This implies that either the luminosity functions in the various volumes surveyed are very different, or that there are systematic variations in selection criteria between the surveys.

Table 2.1: Parameters of the fiducial models.

Model	Ω_0	Λ_0	Ω_b	h	Γ	σ_8	ϵ_*	α_*	V_{hot}	α_{hot}	e	f_{ellip}	f_{df}	IMF	P	R	Υ
Λ CDM	0.3	0.7	0.02	0.7	0.19	0.93	0.0067	-1.5	150.0	2.0	0.0	0.3	1.0	Kennicutt	0.02	0.31	1.35
τ CDM	1.0	0.0	0.08	0.5	0.21	0.60	0.0200	-0.5	300.0	2.0	0.0	0.3	0.1 [†]	Salpeter	0.04	0.28	1.42

[†] f_{df} should be approximately 1 or larger. Here we use an artificially low value in order to obtain a good fit to the local b_J and K-band luminosity functions for the τ CDM model.

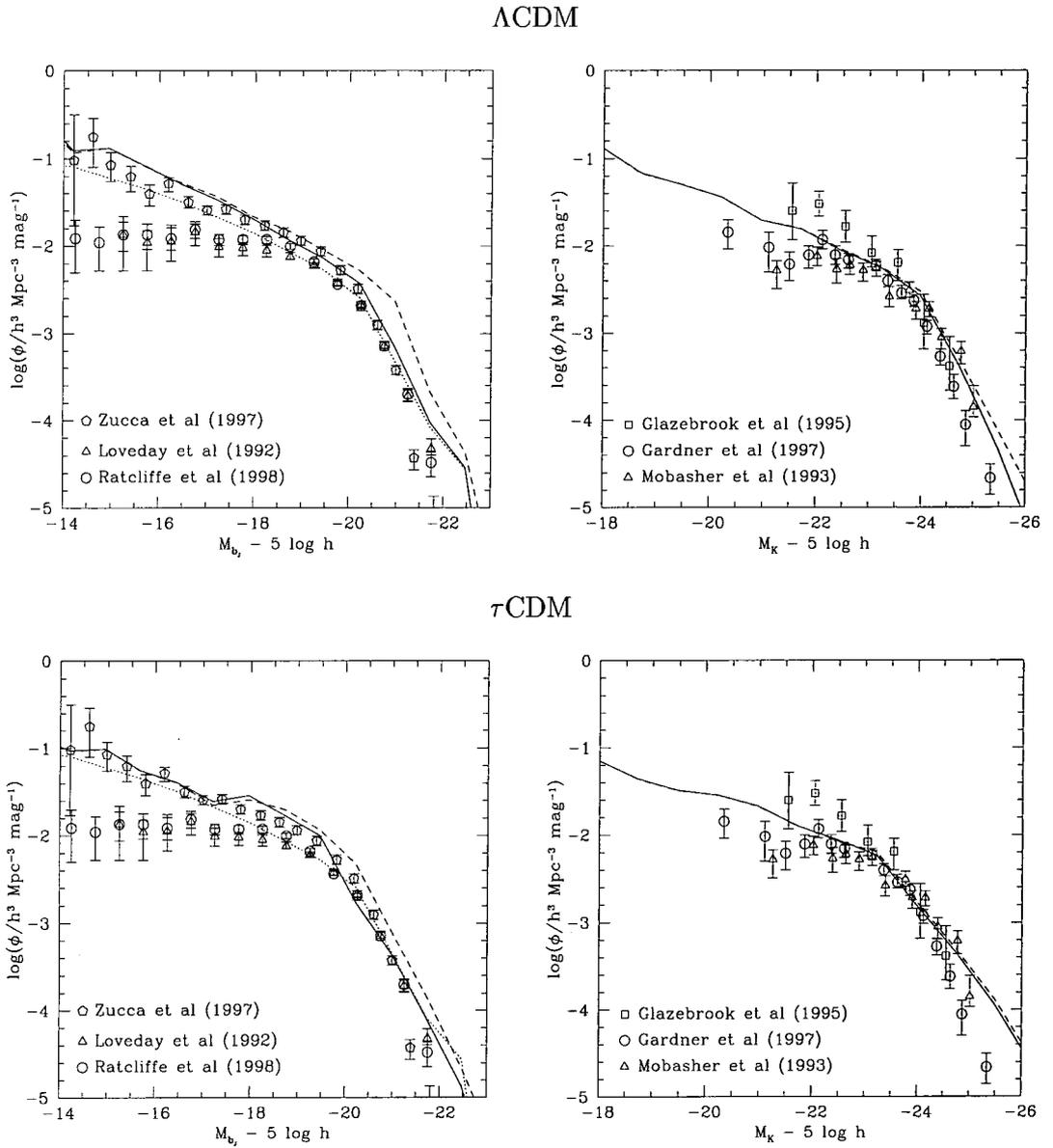


Figure 2.3: The b_J and K-band luminosity functions in the fiducial models compared with a compilation of observational data. The solid line in each case includes the effects of dust and has Υ chosen so as to match the ESP b_J -band luminosity function of Zucca et al. (1997) at $b_J - 5 \log h = -19.8$. The dashed lines show the corresponding luminosity functions before the effects of dust are included. In the left hand panels the dotted line indicates the luminosity function obtained if the 25 mag/arcsec² isophotal magnitude is used rather than the true total magnitude.

For each band we show the model luminosity function with and without the effects of dust. The dust distribution was taken to have the same vertical scaleheight as the gas distribution, but varying this assumption has little effect on the luminosity functions. In the K-band the inclusion of dust has little effect on the luminosity function, which is a reasonable match to the observational data. In the b_J band, however, including dust makes a large difference at the bright end, making galaxies fainter by around 0.5 magnitudes. Without dust the model is a poor fit to the observational data, but with dust included a good match to the bright end of the luminosity function is obtained.

For the b_J band we also show the effects of surface brightness on the model luminosity function. To do this the difference between the total magnitude and the face-on magnitude within the 25 mag/arcsec² isophote was determined, assuming that bulge and disk scalelengths are unaffected by dust extinction. We then assumed that this difference was independent of inclination. Because of these approximations and because some attempt is made in surveys to correct to total magnitudes this line cannot be compared in a quantitative way to the observational data, but does give an indication of the kind of systematic effects which may be present in observational data. This reduces the luminosity of galaxies and lowers the luminosity function. Whilst this has greatest effect at faint magnitudes the shift is still significant (in terms of the observational uncertainties) around L_* .

Tully-Fisher Relation

The Tully-Fisher relation is an observed correlation between the luminosity of spiral galaxies and the peak in their rotation curve. The ability to reproduce this relation is an important test of galaxy formation models as it relates the dynamics of galaxies to their stellar content. We show in Fig. 2.4 the I-band Tully-Fisher relation from the fiducial models compared to the complete diameter limited subset of spiral galaxies selected by de Jong & Lacey (1999) from the catalogue of Mathewson, Ford & Buchhorn (1992). The I-band magnitude of each galaxy, in both the model and the observational data, is corrected to the face-on value including the effects of dust extinction. The velocity plotted here is the maximum in the measured rotation curve for the observed sample, whilst in the model we plot the circular velocity at the half-mass radius of the galaxy, including the self-gravity of disk and bulge, evaluated in the disk mid-plane. Thus the model velocities may be systematically low if the peak occurs at some point other than the half-mass radius, although this effect will be small as the model galaxies have reasonably flat rotation curves.

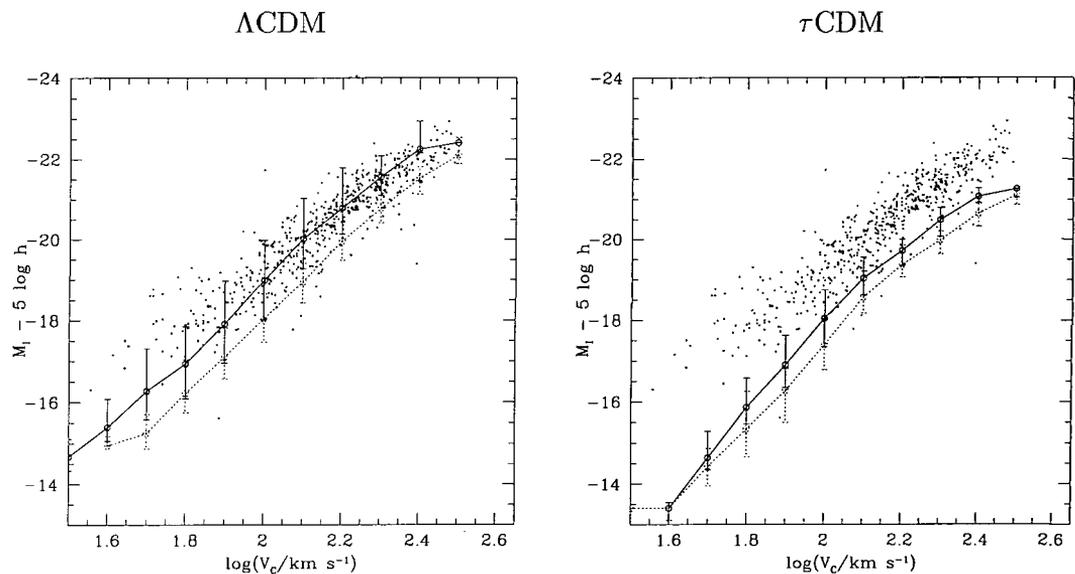


Figure 2.4: The I-band Tully-Fisher relation in the fiducial models. Points show the observed distribution from Mathewson, Ford & Buchhorn (1992), for which all magnitudes were corrected to face-on values, including dust extinction. The dotted line shows the model relation, for galaxies selected to have a bulge-to-total ratio in dust-extincted I-band light of 0.02 to 0.24, and with at least 10% of their disk mass in the form of cold gas. Again, all magnitudes are face-on values including the effects of dust. Circular velocities are measured at the half-mass radius of the disk. The line indicates the median of the model distribution, whilst the error bars show the 10% and 90% intervals of the distribution. The solid line shows the same set of galaxies but now plotted using the circular velocity of the dark matter halo in which they formed rather than the velocity of their disk.

Model galaxies are selected to have a bulge-to-total ratio in dust-extincted I-band light between 0.02 and 0.24, which approximately matches the range of galaxy types (Sb-Sd) included in the Mathewson, Ford & Buchhorn (1992) sample according to the conversion between T-type and bulge-to-total ratio of Baugh, Cole & Frenk (1996b), which was based on the data of Simien & de Vaucouleurs (1986). Model galaxies were also required to have at least 10% of the mass of their disk in the form of cold gas. Galaxies with little cold gas would have no ongoing star formation and would therefore not display prominent spiral arms. Furthermore, gas is required to allow the rotation speed to be measured, either to produce the emission lines needed for optical determinations of the rotation curve, or the 21cm emission needed for HI measurements.

In the Λ CDM cosmology the slope of the model Tully-Fisher relation is in good agreement with the data, and the scatter of 0.75 magnitudes is only slightly smaller than the observed scatter of 1.1 magnitudes. The zero-point, however, is offset from the observational data by approximately 0.8 magnitudes (or alternatively by a factor of 1.25 in velocity). The τ CDM model also has a reasonable slope, but is even further offset from the observed zero-point. This discrepancy is insensitive to most changes in galaxy formation parameters if the model is forced to fit the bright end of the luminosity function by adjustment of Υ (Cole et al. 2000). The predecessor of the Cole et al. (2000) model (Cole et al. 1994), and other semi-analytic models, such as those of Kauffmann et al. (1999a) and Somerville & Primack (1999), constructed the Tully-Fisher relation using the circular velocity of each galaxy's dark matter halo, V_{halo} , rather than the circular velocity of the galaxy itself, V_{disk} . This therefore neglects the self-gravity of each galaxy, which we find results in V_{disk} being approximately 1.3 times V_{halo} . Consequently, when constructed using V_{halo} the model Tully-Fisher zero-point is in much better agreement with that of the data, producing a good fit in the case of our Λ CDM model.

Disk Sizes

Fig. 2.5 compares the distribution of I-band disk scale lengths from the model with the observational data of de Jong & Lacey (1999). In the model the sizes of disks are fundamentally determined by the angular momentum of the proto-disk material, which arises from torques acting on the material as the dark matter halo expands and collapses. The resulting distribution of halo spin parameters, λ_{H} (see §2.1.3), is well understood and reasonably well modelled by the form given in eqn. (2.23). The main uncertainties that arise in calculating the size of disks are therefore the distribution of angular momentum within

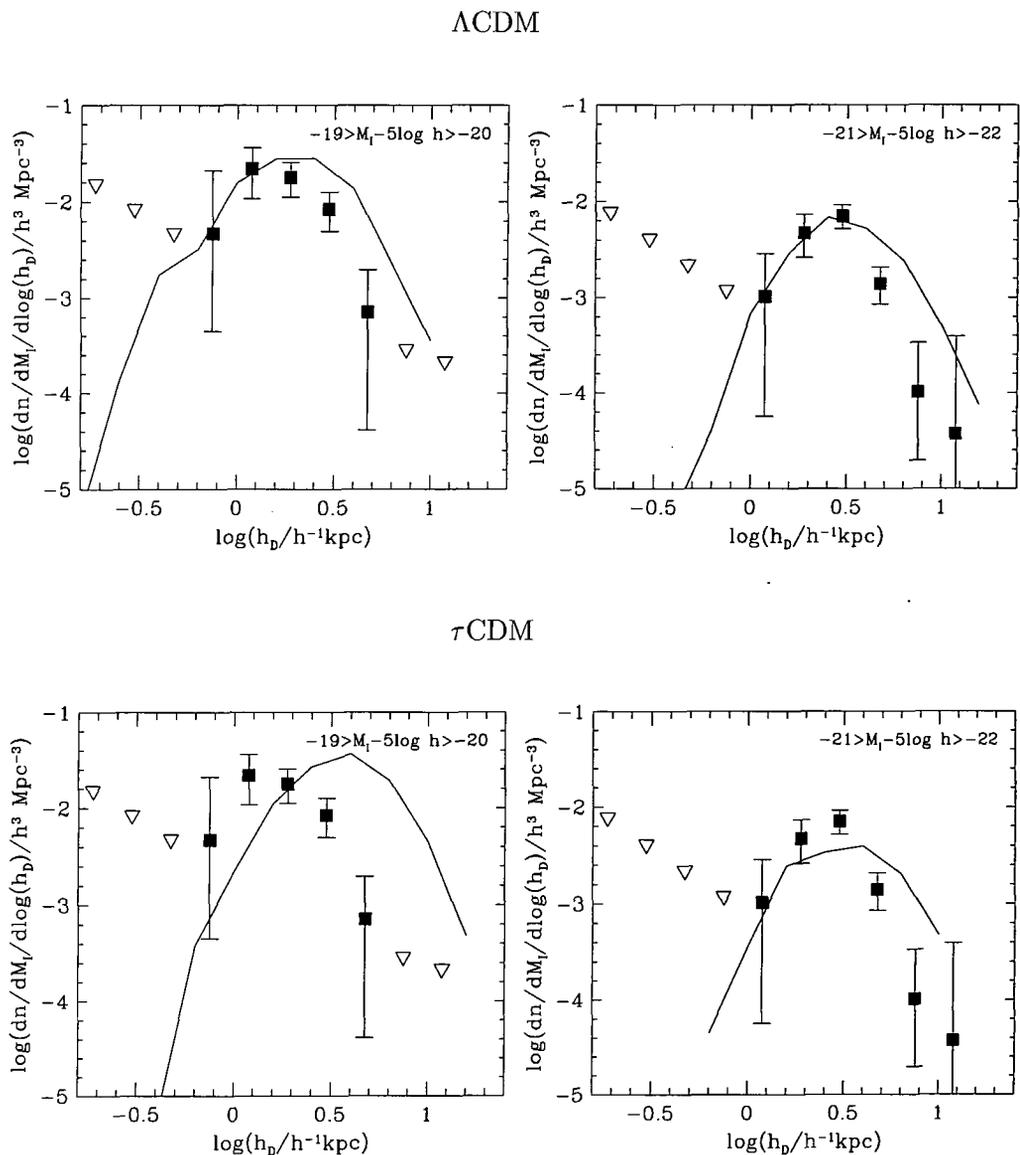


Figure 2.5: The sizes of galaxy disks predicted by the fiducial models compared with the observed distributions of de Jong & Lacey (1999). The two panels for each model show the distribution of exponential scale-lengths, h_D (i.e. the number density as a function of scale-length and magnitude) in luminosity bins straddling L_* . Points with errorbars and triangles are, respectively, the observational data and 95% confidence upper limits for Sb-Sd galaxies, allowing for the dependence of the observational selection function on galaxy size and luminosity. Model galaxies with bulge-to-total ratios in dust-extincted I-band light, $(B/T)_I$, less than 0.33 are shown.

Table 2.2: The percentage of galaxies of each morphological type at $z = 0$ in the two fiducial models.

Model	S	S0	E
Λ CDM	58	7	35
τ CDM	40	11	49

a halo (which determines the angular momentum of the fraction of gas which has been able to cool), and whether the gas conserves its angular momentum as it cools. Whilst the assumptions described in §2.1.3 seem reasonable further investigation is warranted as noted in §2.3.1. Cole et al. (2000) note that beyond these uncertainties the biggest influence on disk sizes is the strength of feedback in the model: if feedback is weak disks form early in small halos and so are themselves small, whilst if feedback is strong disks form later in more massive halos and so are consequently larger. The value of $V_{\text{hot}} = 150$ km/s used in the Λ CDM reference model produces a model disk scale length distribution with a peak close to that in the observational data, and with a width comparable to (though slightly larger than) the data. In the τ CDM model the stronger feedback required to obtain a good match to the luminosity function makes galaxy disks somewhat larger than observed.

Morphology

As described in §2.4.2 bright elliptical galaxies are formed through the violent merger of two disk galaxies. An elliptical galaxy can also be built up gradually via the accretion of lower mass galaxies, as the stars in such galaxies are added to the bulge of the accreting galaxy. Cole et al. (2000) show that the mixture of morphological types in the model is determined by the parameters f_{ellip} and f_{df} which determine, respectively, what fraction of galaxy-galaxy mergers create an elliptical merger remnant and how frequent such mergers are. They also show that the strength of feedback influences the morphological mix, since weak feedback allows disks to form at high redshift giving them a long time in which to undergo mergers and change into ellipticals, whilst strong feedback prevents disks forming until much later, giving them little time in which to merge.

The model does not predict actual morphologies for galaxies, but instead predicts bulge-to-total mass and luminosity ratios. Since there exists a correlation between morphology and bulge-to-total ratio (which admittedly has significant scatter) we will simply

use cuts in bulge-to-total ratio of define morphological types. Ellipticals are defined as those galaxies with more than 60% of their dust-extincted B-band light contributed by the bulge component, whilst spirals are those with less than 40% contributed by the bulge. S0 galaxies occupy the intermediate range.

In Table 2.2 we show the morphological mix in both fiducial models for a volume-limited sample of galaxies brighter than $M_B - 5 \log h = -19.5$. A magnitude limited sample shows a very similar mix. These values can be compared to the morphological mix in the APM Bright Galaxy Survey, which is S+Irr : S0 : E = 67 : 20 : 13 (Loveday 1996, table 10). Here we have lumped together spirals and irregulars and assumed that the 90% of galaxies which were classified in this survey are a representative sample. A compatible result was found in the SSRS2 redshift survey: S + Irr : S0 + E = 66 : 34 (Marzke et al. 1998, table 2).

Considering the crude nature of the morphological classification scheme we have adopted the agreement between this data and the fiducial model results seems reasonable. Cole et al. (2000) note also that since the effects of disk-instabilities are not included in these models (and are likely to alter the morphological mix somewhat, particularly for galaxies fainter than L_*) fine-tuning of the parameters f_{ellip} and f_{df} to match the observational data does not seem warranted.

Cold Gas Content

The cold gas content of spiral galaxies in this model is sensitive to the star formation parameters ϵ_* and α_* (Cole et al. 2000). In Fig. 2.6 we compare the mass of hydrogen in model spiral galaxies with the observational data of Huchtmeier & Richter (1988) and Sage (1993). The Sage (1993) data include measurements of both HI and H₂, and covers a complete sample of galaxies of types Sa-Sd. The Huchtmeier & Richter (1988) data includes only HI measurements, and is for Sa-Im galaxies. Brighter than $M_B - 5 \log h = -16$, Sa-Sc galaxies dominate over Sdm-Im galaxies and so we plot the Sage (1993) data only. Fainter than $M_B - 5 \log h = -16$ the fraction of hydrogen in the form of H₂ appears to be small ($M_{\text{H}_2}/M_{\text{HI}} \lesssim 0.2$), and so here we plot only the Huchtmeier & Richter (1988) data.

For the standard choice of fiducial model parameters the model in both cosmologies is in reasonable agreement with the data. It is interesting to note, as pointed out by Cole et al. (2000), that for a Kennicutt law star-formation timescale (which would correspond to $\alpha_* = 0$ in this model) the model would show a trend of M_{H_2}/L_B vs. M_B in the opposite

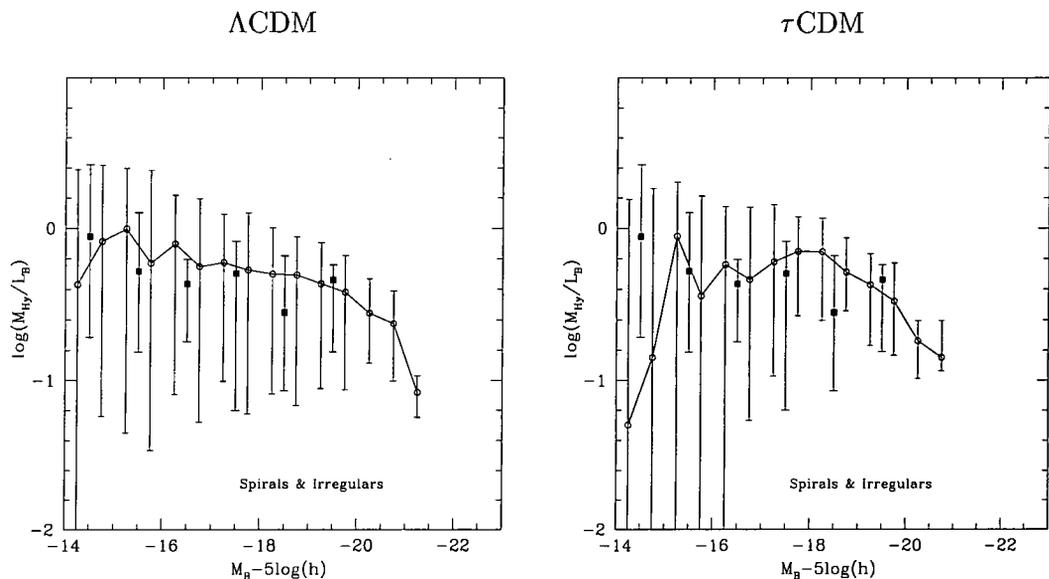


Figure 2.6: The cold gas content of spiral and irregular galaxies as a function of their luminosity. The squares with errorbars show observational data, and indicate the median and 10 and 90 percentiles of the distribution of M_{Hy}/L_B at each luminosity for galaxies of class Sa to Im. $M_{\text{Hy}} = M_{\text{H}_1} + M_{\text{H}_2}$ is the mass of cold hydrogen including both atomic and molecular hydrogen. These points have been estimated from a combination of data from Huchtmeier & Richter (1988) and Sage (1993), as described in the text. Circles with errorbars show results from the models. Galaxies of comparable morphological type are selected by requiring their bulge-to-total ratio in B-band light to be less than 0.4. The model cold gas masses are expressed in the observational units, $h^{-2}M_{\odot}$, and we set $M_{\text{Hy}} = 0.7M_{\text{cold}}$ to account for the presence of helium in the cold gas.

direction to the observed trend.

Metallicity

Chemical evolution in this model is followed using the instantaneous recycling approximation. However, the model differs from the simple “closed-box” scenario since gas can be ejected from a galaxy through the action of supernovae feedback, and fresh gas can flow into a galaxy as it cools from the surrounding halo. As such the effective yield, p_{eff} , varies from galaxy to galaxy through the variation of the feedback parameter β as shown in eqn. (A.87). Model predictions for metallicity scale linearly with p , except for the effect of metallicity on the cooling rate of halo gas.

Both cosmologies are able to match the data quite well. The model dependence of metallicity on absolute magnitude appears somewhat weaker than is observed, but the two are consistent within the errors.

Star Formation History

In Fig. 2.8 we show the star formation history of the two fiducial models, compared to observational data. Solid lines indicate the luminous star formation rate (i.e. brown dwarfs are excluded). Dotted lines show the contribution from quiescent star formation in galaxy disks, whilst the dashed lines indicate the contribution from bursts of star formation triggered by major mergers. The basic form of these curves is generic to CDM cosmologies: the rate initially rises as structure (and so galaxies) form, but eventually begins to decline as the supply of gas needed for star formation is used up. However, it is obvious from Fig. 2.8 that the actual form is strongly dependent on the specific model. For example, in the τ CDM cosmology the drop in star formation rate at high redshift is much more rapid than in the Λ CDM cosmology, due to the much stronger feedback employed in the τ CDM model.

Recent measurements of the star formation rate per unit volume over a large redshift baseline have made this quantity an important test for theories of galaxy formation. However, it is worth noting that measuring the star formation rate is a difficult observational task. For example, the conversion from emission line luminosity to star formation rate depends not only on the assumed IMF and stellar population model, but also on the degree to which the emission line light is attenuated by dust and the extrapolation from the observed part of the star formation rate distribution to the global rate.

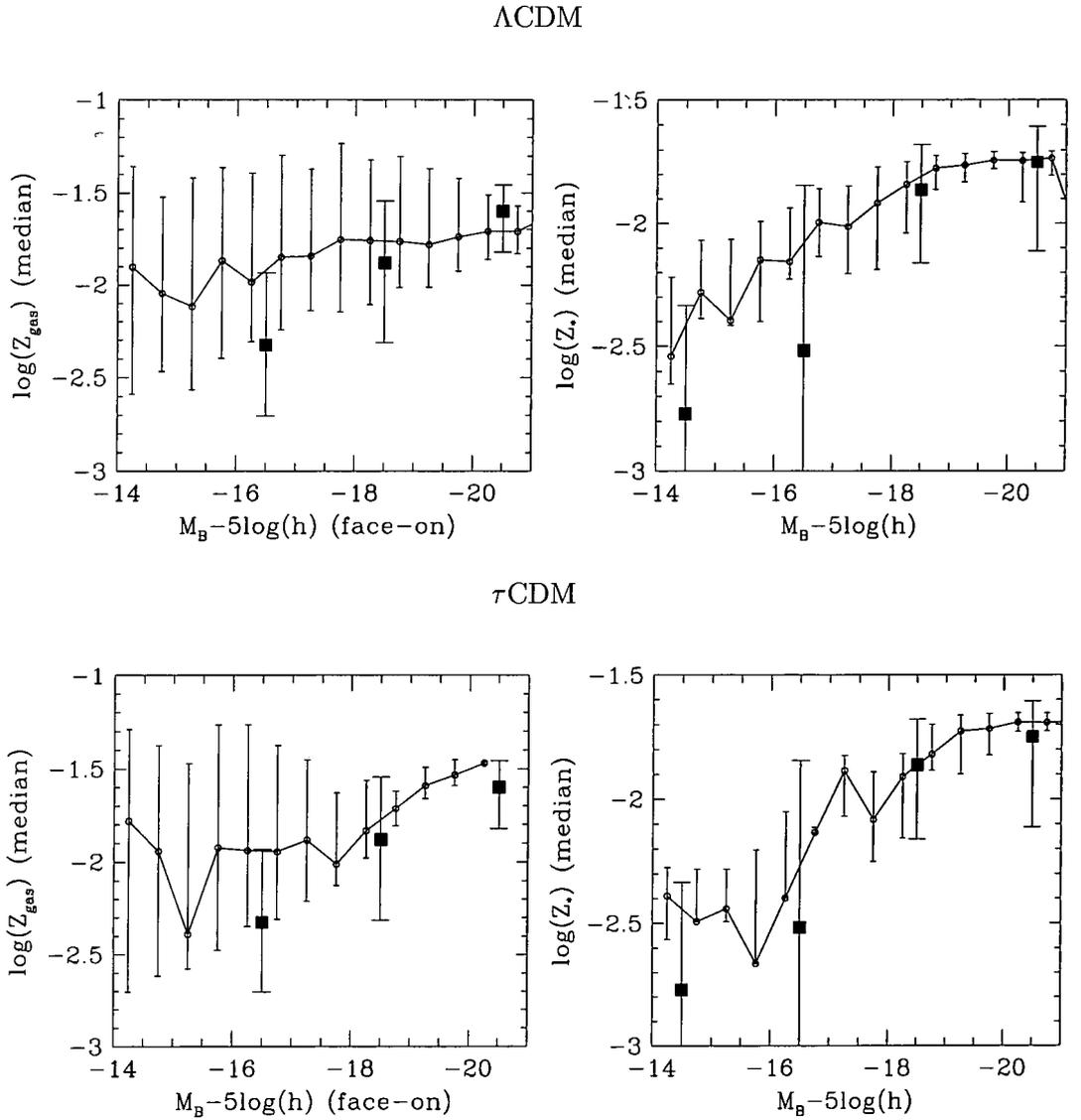


Figure 2.7: The variation in metallicity with luminosity in the fiducial models is compared to observational data. Lines indicate the median metallicity in the model with errorbars showing the 10 and 90 percentiles of the distribution. Panels on the left show the metallicity of the cold star-forming gas in disk-dominated galaxies compared to data for spiral and irregular galaxies, whilst the panels on the right show the stellar metallicities of bulge-dominated galaxies, compared to observations of elliptical galaxies. All observational data was taken from the compilation by Zaritsky, Kennicutt & Huchra (1994).

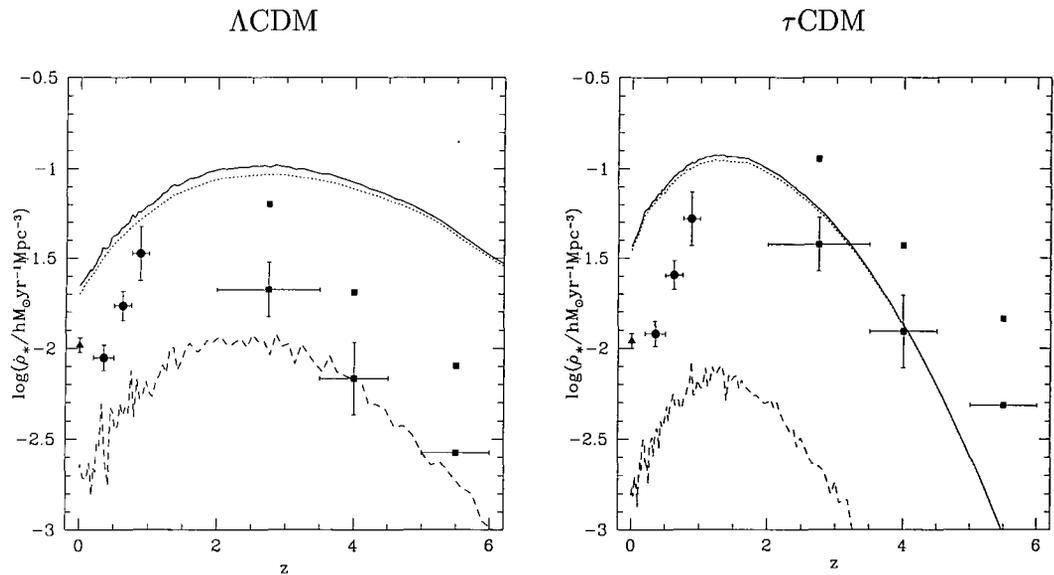


Figure 2.8: The luminous star formation rate (i.e. that excluding brown dwarfs) per unit comoving volume as a function of redshift. In each panel the solid line gives the star formation rate from the fiducial model, whilst the dotted and dashed lines show the quiescent star formation rate (in galaxy disks) and the rate of star formation in bursts respectively. Points with errorbars show observational data from Gallego et al. (1995a,b) (triangle), Lilly et al. (1996) (circles) and Madau et al. (1996) (squares). Points without error bars are raised by a factor of 3 to indicate the likely effects of correcting for the presence of dust. Note also that the data points are different for the two cosmologies due to the differences in the comoving volume-redshift relation, the luminosity distance and the assumed stellar IMF.

2.8 Discussion and Conclusions

We have described the implementation and use of a complete model for galaxy formation in hierarchical universes. Given assumptions about the background cosmology this model is able to predict key physical properties of the galaxy population at any redshift. We have presented results from this model in two different cosmological models and compared these to the available observational data, finding generally good agreement, but highlighting some discrepancies.

In Chapters 6 and 7 we will test one of the key assumptions of this model and compare its predictions to an alternative model of galaxy formation. Firstly, in Chapters 3, 4 and 5 we will employ this model to study the problems of galaxy clustering and the reionization of the Universe.

Chapter 3

The Nature of Galaxy Bias and Clustering

3.1 Introduction

Studies of the clustering of cosmological dark matter have progressed enormously in the past twenty years. The dynamical evolution of the dark matter is driven by gravity and fully specified initial conditions are provided in current cosmological models. This problem can therefore be attacked quite cleanly using N-body simulations (see Gross et al. 1998, Jenkins et al. 1998, and references therein). Studies of the clustering properties of galaxies, on the other hand, are much more complicated because galaxy formation includes complicated and poorly understood astrophysical processes such as gas cooling, star formation and feedback from supernovae. These processes couple with the gravitational evolution of the dark matter to produce the clustering pattern of galaxies. Because of this complexity, progress in understanding galaxy clustering has been slow. Yet, theoretical modelling of galaxy clustering is essential if we are to make the most of the new generation of galaxy redshift surveys, the Two-degree Field (2dF, Colless 1996) and Sloan Digital Sky Survey (SDSS, Gunn & Weinberg 1995), and of the new data on galaxy clustering at high redshift that has been accumulating recently (e.g. Adelberger et al. 1998, Giavalisco et al. 1998)

Two kinds of simulation techniques are being used to approach galaxy clustering from a theoretical standpoint. The first of these attempts to follow galaxy formation by simulating directly dark matter and gas physics in cosmological volumes (e.g. Katz, Hernquist & Weinberg 1992, Evrard, Summers & Davis 1994, Pearce et al. 1999, Blanton et al. 2000). Because the resolution of such simulations is limited, phenomenological models are required to decide when and where stars and galaxies form and to include the associated feedback effects. The advantage of this approach is that the dynamics of cooling gas are calculated approximately without the need for simplifying assumptions.

The disadvantage is that even with the best codes and fastest computers available, the attainable resolution is still some orders of magnitude below what is required to resolve the formation and internal structure of individual galaxies in cosmological volumes. For example, the gas resolution element in the large Eulerian simulations of Blanton et al. (2000) is around half a megaparsec. Lagrangian hydrodynamic methods offer better resolution, but even in this case, this is poorer than the galactic scales on which much of the relevant astrophysical processes occur.

A different and complementary approach to studying galaxy clustering is to use a semi-analytic model of galaxy formation such as that described in Chapter 2. In this case, resolution is generally not a major issue. The disadvantage of this technique, compared to hydrodynamic simulations, is that, in calculating the dynamics of cooling gas, a number of simplifying assumptions, such as spherical symmetry or a particular flow structure, need to be made (some of these assumptions are tested against smoothed particle hydrodynamics simulations in Chapter 7). As in the direct simulation approach, a model for star formation and feedback is required. In addition to adequate resolution, semi-analytic modelling offers a number of advantages for studying galaxy clustering. Firstly, it is a much more flexible approach than full hydrodynamic simulation and so the effects of varying assumptions or parameter choices can be readily investigated. Secondly, with detailed semi-analytic modelling it is possible to calculate a wide range of galaxy properties such as luminosities in any particular waveband, sizes, bulge-to-disk luminosity and mass ratios, masses, circular velocities, etc. This makes it possible to construct mock catalogues of galaxies that mimic the selection criteria of real surveys and to investigate clustering properties as a function of magnitude, colour, morphological type or any other property determined by the model.

Semi-analytic modelling has been used in two different modalities to study galaxy clustering. In the first approach, the semi-analytic model is used to populate a grid of halo masses drawn from the Press-Schechter halo distribution, with galaxies, and the galaxy correlation function is computed using the analytic model of Mo & White (1996). This technique has been extensively applied by Baugh et al. (1998b, 1999), and is essentially unlimited in the halo mass resolution that it can reach. In the second, more direct, approach, the semi-analytic modelling is applied to dark matter halos grown in a cosmological N-body simulation. The advantages of this latter strategy are that it allows a proper treatment of the small scale regime where the Mo & White model breaks down and it bypasses any inaccuracies in the analytic (Press-Schechter) model used to

compute the mass function of dark halos in the pure semi-analytic approach. This technique has been implemented in two ways. In the simplest case (Kauffmann, Nusser & Steinmetz 1997, Roukema et al. 1997, Governato et al. 1998), a merger tree for each halo identified in the N-body simulation is generated in a Monte-Carlo manner. In the second implementation (Kauffmann et al. 1999a,b, Diaferio et al. 1999), the halo merger trees are extracted directly from the N-body simulation.

In this Chapter, we adopt the first approach to the combined use of semi-analytic and N-body techniques (i.e. with Monte-Carlo merger trees) to study galaxy clustering. We focus on the specific question of how the process of galaxy formation couples with the large scale dynamics of the dark matter to establish the clustering properties of the galaxy population. We investigate in detail processes that bias galaxies to form preferentially in certain regions of space. Previous cosmological dark matter simulations have established that the dark matter in popular CDM models tends to be more strongly clustered on small scales than the observed galaxy population (Gross et al. 1998, Jenkins et al. 1998). We investigate whether the required antibias arises naturally in these cosmologies. More generally, we compare the predictions of these models with observations over a range of scales. The techniques that we use are described in §3.2. The clustering properties of galaxies in this model are presented in §3.3. The various processes that play a role in determining how the galaxy distribution is biased relative to the mass are discussed in §3.4. In §3.5, we show that our results are robust to changes in model parameters and finally in §3.6 we discuss our main conclusions.

3.2 Description of the Model

The two techniques that we employ in this Chapter, N-body simulations and semi-analytic modelling, are both well established and powerful theoretical tools. The semi-analytic model is described in detail in Chapter 2. We do not intend to describe the techniques of N-body simulations in detail here, but instead refer the reader to the appropriate sources.

3.2.1 Semi-Analytic Models

In this Chapter we extend this list of predictions made by the semi-analytic model of Chapter 2 to include the spatial clustering of galaxies, in particular the two-point correlation function. Whilst these models are similar in principle to those of Kauffmann et al. (1999a) they use different rules for star formation and feedback and also include the

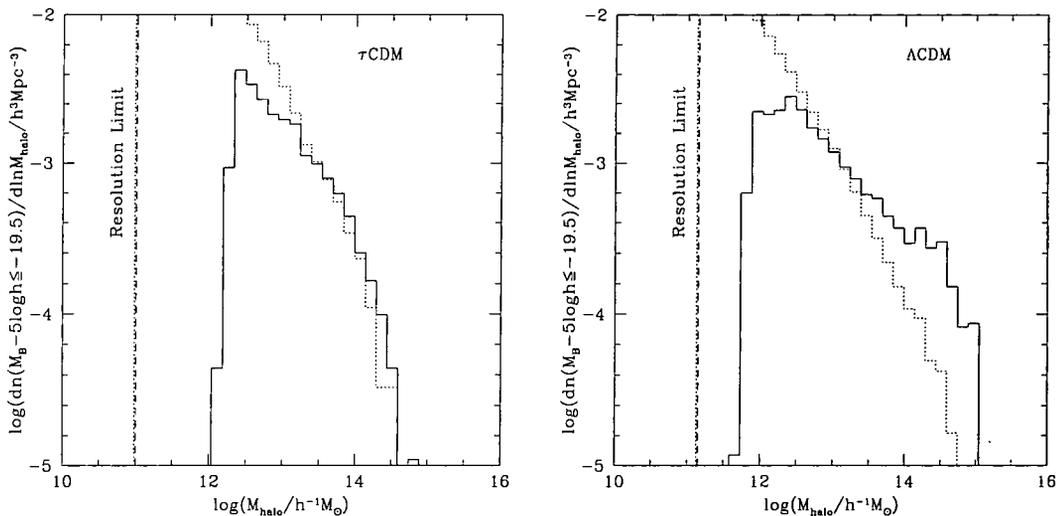


Figure 3.1: Mass function of halos containing galaxies with $M_B - 5 \log h \leq -19.5$ in our τ CDM (left hand panel) and Λ CDM (right hand panel) models. Mass functions weighted by number of galaxies are shown by the solid lines while unweighted mass functions are shown by dotted lines. These halos are well above the resolution limit (equal to the mass of a group of 10 particles in each simulation).

effects of chemical enrichment due to star formation. We also choose to constrain the model parameters in a rather different way, as discussed below.

3.2.2 Incorporation into N-body Simulations

We first locate dark matter halos within the simulation volume by use of a group finding algorithm. This provides a list of approximately virialised objects within the simulation. For each such halo we determine the position and velocity of the centre of mass and also record the positions and velocities of a random sample of particles within the halo. The list of halo masses from the simulation is fed into the semi-analytic model of galaxy formation in order to produce a population of galaxies associated with each halo. The merger tree for each dark matter halo is generated using a Monte-Carlo method, as opposed to being extracted directly from the N-body simulation (as was done by Kauffmann et al. 1999a).

Each galaxy is assigned a position and velocity within its halo. Since the semi-analytic model distinguishes between central and satellite galaxies, we locate the central galaxy at the centre of mass of the halo and assign it the velocity of the centre of mass. Any satellite galaxies are located on one of the randomly selected halo particles and are assigned the velocity of that particle. In this way, by construction, satellite galaxies always trace the density and velocity profile of the dark matter halo in which they reside.

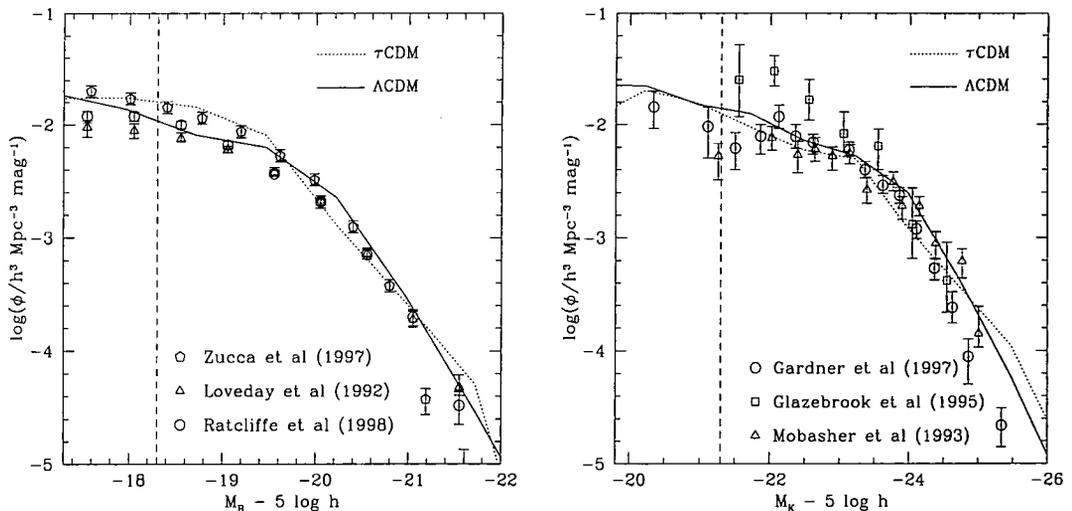


Figure 3.2: B and K band luminosity functions for the τ CDM (dotted line) and Λ CDM (solid line) reference models. Points with error bars show a selection of observational determinations of the luminosity functions. The luminosity functions are shown only as faint as the τ CDM resolution limit and the vertical dashed line shows the resolution limit for Λ CDM models.

Once galaxies have been generated and assigned positions and velocities within the simulation it is a simple process to produce catalogues of galaxies with any desired selection criteria (e.g. magnitude limit, colour, etc.) complete with spatial information or, equally simply, with redshift space positions to enable the study of redshift space distortions (see Chapter 4).

3.2.3 Reference Models

We have made use of the “GIF” simulations carried out by the Virgo Consortium. These are high resolution simulations of cosmological volumes of dark matter carried out in four different cosmologies: τ CDM and Λ CDM (which are used as our reference models), SCDM and OCDM (which we consider briefly in §3.5). These models are described in detail by Jenkins et al. (1998) and the simulations are described by Kauffmann et al. (1999a). Briefly, the simulations model boxes of order $100 h^{-1}$ Mpc in size with nearly 17 million particles, each of mass approximately $10^{10} h^{-1} M_{\odot}$. The critical density models, SCDM and τ CDM, have $h = 0.5$ and spectral shape parameter $\Gamma = 0.5$ and 0.21 respectively, whilst the low density models, Λ CDM and OCDM, have $h = 0.7$, $\Omega_0 = 0.3$ and $\Gamma = 0.21$. The Λ CDM model is made to have a flat geometry by inclusion of a cosmological constant. All the models are normalised to produce the observed abundance of rich clusters today

(Eke, Cole & Frenk 1996, Viana & Liddle 1996). Dark matter halos were identified using the “Friends-of-Friends” algorithm (Davis et al. 1985) with a linking length of $b = 0.2$; only halos containing 10 or more particles are considered. The ability to resolve halos of this mass allows us to determine the properties of galaxies up to one magnitude fainter than L_* .

We construct two reference semi-analytic models with the same cosmological parameters as the corresponding GIF simulations. The smallest halo that can be resolved in the N-body simulation determines the faintest galaxies for which the model catalogues are complete. We consider only galaxies brighter than $M_B - 5 \log h = -19.5$ and we have checked, in each case, that the model is complete to this magnitude. A model is complete if the lowest mass halo which can contain a galaxy of interest is above the group resolution limit in the simulation (which is 10 times the particle mass). Fig. 3.1 displays the halo mass functions for galaxies brighter than $M_B - 5 \log h = -19.5$ in the two reference models and shows that these two models are complete to this magnitude limit. The minimum mass halo occupied by galaxies is $5.3 \times 10^{11} h^{-1} M_\odot$ in Λ CDM and $1.5 \times 10^{12} h^{-1} M_\odot$ in τ CDM. The faintest galaxies which are fully resolved in the semi-analytic models have $M_B - 5 \log h \approx -18.3$, $M_K - 5 \log h \approx -21.3$ and $M_B - 5 \log h \approx -17.3$, $M_K - 5 \log h \approx -19.8$ in Λ CDM and τ CDM respectively. When varying model parameters we have checked that the galaxy samples remain complete.

The τ CDM and Λ CDM models both reproduce the local B and K-band luminosity functions reasonably well, including the exponential cut-off at bright magnitudes, as shown in Fig. 3.2. The Λ CDM model also produces a close fit to the I-band Tully-Fisher relation constructed using the circular velocities of the dark matter halos in which they formed, as may be seen in Fig. 3.4. (When the circular velocities of the galaxies themselves are used instead, the model velocities are about 30% too large; see §2.7.2 and Cole et al. (2000) for a full discussion.) In contrast the τ CDM model misses the Tully-Fisher zero-point by nearly 1 magnitude. (The model Tully-Fisher relations plotted in Fig. 3.4 are for galaxies selected as described in §2.7.2.)

The Λ CDM model is similar to the fiducial model of Chapter 2. In both cases, the model parameters were chosen so as to obtain a reasonable match to a subset of local data, most notably the galaxy luminosity function. The model used in this Chapter was selected before the fiducial model of Chapter 2 (which is very similar to that of Cole et al. 2000) had been fully specified and so there are small differences in the values of some of the parameters in the two models. These differences are immaterial for our present

purposes. We show in Figure 3.3 galaxy two-point correlation functions from the reference model of this Chapter (solid line) and from the fiducial model of Chapter 2 (dashed line). Dotted lines indicate the Poisson errors on the solid curve. For the fiducial model of Chapter 2 the semi-analytic galaxies were placed into an N-body simulation identical to the GIF Λ CDM simulation except that the cosmological parameters σ_8 and Γ were set to be 0.93 and 0.19 respectively in order to match the parameters of the fiducial model. Once the difference in σ_8 between the two models is taken into account the two galaxy correlation functions differ by an amount entirely comparable to the scatter seen in Fig. 3.14 (for models which are good fits to the luminosity function).

All semi-analytic models considered in this Chapter include the effects of dust on galaxy luminosities calculated using the models of Ferrara et al. (1999), unless otherwise noted. The model parameters of the reference models used in this Chapter are listed in Table 3.1. The role of each, and the way in which these parameters are constrained by a set of observations of the local Universe, are discussed in detail in Chapter 2 and by Cole et al. (2000). Note also that these models used a somewhat smaller core radius for the hot gas in halos, with r_{core} being initially set to 0.1 times r_{NFW} .

As noted in Table 3.1 an artificially low value of f_{df} is required in the τ CDM model in order to obtain a good fit to the local B and K-band luminosity functions. The rapid galaxy merger rate that results from this choice will deplete the number of galaxies living in high mass halos, and so may affect the correlation function of galaxies. However, in §3.5.1 we show that altering this parameter produces no significant change in the model correlation function.

3.3 Clustering of Galaxies

3.3.1 The Galaxy Two-Point Correlation Function

The evolution of dark matter in the linear regime is well understood analytically (Appendix A.1.2), and can be followed into the non-linear regime using N-body simulations (e.g. Jenkins et al. 1998), or theoretically inspired model fits to the simulation results (Hamilton et al. 1991, Peacock & Dodds 1996). The case for galaxies is very different. Galaxies are generally believed to form near regions of high density (as in the heuristic “peaks bias” model of galaxy formation; see, for example, Bardeen et al. 1986). If galaxies formed only in halos with masses greater than the characteristic clustering mass, M_* (the mass for which the r.m.s. density fluctuation in the Universe equals the critical over-

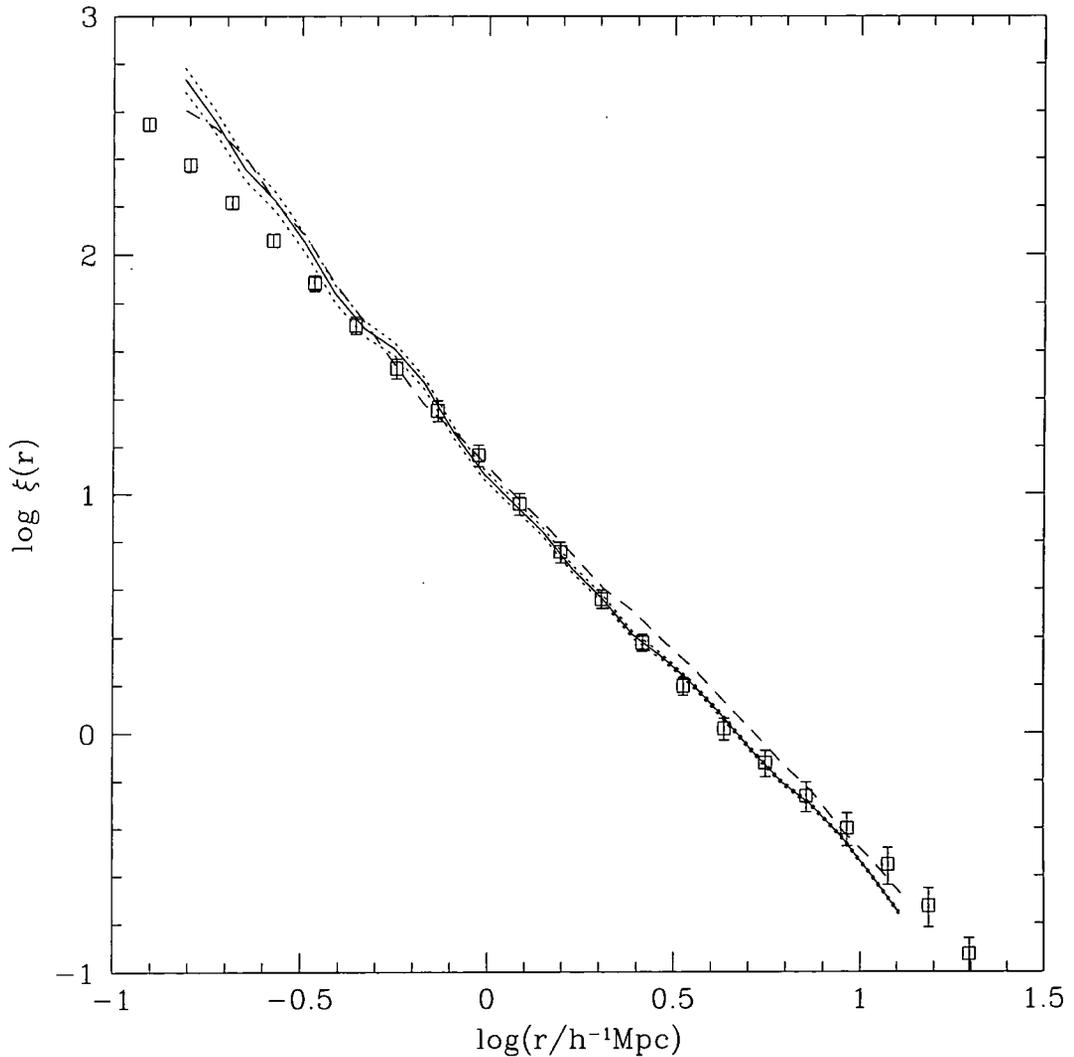


Figure 3.3: Two-point correlation functions for galaxies brighter than $M_B - 5 \log h = -19.5$ in a Λ CDM cosmology. The solid line indicates the correlation function of galaxies in the model used in this Chapter whilst the dashed line indicates that of galaxies in the fiducial model of Chapter 2. Dotted lines indicate Poisson errors for the solid curve (errors for the dashed curve are comparable). Open squares indicate the observed APM real-space correlation function (Baugh 1996).

Table 3.1: The parameters of our two reference models, using the notation described in the text and described in detail in Chapter 2.

Model	Ω_0	Λ_0	Ω_b	h	Γ	σ_8	ϵ_*	α_*	V_{hot}	α_{hot}	e	f_{ellip}	f_{df}	IMF	p	R	Υ
Λ CDM	0.3	0.7	0.02	0.7	0.21	0.90	0.0100	-0.5	150.0	2.0	0.0	0.3	1.0	Kennicutt (1983)	0.02	0.41	1.63
τ CDM	1.0	0.0	0.08	0.5	0.21	0.60	0.0200	-0.5	300.0	2.0	0.0	0.3	0.1 [†]	Salpeter (1955)	0.04	0.28	1.23

[†] As noted in Chapter 2, f_{df} should be approximately 1 or larger. Here we use an artificially low value in order to obtain a good fit to the local B and K-band luminosity functions for the τ CDM model.

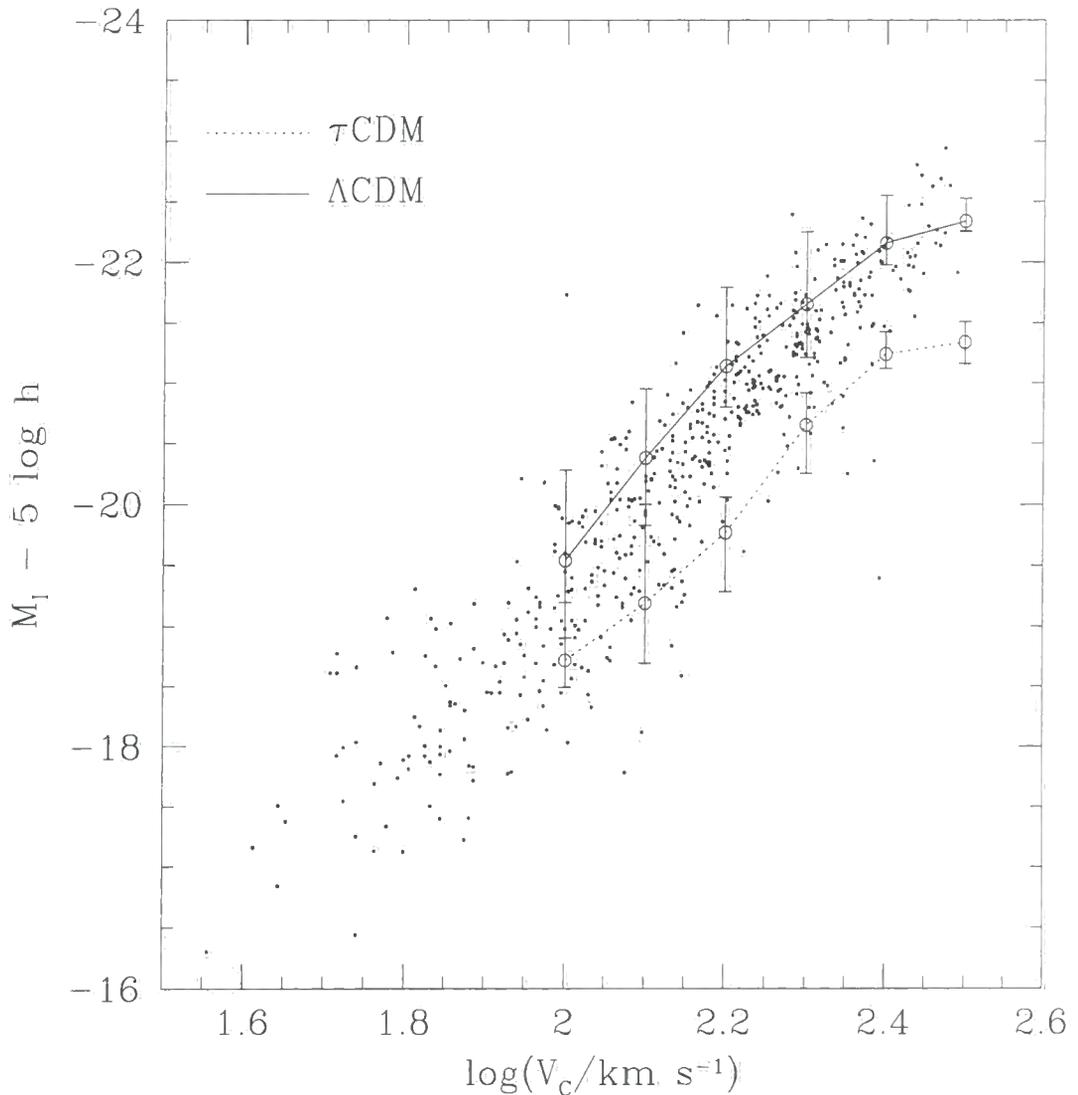


Figure 3.4: Tully-Fisher relations in the τ CDM (dotted line) and Λ CDM (solid line) reference models. These models are constrained by the luminosity function. Points are the observational data of Mathewson, Ford & Buchhorn (1992). Each line is plotted using the circular velocity of each galaxy's dark matter halo and indicates the median of the distribution, while the error bars indicate the 10% and 90% intervals. Galaxies are selected by their bulge-to-total ratio in dust-extinguished I-band light, which is required to be in the range 0.02 to 0.24, and are required to have at least 10% of the mass of their disk in the form of cold gas.

density for collapse in the spherical top-hat model), at birth they would be biased with respect to the dark matter. However, galaxy formation is an ongoing process occurring in a range of halo masses, so any initial bias will evolve with time.

Several authors (Davis et al. 1985, Bagla 1998, Tegmark & Peebles 1998) have shown that if galaxies could be assigned permanent tags at birth, then their correlation function would approach that of the dark matter at late times because the clustering due to gravitational instability eventually becomes much greater than that due to the initial formation sites of galaxies. They show that this is true even in simple, continuous models of galaxy formation.

However, the Universe is more complex than this. It is difficult, if not impossible, to assign a permanent tag to a galaxy since galaxies evolve and sometimes merge. Therefore, as we look to higher redshifts, it is unlikely that we will be observing the same population of galaxies that we see at $z = 0$. For example, in a survey with a fixed apparent magnitude limit we should expect to see the galaxy correlation function initially decreasing to higher z , as the characteristic clustering mass decreases. Eventually, however, the correlation function should begin to rise as the apparent magnitude limit selects only the brightest and most massive galaxies at high redshift which are intrinsically more clustered than the average galaxy. These points have been discussed in detail by Kauffmann et al. (1999a) and Baugh et al. (1999). Thus, the apparent evolution of the galaxy clustering pattern depends on the internal evolution of the galaxies themselves as well as on the variation of their positions with time. In the semi-analytic model both of these forms of evolution are explicitly included.

The results of the techniques described in the previous section are shown in Fig. 3.5 & 3.6. Fig. 3.5 shows slices through the GIF Λ CDM and τ CDM dark matter simulations on which we have overlaid the positions of galaxies from the models. The galaxies can be seen to trace out structure in the dark matter and to avoid the underdense regions in the dark matter distribution. The galaxies clearly follow the large scale structure of the dark matter, but as we will show, they are biased tracers of the mass. The most obvious difference between the two diagrams is the smaller number of galaxies in the τ CDM model. This is simply due to the smaller volume of the τ CDM slice (approximately five times smaller than the Λ CDM slice), since the number of galaxies per unit volume is constrained to be very similar in each model by the requirement that they match the observed luminosity function.

Fig. 3.6 shows the two-point correlation function of the model galaxies, and compares

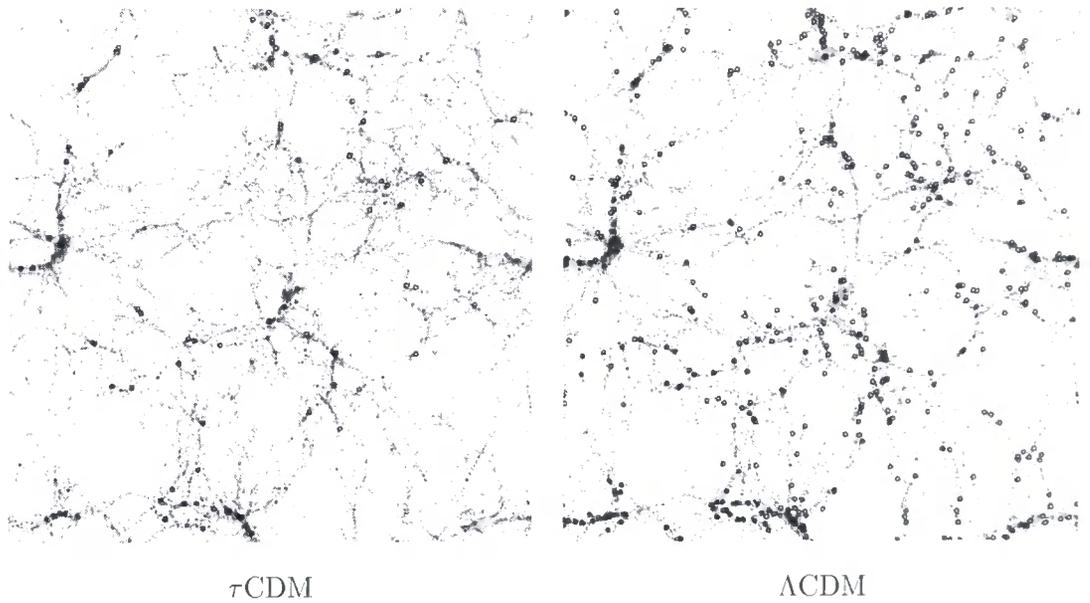


Figure 3.5: The left hand panel shows the locations of galaxies brighter than $M_B - 5 \log h = -19.5$ in a τ CDM model. The figure shows a slice of the dark matter simulation $85 \times 85 \times 4.7h^{-3} \text{ Mpc}^3$ in size. The density of dark matter is indicated by the greyscale (with the densest regions being the darkest). Overlaid are the positions of the galaxies, indicated by open circles. The right hand panel shows the equivalent slice from a Λ CDM model (the initial density fields of the GIF simulations all have the same Fourier mode phases, hence the similarity of the structure), the slice in this case being $141 \times 141 \times 8h^{-3} \text{ Mpc}^3$ in size.

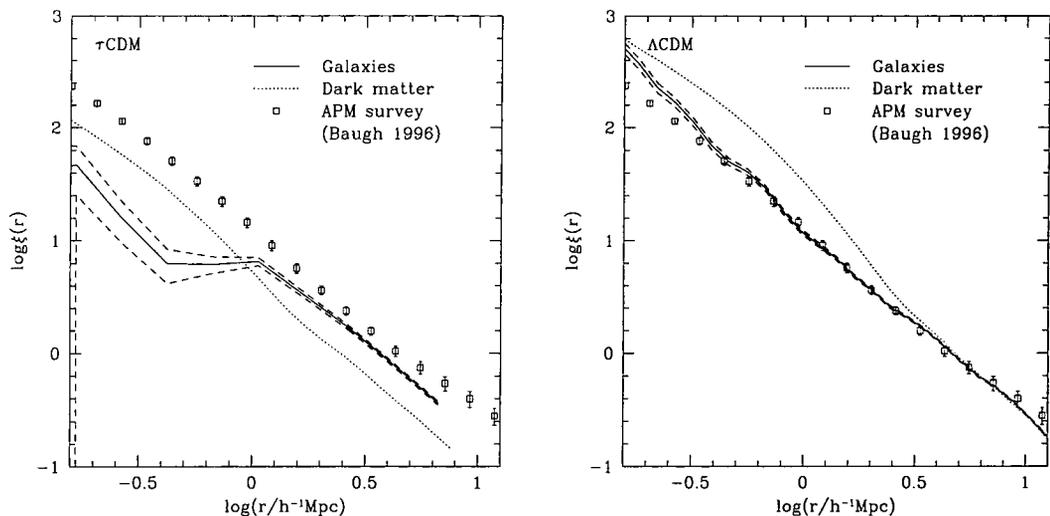


Figure 3.6: The left hand panel shows the two-point correlation function of galaxies brighter than $M_B - 5 \log h = -19.5$ in the τCDM model as a solid line. The dashed lines to either side indicate the Poisson sampling errors. This is compared to the observed APM real-space correlation function (points with error bars) and to the mass correlation functions in the N-body simulations (dotted line). The right hand panel shows the equivalent plot for the ΛCDM model.

them to the observed APM correlation function (in real-space) and to the correlation function of the underlying dark matter. The two models show distinct differences in their behaviour. Most obviously, the ΛCDM model is very close to the APM data from $r \approx 0.3h^{-1}$ Mpc to $r \approx 10h^{-1}$ Mpc, whilst the τCDM model fails to achieve a large enough amplitude on scales $\geq 1.0h^{-1}$ Mpc and drops even further below the observed correlation function on smaller scales. The τCDM model shows a strong bias on large scales. The bias parameter, defined as the square root of the ratio of the mass and galaxy correlation functions, is approximately 1.4. The ΛCDM model, on the other hand, is essentially unbiased on large scales. Both models show an anti-bias on smaller scales. It is interesting to note that the galaxy correlation functions do not display the same features as the dark matter. For example the shoulder in the ΛCDM dark matter correlation function at $3 h^{-1}$ Mpc is not present in the galaxy correlation function. Instead, the latter is remarkably close to a power-law form over about four orders of magnitude in amplitude.

3.3.2 Systematic Effects

In this section we consider two systematic effects which may affect the clustering properties of galaxies in these models: dynamical friction in groups and clusters and our procedure for constructing merger trees for the dark matter halos. We show that neither of these significantly affects the two-point correlation function.

Dynamical Friction

These models do not accurately account for the effects of dynamical friction on the spatial position of satellite galaxies in halos. The simulations lack the resolution to follow this process directly. They can, however, correctly model the two extremes of this effect. If the dynamical friction timescale is much longer than the age of the halo, then the galaxy orbit is close to its original orbit, and so our placement scheme, consisting of identifying galaxies with randomly chosen halo particles, is correct on average. Conversely, if the dynamical friction timescale were much shorter than the halo lifetime, the satellite galaxy would have sunk to the bottom of the halo potential well and merged with the central galaxy. This effect is included in the semi-analytic model. Therefore, it is only in the intermediate range where the dynamical friction timescale is of the same order as the halo lifetime that these models do not accurately reproduce the galaxy positions within clusters.

To estimate the effect of dynamical friction on the correlation function we have tried perturbing the galaxy positions using the following simple model. From the calculation of the dynamical friction timescale in an isothermal halo given in §2.4.1, and specifically eqn. (A.106), it can be seen that the orbital radius of a galaxy in a circular orbit, r , decays with time as

$$r = r_i \sqrt{\frac{\tau_{\text{mrg}} - t}{\tau_{\text{mrg}}}}, \quad (3.1)$$

where r_i is the initial orbital radius of the galaxy when the halo forms, at $t = 0$, and τ_{mrg} is the dynamical friction timescale of the galaxy, given by eqn. (2.35). Here, to mimic this behaviour, each satellite galaxy is first assigned a position in the halo tracing the dark matter as before and then its distance from the halo centre is reduced by a factor r/r_i , which we can compute as we know both τ_{mrg} and t for each galaxy.

Fig. 3.7 shows the correlation function in the Λ CDM reference model with and without this dynamical friction effect included. Dynamical friction causes only a slight increase in the clustering amplitude on small scales, $< 0.5h^{-1}$ Mpc, (since galaxies are drawn closer

together inside halos). However the effect is small and can be safely neglected.

Merger Tree Construction

As noted in §3.1, one difference between this work and that of Kauffmann et al. (1999a) is that they extract merger trees for dark matter halos directly from the N-body simulation, whereas we extract the final mass of the halo and generate the merger tree using a Monte-Carlo technique based on the extended Press-Schechter formalism. There are advantages to both techniques. Extracting the halo trees from the simulation circumvents any possible discrepancy between the extended Press-Schechter predictions and the merging histories in the N-body simulation, although it has been shown that the two are statistically similar (see, for example, Lacey & Cole 1994, Somerville et al. 1998, Lemson & Kauffmann 1999, and Chapter 7). In particular, Lemson & Kauffmann (1999) have studied the statistical properties of halo formation histories in N-body simulations and find no detectable dependence of formation history on environment, as expected in the Press-Schechter theory. Thus, the fact that we construct merger trees similarly for halos in high and low-density regions should make little or no difference to our results. Furthermore, since here we are only interested in the statistical properties of the galaxy population, our approach is justified.

One drawback of the direct extraction technique is that the merger trees become limited by the resolution of the simulation. Like ourselves, Kauffmann et al. identified halos containing at least 10 particles. Since this mass resolution limit applies at all times in the simulation, such a halo cannot have been formed by merging, as it might have done in a higher resolution simulation. Furthermore, even large mass halos might have significantly modified merging histories due to this artificial resolution limit. The analytic merging trees that we generate do not suffer from this problem. The effective mass and time resolutions can be made as small as desired, until convergence is reached. We demonstrate the effects of the resolution limit by considering a Λ CDM model in which we artificially impose an effective mass resolution equivalent to that in the Kauffmann et al. models. Fig. 3.8 shows that the differences between the models with and without the artificial mass resolution limit are, in general insignificant, although there is a region (approximately from separations of 0.4 to 1.5 h^{-1} Mpc) where the disagreement between the two is significant.

Finally, it should be noted that since we extract the final masses of halos from the N-body simulation, our models do not suffer from the well-documented (but small) dif-

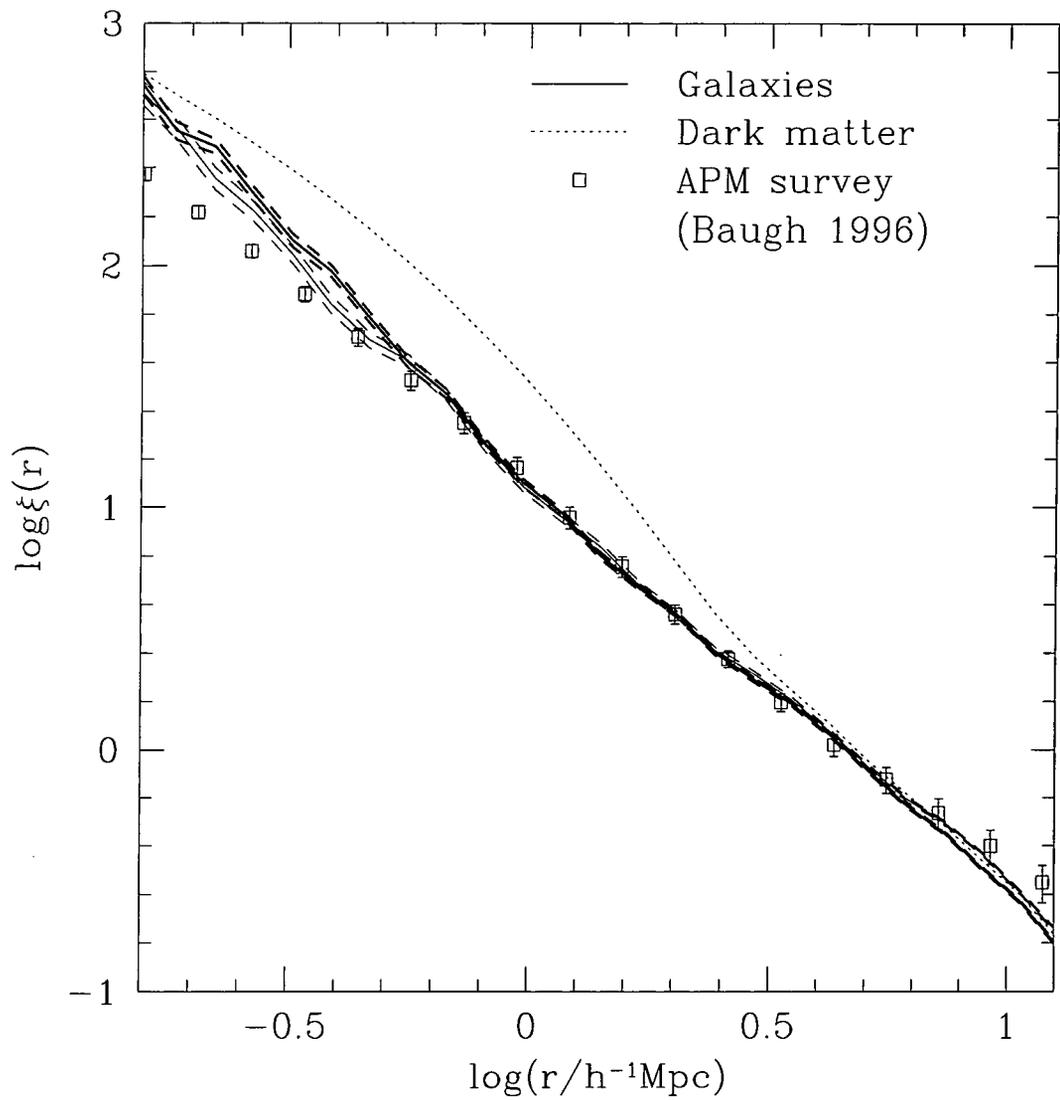


Figure 3.7: The galaxy correlation function in the Λ CDM reference model, with and without the effects of dynamical friction on satellite galaxy positions. The thin solid line shows the standard model (the dashed lines indicating the Poisson errors), whilst the thick lines show the same model with an estimate of dynamical friction effects included.

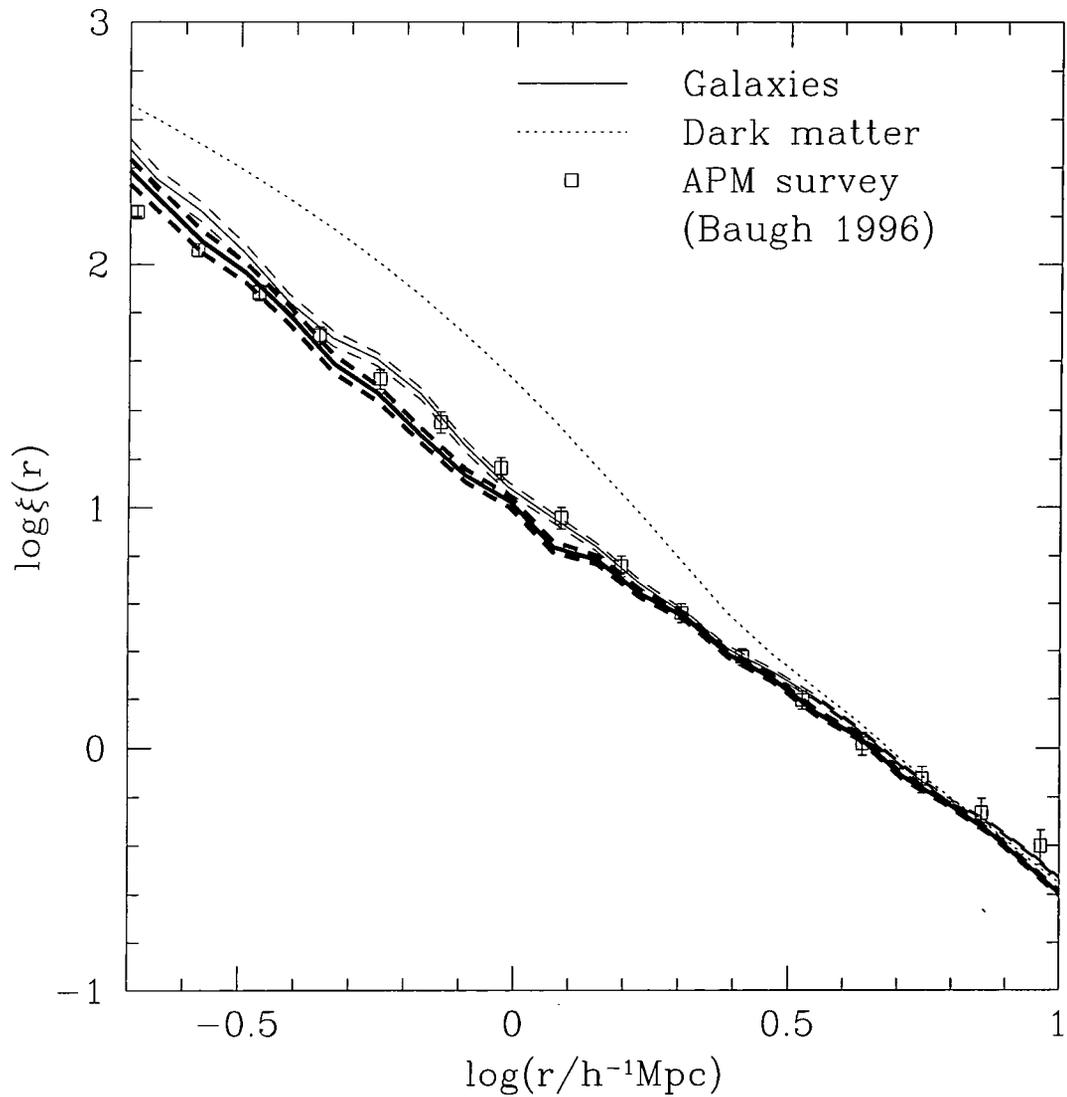


Figure 3.8: A comparison of correlation functions for galaxies brighter than $M_B - 5 \log h = -19.5$ in the Λ CDM reference model with the results of a model with an artificial mass resolution designed to mimic the models of Kauffmann et al. (1999a). The low resolution model is shown as a thick solid line whilst our reference model is shown by thin solid lines. The dashed lines indicate the Poisson sampling errors.

ferences between the Press-Schechter and N-body mass functions at low mass (e.g. below $\sim 10^{14}h^{-1}M_{\odot}$), discussed by Efstathiou et al. (1988), Lacey & Cole (1994) and Somerville et al. (1998).

3.4 The Nature of Bias

In the models explored here, galaxies do not trace the mass exactly because galaxy formation proceeds with an efficiency which depends on halo mass. In the lowest mass halos, feedback from supernovae prevents efficient galaxy formation, whilst in the high mass halos, gas is unable to cool efficiently by the present day thereby inhibiting galaxy formation. These effects can be seen in the mass-to-light ratios (in the B-band) of halos in the reference models plotted in Fig. 3.9. The mass-to-light ratio is strongly dependent on halo mass. Initially, it decreases as halo mass increases, before turning upwards and levelling off at close to the universal value for the highest mass halos in the simulations. The minimum, at around $10^{12}h^{-1}M_{\odot}$, marks a preferred mass scale at which the efficiency of galaxy formation is greatest. The mass-to-light ratio varies by more than a factor of 3 over the range of masses plotted here. As a result of this varying mass-to-light ratio we expect a complex, scale-dependent bias to arise and this is, in fact, seen in the two reference models. The clustering of galaxies is controlled by the intrinsic bias of their host halos, the non-linear dynamics of the dark matter and the processes of galaxy formation.

Fig. 3.10 shows the mean number of galaxies per halo as a function of halo mass in the two models. (For future reference we also plot the mean number of pairs of galaxies per halo as defined by eqn. (3.3) below.) Below the $10^{12.5}h^{-1}M_{\odot}$ and $10^{13}h^{-1}M_{\odot}$ bins in the τ CDM and Λ CDM models respectively, halos always contain zero or one galaxy (i.e. the number of pairs is zero). This is simply because there is not enough cold gas in the halo to form two or more galaxies of the required luminosity by the present day. At higher halo masses there is a trend of increasing number of galaxies per halo. The average occupation increases less rapidly than the halo mass, indicating once again that the halo mass-to-light ratio increases with increasing mass.

Galaxies brighter than some given absolute magnitude only form in halos above a certain mass, M_h . On scales much larger than the radii of these halos, the correlation function of these galaxies will be proportional to that of the dark matter, with some constant, asymptotic, large-scale bias, as has been shown by Mo & White (1996). Behaviour of this type is seen in both of the reference models. This large scale bias can be

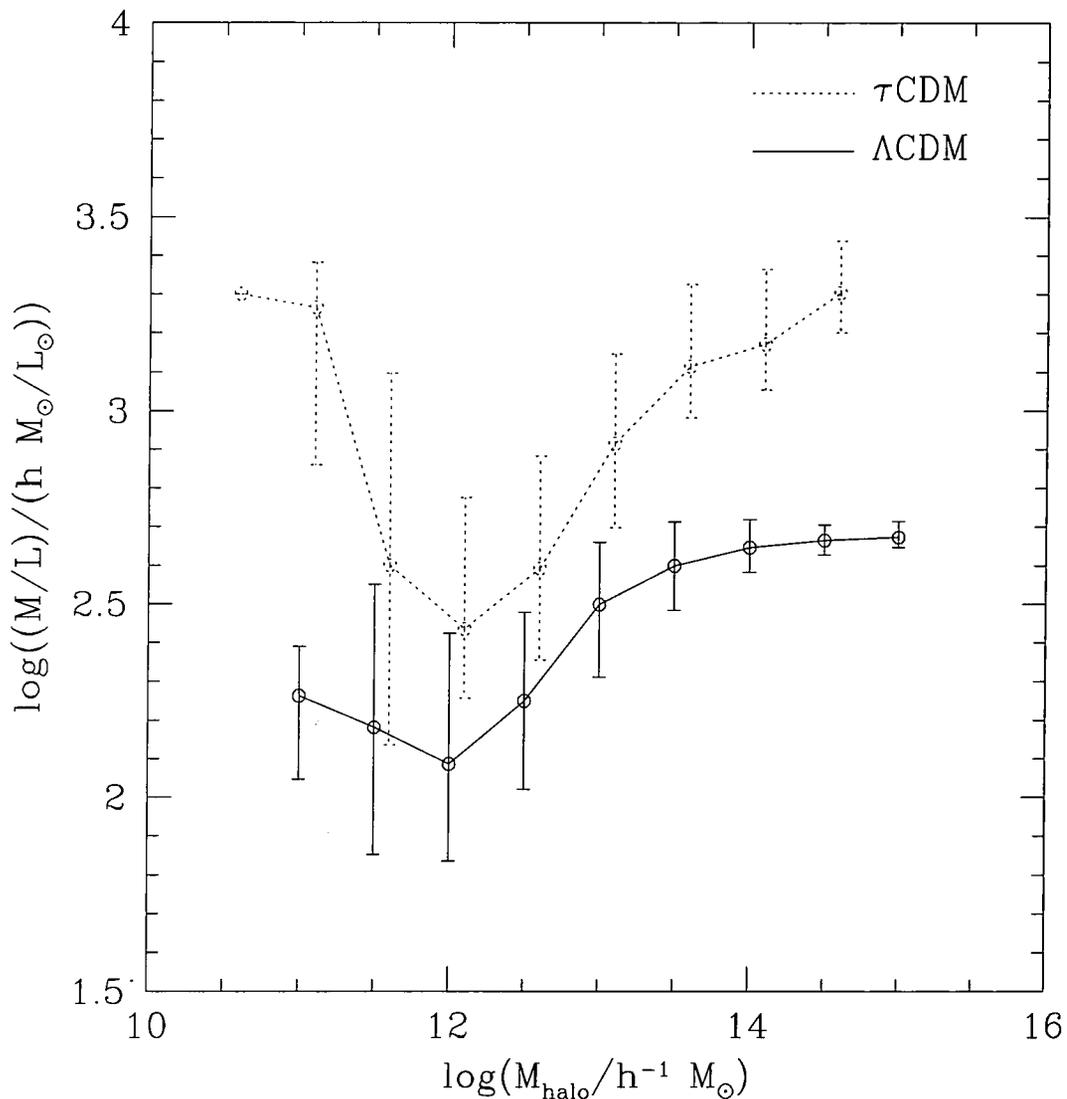


Figure 3.9: The B-band mass-to-light ratio of halos in the models. The dotted line corresponds to τ CDM and the solid line to Λ CDM. Lines show the median mass-to-light ratio, whilst the error bars indicate the 10 and 90 percentiles of the distribution. For reference, the mean mass-to-light ratio in the simulation as a whole is about 1440 and 470 hM_{\odot}/L_{\odot} in the τ CDM and Λ CDM cosmologies respectively, with an uncertainty of about 20% due to unresolved galaxies.

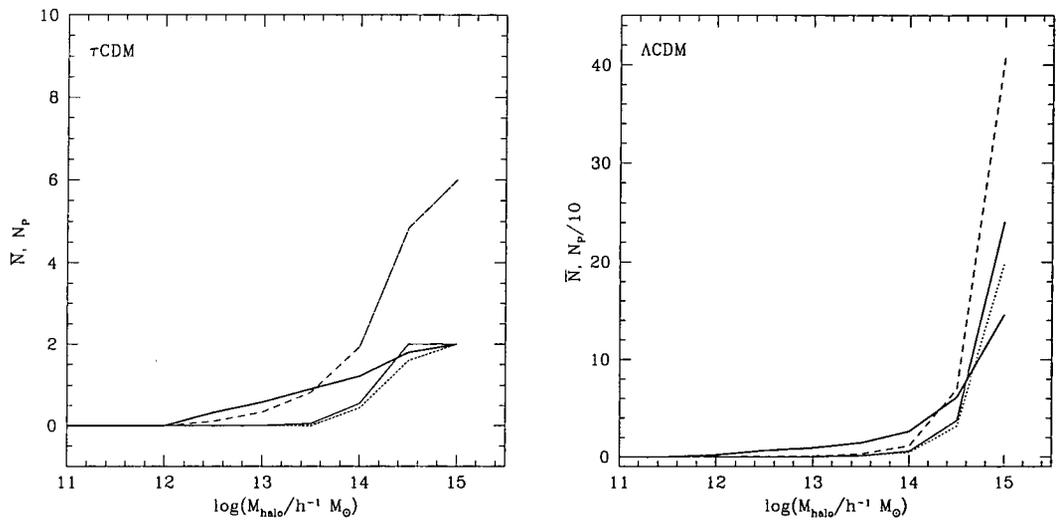


Figure 3.10: The thick solid lines show the mean number of galaxies, $\bar{N}(M)$, brighter than $M_B - 5 \log h = -19.5$ per halo as a function of halo mass. The plots are for the τCDM (left hand panel) and the ΛCDM (right hand panel) models. Note that unoccupied halos are included when computing the mean. The remaining lines indicate the mean number of galaxy pairs per halo as defined by eqn. (3.3), for three different probability distributions, $P(N; M)$: “true” (thin solid line); “average” (dotted line) and “Poisson” (dashed line). Note the different scales in the two plots.

estimated by averaging the Mo & White analytic bias for all halos of mass greater than M_h , weighting by the abundance of those halos and by the number of galaxies residing (on average) within them (see Baugh et al. 1999).

On smaller scales the situation is more complex. The Mo & White calculations break down on scales comparable to the pre-collapse (Lagrangian) radius of the host halos. If halos of mass M have a Lagrangian radius R , then we expect a reduction in the correlation of these halos on scales $\leq R$, since these objects must have formed from spatially exclusive regions of the universe. Halos may have moved somewhat after their formation and so will not be completely exclusive below this scale. However, they must be completely exclusive below their post-collapse (virial) radius (R_{vir}), since no two halos can occupy the same region of space. Galaxies, however, resolve the internal structure of the halos and so we should not necessarily expect the same degree of anti-bias on sub- R_{vir} scales in the galaxy distribution although these exclusion effects may still be apparent to some extent. Instead, the correlation function will begin to reflect the distribution of dark matter within the halos since, in these models, galaxies always trace the halo dark matter.

However, this is still not the whole picture. If $\bar{N}(M)$ is the average number of galaxies per halo of mass M , then we can define a mass $M' > M_h$, where $\bar{N}(M') = 1$ (here M_h is the minimum mass of a halo that can host a galaxy brighter than $M_B - 5 \log h = -19.5$). We find $M' = 10^{13}$ and $10^{12}h^{-1}M_\odot$ for the τ CDM and Λ CDM models respectively. In halos less massive than M' , we typically find at most a single galaxy and so the distribution of dark matter within these halos is not resolved. Instead, the galaxy catalogue contains information only about the position of the halo centre. In general, the clustering will depend upon $P(N; M)$, the probability of finding N galaxies in a halo of mass M . In particular, the small-scale clustering will depend upon the mean number of pairs per halo, which is itself determined by the form of the $P(N; M)$ distribution. Since the correlation function is a pair weighted statistic it gives extra weight to distributions with a tail to high N .

The correlation function of galaxies is thus the result of a complex interplay of several effects: (i) asymptotic constant bias on large scales; (ii) spatial exclusion of halos; (iii) the number of galaxies per halo which controls whether the internal structure of the halos is resolved or not and (iv) the form of the $P(N; M)$ distribution (as this determines the mean number of pairs of galaxies per halo), which we discuss below. It is difficult, therefore, to construct an empirical model that reproduces the results of the full semi-analytic plus N-body models (but see Peacock & Smith 2000). It is, however, instructive to plot several

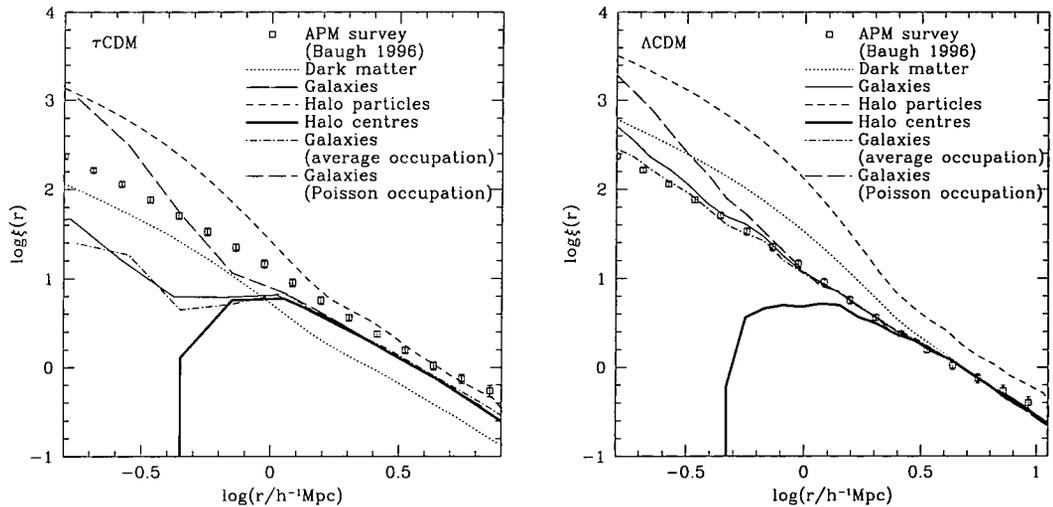


Figure 3.11: Correlation functions constructed from different samples of dark matter particles compared to the observed and model galaxy correlation functions in the τ CDM reference model (left-hand panel). The various curves, labelled in the legend, are described in detail in the text. The right hand panel shows the same plots for the Λ CDM reference model.

correlation functions which act as bounds on the true galaxy correlation function.

Fig. 3.11 shows the correlation functions of galaxies in our model (thin solid line), dark matter in the simulation (dotted line), and observed galaxies in the APM survey, as measured by Baugh (1996) (squares with error bars). The short-dashed line is computed from all dark matter particles that are part of halos of mass greater than M_h (i.e. halos sufficiently massive to contain galaxies at least some times). This curve is highly biased with respect to the full dark matter distribution, a fact that is not surprising given that it excludes the least clustered mass. We would expect the galaxy correlation function to be similar to this if the number of galaxies per halo were drawn from a Poisson distribution with mean proportional to the halo mass, that is, if the mass-to-light ratio were independent of halo mass. Evidently this is not the case. (The asymptotic bias of this correlation function is greater than that of the model galaxies (thin solid line), as weighting by halo mass gives more weight to the highly biased, most massive halos than does weighting by galaxy number.) The heavy solid line is the correlation function of halo centres, with each centre weighted by the model $P(N; M)$ distribution. The spatial exclusion of halos is evident, causing this curve to drop below that of the galaxies and finally to plummet to $\xi(r) = -1$ at a scale comparable to twice the virial radius of the smallest occupied halos.

The dot-dash line shows the correlation function found by placing in each halo the

average number of galaxies per halo of each mass $[\sum_{N=1}^{\infty} NP(N; M)]$, using our usual placement scheme (i.e. the first galaxy is placed at the halo centre and the others are attached to random particles in the halo). We refer to this as the “average” model. Obviously, we cannot place the average number per halo if this is not an integer. In this case, we place a number of galaxies equal to either the integer immediately below or immediately above the actual mean with the relative frequencies needed to give the required mean, which results in a small scatter in the occupation number. Finally, the long-dashed line shows the correlation function obtained when the number of galaxies in a halo is drawn from a Poisson distribution with the same mean as the model distribution (the “Poisson” model).

The differences between the correlation functions of the full semi-analytic model and models in which halos are occupied according to a Poisson distribution or simply with the average galaxy number (thin solid line, long-dashed line and dot-dash line respectively) must be due entirely to the form of $P(N; M)$ (i.e. the frequency with which a halo of a given mass is occupied by N galaxies), since all these models are, by construction, identical in all other respects, including the mean halo occupation number. Fig. 3.12 illustrates the difference between the actual distribution of galaxies in all halos resolved in the simulations,

$$P(N) = \int_{M_{\min}}^{\infty} P(N; M)n(M)dM / \int_{M_{\min}}^{\infty} n(M)dM, \quad (3.2)$$

and the “Poisson” and “average” models. Here M_{\min} is the mass of the smallest halo that can be resolved in the simulation. The values of $P(N)$ in this plot are multiplied by $N(N - 1)$ so that the area under the histogram gives the mean number of pairs per halo. The number of galaxies present in a halo is related to the structure of the merger tree for that halo. Although the merger tree is generated by a Monte-Carlo method, this does not produce a Poisson distribution of progenitor halos. Furthermore, whilst the Poisson distribution always possesses a tail to arbitrarily high numbers, the real distribution cannot as there is only enough cold gas in any one halo to make a limited number of bright galaxies.

Table 3.2 gives the mean number of pairs found within a single halo in the two reference models for all halos resolved in the simulation. This is given by

$$N_P = \sum_{i=0}^{\infty} i(i - 1) \frac{\int_{M_{\min}}^{\infty} P(i; M)n(M)dM}{\int_{M_{\min}}^{\infty} n(M)dM}, \quad (3.3)$$

where $P(i; M)$ is the distribution of occupancies (normalised such that $\sum_{i=0}^{\infty} P(i; M) = 1$) for halos of mass M and $n(M)$ is the abundance of halos of mass M . The number of pairs

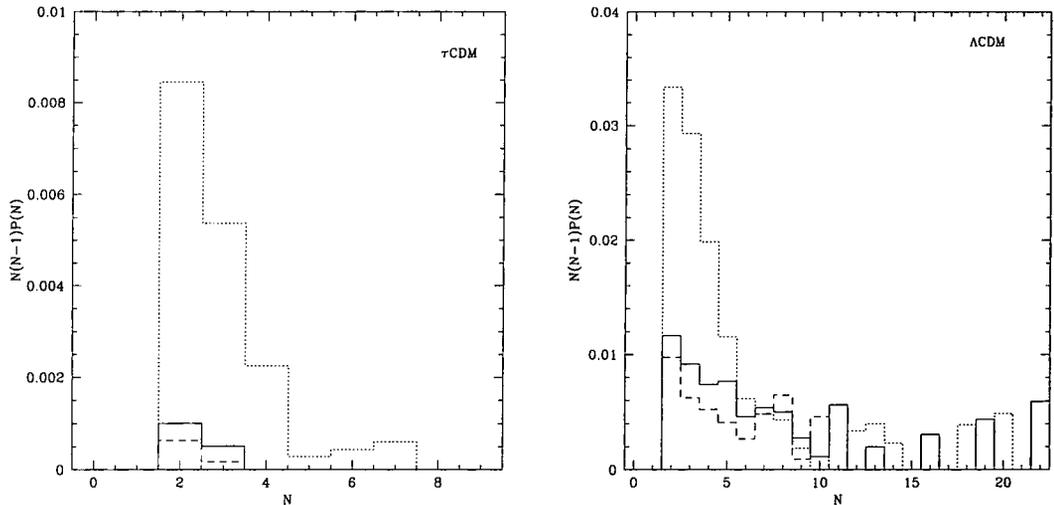


Figure 3.12: The probability, $P(N)$, of occupation by N galaxies (multiplied by $N(N-1)$ for clarity) for halos in the τ CDM model (left hand panel) and Λ CDM model (right hand panel). All halos resolved in the simulations are considered. The solid line shows the distribution from the actual model, the dotted line shows the “Poisson” model distribution and the dashed line shows the “average” model distribution. Note that the $N = 0$ and $N = 1$ bins are always zero because we choose to weight $P(N)$ by $N(N-1)$.

in the “true”, “average” and “Poisson” distributions is shown in Figure 3.10. Note that if we consider two such halos separated by some distance $\sim r$ then the mean number of pairs at separation $\sim r$ is

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} ijP(i; M)P(j; M) = \bar{i}^2 \quad (3.4)$$

and \bar{i} is constrained to be equal in all three distributions.

Thus, we can understand the difference in the clustering amplitudes of the three correlation functions at small scales (they all agree within the errors at large scales) simply on the basis of the form of their $P(N; M)$ function which determines the mean number of pairs per halo, N_P . For distributions with the same mean, the one with the lowest number of pairs per halo, N_P , will have the lowest clustering amplitude, whilst the one with the largest number of pairs will have the highest clustering amplitude. (In the case of the “average” and “true” distributions, the correlation functions are very similar on small scales as the contribution from pairs of galaxies within a single halo is small compared to that from pairs in distinct halos.) The consequence of this is that the amplitude of the small scale end of the observed correlation function tells us something interesting about $P(N; M)$, namely that it has fewer pairs than a Poisson distribution and

Table 3.2: The mean number of pairs per halo, N_P , calculated for three different distributions of halo occupancy, all with the same mean. “Average” has the same number of galaxies in all halos in a given mass range (or as close to this distribution as possible if the mean is not an integer), “true” has the distribution found in the reference models and “Poisson” has a Poisson occupation.

Model	average	true	Poisson
τ CDM	0.0005	0.0010	0.0116
Λ CDM	0.0538	0.0678	0.1262

is in reasonable agreement with the distribution predicted from the semi-analytic model. Thus, the behaviour of the small separation end of the correlation function is determined by the physics of galaxy formation. We have checked that our choice of placing central galaxies at the centre of mass of their halo does not affect these results. If instead each central galaxy is placed on a randomly chosen dark matter particle (i.e. if treated just like a satellite galaxy) the correlation function is unaltered within the error bars.

The τ CDM reference model shows a break on sub-Mpc scales. As can be seen in Fig. 3.11 this coincides with the turnover in the correlation function of halo centres. This turnover is reflected in the galaxy correlation function because in this model there are too few bright galaxies in cluster halos to adequately resolve their internal structure. In the Λ CDM models, on the other hand, halos are adequately resolved and the galaxy correlation function remains almost a power-law, even though that of halo centres turns over. To remove this feature from the τ CDM model would require more bright galaxies to form in cluster halos. However, this would have to be accomplished without significantly increasing the number of bright galaxies in lower mass halos since these would quickly come to dominate the asymptotic bias which would therefore become lower than its present value thus exacerbating the discrepancy between model and observations at large separations. We have been unable to find a model constrained to match the local luminosity function which succeeds in removing the sub-Mpc feature in the τ CDM model whilst simultaneously producing the required asymptotic bias.

The reason for the differences between the two reference models is illustrated in Fig. 3.13, where we compare the models to an observational determination of the “luminosity function of all galactic systems.” This function, estimated by Moore, Frenk & White (1993), gives the abundance of halos as a function of the *total* amount of light

they contain, regardless of how it is shared amongst individual galaxies. This quantity is difficult to determine observationally, since one must establish which galaxies are in the same dark matter halo. Moore, Frenk & White (1993) approached this problem by analysing the “CfA-1” galaxy redshift survey (Davis et al. 1982, Huchra et al. 1983) using a modified friends-of-friends group finding algorithm which was allowed to have different linking lengths in the radial and tangential directions to account for redshift-space distortions. These linking lengths were also allowed to vary with distance, to reflect the changing number density of galaxies in the survey. It is entirely possible, however, that in some instances this technique may have grouped together galaxies which actually reside in distinct dark matter halos. Finally, Moore, Frenk & White (1993) made a correction to the luminosity of each identified group to account for the light from unseen galaxies (i.e. those below the magnitude limit of the survey). This was done assuming a universal form for the galaxy luminosity function. Such a form may not, in fact, be applicable to the real Universe, and is not guaranteed to arise in our models.

Bearing these caveats in mind, it can be seen in Fig. 3.13 that the Λ CDM reference model and the data are in excellent agreement except at the faint end where the discrepancy reflects the fact that the data come from the CfA-1 Survey which has a flatter luminosity function than the ESO Slice Project (ESP) luminosity function we used to constrain the semi-analytic model. By contrast, the τ CDM model fails to match the luminosity function of all galactic systems, in spite of the fact that it agrees quite well with the bright end of the galaxy luminosity function (c.f. Fig. 3.2). This model does not make enough bright galaxies in high mass, highly clustered halos and it makes too many in low mass, weakly clustered halos. It is not surprising therefore that the τ CDM galaxy correlation function falls below the observed data on all scales (c.f. Fig. 3.14). A similar conclusion applies to the standard $\Omega_0 = 1$ CDM model although in this case the disagreement with the observed correlation function on large scales is even worse than in the τ CDM model. Matching the luminosity function of all galactic systems is, therefore, an important prerequisite for a model to match the two-point correlation function.

3.5 Testing the Robustness of the Predictions

The semi-analytic model of galaxy formation is specified by several parameters. These determine the cosmological model and control astrophysical processes such as star formation, supernovae feedback and galaxy merging. Whilst these parameters can be constrained by

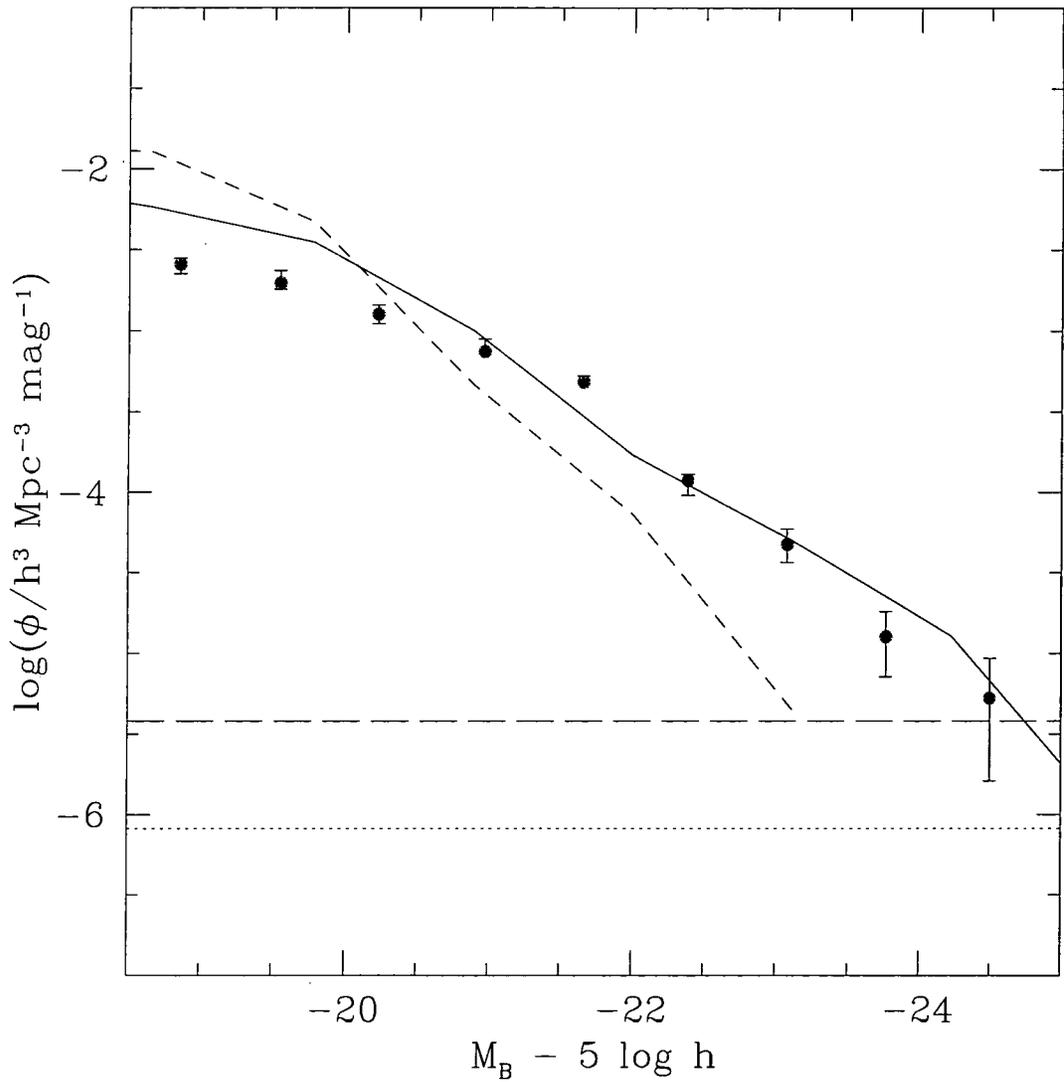


Figure 3.13: The luminosity function of all galactic systems. Results for the τ CDM model are shown by the short-dashed line and for the Λ CDM model by the solid line. The symbols with error bars are the observational data from Moore, Frenk & White (1993). The horizontal lines indicate the abundance below which the probability of finding one or more such objects in the entire volume of the simulation is less than 10% in the τ CDM (long dashed line) and Λ CDM (dotted line) models.

requiring the model to reproduce certain local observations (such as the B and K-band luminosity functions; see Chapter 2 and Cole et al. 2000), we wish to explore here what effect altering these parameters has on our estimate of the correlation function.

Thus, we alter the parameters of the reference model one at a time. We try to preserve as good a match as possible to the local B-band luminosity function by giving ourselves the freedom of adjusting the value of Υ so that the model B-band luminosity function has the correct amplitude at L_* . Since the reference models give a good match not only to the B-band luminosity function, but also to a variety of other observational data (such as the distribution of colours, sizes, star formation rates, etc.), the modified models will, in general, not be as good as the reference models. Furthermore, in some cases matching the $M_B - 5 \log h = -19.6$ point of the ESP luminosity function requires $\Upsilon < 1$ which is unphysical (as it implies negative mass in brown dwarfs). However, this is not a serious concern here since we are only interested in testing the robustness of clustering properties to changes in model parameters. We also consider a few models in which Υ is set so as to match the zero-point of the I-band Tully-Fisher relation, rather than the amplitude of the luminosity function at L_* . These are closer to the models of Kauffmann et al. (1999a).

The semi-analytic model with the altered parameters is used to populate the N-body simulation with galaxies. We then measure the bias of galaxies brighter than $M_B - 5 \log h = -19.5$ in each model.

3.5.1 Models Constrained by the Luminosity Function

Both of the reference models which are constrained to match the $M_B - 5 \log h = -19.6$ point of the ESP luminosity function (Zucca et al. 1997), also reproduce the observed exponential cut-off at the bright end of the local B and K-band luminosity functions (cf. Fig. 3.2). This fact turns out to be of importance when studying the clustering of these galaxies.

Tables 3.3 & 3.4 list the variant models that we have studied in the τ CDM and Λ CDM cosmologies respectively. The first column of each table lists those parameters that have changed from the reference model (which is listed in the first row of the tables). “Recooling” models allow enriched gas reheated by supernovae to recool within a dark matter halo. All but one model use the “No recooling” algorithm in which this gas is not allowed to recool until its halo doubles in mass. In the “No dust” models we do not account for extinction by internal dust (see §2.5.3). Also listed in the tables is the value of Υ for each model and the average analytic asymptotic bias of the galaxies calculated

Table 3.3: Variant models in the τ CDM cosmology. The first column gives the value of the parameter which is varied relative to the reference model. The remaining columns give the values of Υ used to match a point in the luminosity function for each model and the asymptotic bias of the galaxies estimated from the fitting formula of Jing (1998), b_{analytic} , and from our models, b_{model} .

Model	Υ	b_{analytic}	b_{model}
Reference	1.23	1.27	1.65 ± 0.37
$v_{\text{hot}} = 350$ km/s	1.06	1.27	1.62 ± 0.36
$v_{\text{hot}} = 200$ km/s	1.41	1.26	1.57 ± 0.29
$\alpha_{*} = -0.25$	1.20	1.27	1.62 ± 0.33
$\alpha_{*} = -1.50$	1.33	1.27	1.63 ± 0.40
$\epsilon_{*} = 0.01^{\dagger}$	0.98	1.27	1.64 ± 0.28
$\epsilon_{*} = 0.04$	1.52	1.29	1.72 ± 0.42
$f_{\text{df}} = 0.5^{\dagger}$	1.17	1.27	1.67 ± 0.44
$f_{\text{df}} = 0.03$	1.23	1.25	1.59 ± 0.35
IMF: Kennicutt (1983)	2.01	1.29	1.65 ± 0.38
$r_{\text{core}} = 0.2$	1.22	1.26	1.60 ± 0.39
$r_{\text{core}} = 0.02$	1.23	1.28	1.65 ± 0.36
$\Omega_{\text{b}} = 0.10$	1.82	1.28	1.59 ± 0.31
$\Omega_{\text{b}} = 0.05$	0.55	1.26	1.71 ± 0.36
$p = 0.02$	1.20	1.29	1.69 ± 0.38
Recooling	1.68	1.28	1.64 ± 0.36
No dust	1.89	1.30	1.68 ± 0.36

[†] The two models with the greatest deviation from the mean two-point correlation function.

Table 3.4: Variant models in the Λ CDM cosmology. The first column gives the value of the parameter which is varied relative to the reference model. The remaining columns give the values of Υ used to match a point in the luminosity function for each model and the asymptotic bias of the galaxies estimated from the fitting formula of Jing (1998), b_{analytic} , and from our models, b_{model} .

Model	Υ	b_{analytic}	b_{model}
Reference	1.63	1.07	1.01 ± 0.03
$v_{\text{hot}} = 200$ km/s	1.45	1.07	0.98 ± 0.01
$v_{\text{hot}} = 100$ km/s	1.53	1.06	0.97 ± 0.02
$\alpha_* = -0.25$	1.63	1.08	0.98 ± 0.01
$\alpha_* = -1.50$	1.58	1.06	0.97 ± 0.01
$\epsilon_* = 6.67 \times 10^{-3}$	1.33	1.06	0.98 ± 0.01
$\epsilon_* = 0.02$	1.81	1.06	0.97 ± 0.01
$f_{\text{df}} = 5.0^\dagger$	1.29	0.93	0.88 ± 0.02
$f_{\text{df}} = 0.2^\dagger$	1.31	0.98	0.91 ± 0.01
IMF: Salpeter (1955)	0.90	1.01	1.00 ± 0.02
$r_{\text{core}} = 0.2$	1.63	1.07	0.98 ± 0.01
$r_{\text{core}} = 0.02$	1.63	1.07	0.99 ± 0.01
$\Omega_{\text{b}} = 0.04^\dagger$	2.04	1.13	1.02 ± 0.01
$\Omega_{\text{b}} = 0.01^\dagger$	0.70	0.96	0.95 ± 0.02
$p = 0.04^\dagger$	1.04	1.11	1.01 ± 0.01
Recooling [†]	1.67	1.11	1.00 ± 0.02
No dust	2.31	1.04	0.95 ± 0.01

[†] The six models with the greatest deviation from the mean two-point correlation function.

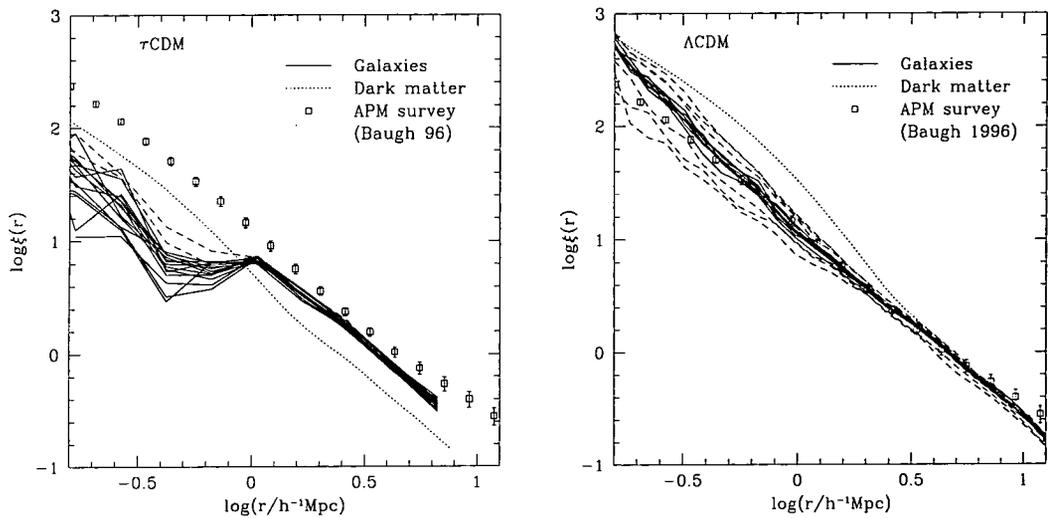


Figure 3.14: Correlation functions in the τ CDM (left hand panel) and Λ CDM (right hand panel) cosmologies. All the models are constrained to match the abundance of L_* galaxies in the ESP B-band luminosity function. In each plot the points with error bars show the observed APM real-space correlation function of Baugh (1996), whilst the dotted line shows the correlation function of the dark matter. The model galaxy correlation functions are shown as solid lines except in the case of models which deviate substantially from the average of all models which are shown as dashed lines.

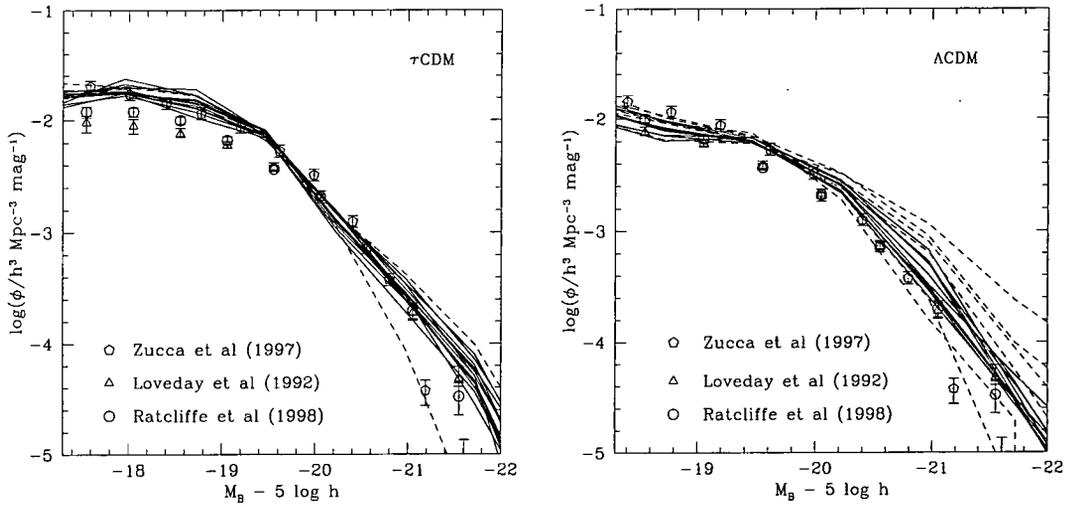


Figure 3.15: B-band luminosity functions in the τ CDM (left hand panel) and Λ CDM (right hand panel) models. All models are constrained to match the ESP luminosity function of Zucca et al. (1997) at $M_B - 5 \log h = -19.6$. Symbols with error bars show a selection of observational determinations of the luminosity functions, from the sources indicated in the legend. The solid lines show results for our models, except that the outliers identified in Fig. 3.14 are shown as dashed lines. Each luminosity function is plotted only to the completeness limit of the simulations.

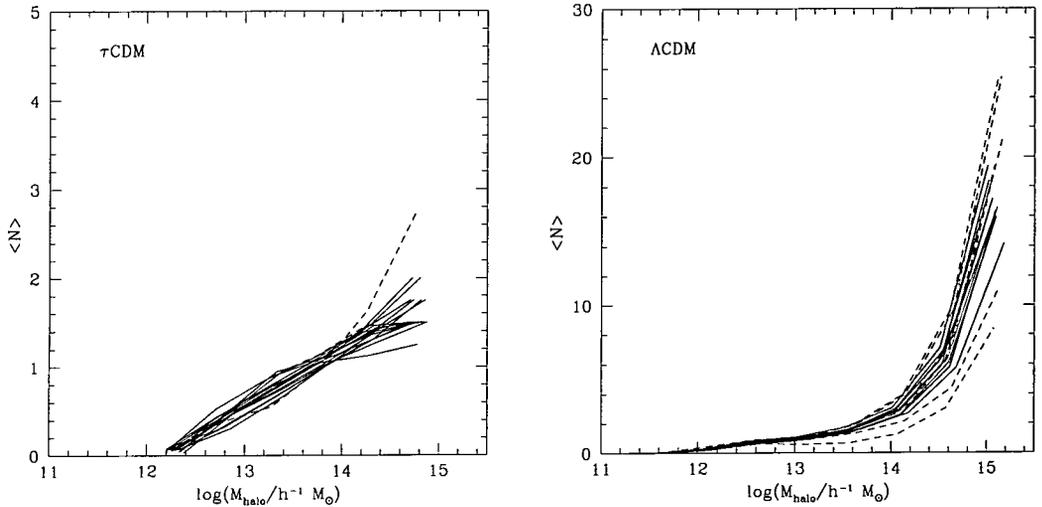


Figure 3.16: The average number of galaxies brighter than $M_B - 5 \log h = -19.5$ per halo as a function of halo mass. The panels refer to the τ CDM models (left) and the Λ CDM models (right). Note that unoccupied halos are counted. Dashed lines indicate the outlier models identified in Fig. 3.14. (Some of the lines lie on top of one another because the corresponding changes in model parameters do not alter the number of galaxies per halo.)

as follows:

$$b_{\text{analytic}} = \sum_{i=1}^N b(M_i)/N \equiv \frac{\int_0^\infty b(M) \bar{N}(M) n(M) dM}{\int_0^\infty \bar{N}(M) n(M) dM}, \quad (3.5)$$

where N is the number of galaxies in the catalogue, M_i is the mass of the halo hosting the i^{th} galaxy, $\bar{N}(M)$ is the mean number of galaxies per halo of mass M , and $n(M)$ is the dark matter halo mass function in the simulation. The function $b(M)$ is the asymptotic bias of halos of mass M which we estimate using Jing's (1998) formula obtained from fitting the results of N-body simulations. This formula tends to the analytic result of Mo & White (1996) for masses much greater than M_* . The final column gives the asymptotic bias estimated directly from the models, on scales where $\xi_{\text{matter}}(r) < 1$ (≈ 2.5 and $5.0h^{-1}$ Mpc in the τ CDM and Λ CDM cosmologies respectively), as described by Jing (1998). Note that the analytic biases are consistently lower than those measured in the τ CDM models. The halos in the τ CDM GIF simulation show a similar disagreement with the fitting formula of Jing (1998) which was tested on SCDM, OCDM and Λ CDM cosmologies only. The models marked by a dagger are those showing large deviations from the mean clustering amplitude of all models (six and two such models are identified in the Λ CDM and τ CDM cosmologies respectively).

In the remainder of this section we show the correlation functions obtained from these variant models and discuss how the form of the correlation function is related to other properties of the galaxy population. The correlation functions are displayed in Fig. 3.14. All cases show antibias on small scales and a constant bias on large scales. Most of the models in both cosmologies have similar correlation functions but the scatter is somewhat greater in the τ CDM case than in the Λ CDM case. The models that deviate most from the average are shown as dashed lines in Fig. 3.14 and also in all other plots in this section. In the Λ CDM cosmology, where the reference model is well fit by a power-law correlation function, the deviant models have slopes which are somewhat different from the other models.

The luminosity functions in most of the Λ CDM models, plotted as solid lines, in Fig. 3.15, are quite similar. (They are all forced to go through the same point at $M_B - 5 \log h = -19.6$.) The ones that deviate the most are those plotted as dashed lines, that is, those that were identified in Fig. 3.14 as giving the most discrepant correlation functions. Thus, we see that the main factor that determines the sensitivity of the correlation function to model parameters is the ability of the model to reproduce the exponential cut-off observed in the luminosity function. Models that achieve this all

give similar galaxy correlation functions. A similar conclusion applies in the τ CDM case, although here the distinction is less clear cut due to the noisy estimates of the correlation function on small scales.

We have also considered two models that have different dark matter power spectra. These make use of the GIF SCDM and OCDM simulations (described in §3.2) and, apart from the values of the cosmological parameters, they have the same parameter values as the reference τ CDM and Λ CDM models respectively. Whilst the OCDM model shows very little difference from the Λ CDM model, the SCDM model has a significantly different clustering amplitude than our reference τ CDM model (approximately 40% lower on scales larger than 1 Mpc). Despite this its luminosity function is in fairly close agreement with that of the τ CDM model at the bright end. This demonstrates that models with the same luminosity function only produce the same correlation function if they have the same underlying dark matter distribution.

As described in §3.3 the distribution of the number of galaxies per halo as a function of halo mass is very important in determining the behaviour of the correlation function. On small scales, the full distribution determines the amplitude and slope of the correlation function, whilst on large scales the number of galaxies per halo determines the asymptotic bias of the galaxy distribution by selecting the range of host halo masses that dominates the correlation function. Fig. 3.16 shows the number of galaxies per halo as a function of halo mass in the models. It is apparent, particularly for Λ CDM, that the models identified as outliers in the correlation function plot (Fig. 3.14) are also the ones that deviate the most from the reference models in these plots as well.

3.5.2 Models Constrained by the Tully-Fisher Relation

We have shown that matching the local galaxy luminosity function – our preferred method for constraining the parameters of the semi-analytic model – leads to model predictions for galaxy clustering that are robust to reasonable changes in these parameters. Kauffmann et al. (1999a) adopted a different philosophy: they chose to constrain their models by matching the I-band Tully-Fisher relation, rather than the luminosity function. We explore the effect of this choice by constraining our own models in a similar way. Specifically, we require the median magnitude of central spiral galaxies with halo circular velocities in the range 215.0 to 225.0 km s⁻¹ to be $M_I - 5 \log h = -22.0$ (where M_I is the dust-extincted I-band magnitude of each galaxy corrected to the face-on value). This can be achieved by a suitable choice of the luminosity normalisation parameter, Υ , and leads

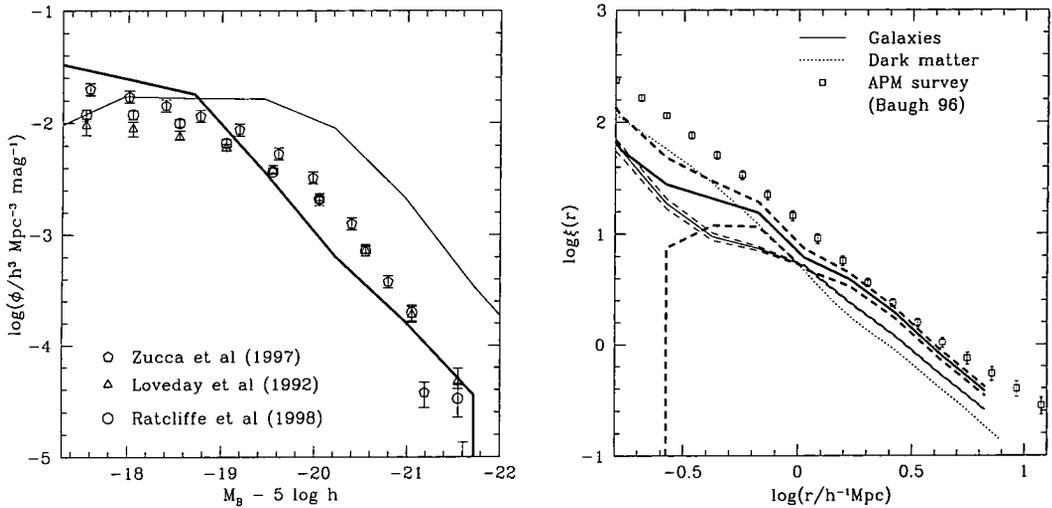


Figure 3.17: The B-band luminosity function (left) and galaxy correlation function (right) for two τ CDM models constrained to match the I-band Tully-Fisher relation. The thin line in both panels corresponds to the reference model (but with the value of Υ required by the Tully-Fisher relation) and the thick lines corresponds to a model with very weak feedback. The luminosity function is shown only to the completeness limit of this model. In the right hand panel the symbols with error bars show Baugh's (1996) APM correlation function, the dashed lines the Poissonian errors on the correlation functions, and the dotted line the dark matter correlation function.

to a model Tully-Fisher relation that agrees well with data from Mathewson, Ford & Buchhorn (1992).

Since the original Λ CDM models (that is, the reference model and its variants) already agreed quite well with the Tully-Fisher relation, (see Fig. 3.4), this different choice of constraint has only a minor effect on the correlation function. The only noticeable change is an increase in the scatter of the asymptotic bias in the variant models. In the τ CDM models, on the other hand, the new constraint has an important effect because the original models that agreed well with the luminosity function, missed the Tully-Fisher relation by about 1 magnitude. Forcing a fit to the Tully-Fisher relation destroys the good agreement of the reference model with the luminosity function, as may be seen in Fig. 3.17. This figure also shows a τ CDM model in which we have attempted to obtain a better luminosity function by dramatically reducing the amount of supernovae feedback into the interstellar gas.

Fig. 3.17 shows that the two τ CDM models that match the Tully-Fisher relation have different correlation functions. In other words, this exercise demonstrates that when models are constrained in this way, the resulting correlation functions are rather sensitive to the choice of model parameters. This explains why Kauffmann et al. (1999a) concluded that their clustering predictions depended strongly on the way they parameterised star formation, feedback and the fate of reheated gas in their model. By contrast, we have found that our predictions for the correlation function are robust to changes in model parameters, *so long as the models match the bright end of the galaxy luminosity function.*

3.6 Discussion and Conclusions

In this Chapter, we have considered some of the physical and statistical processes that determine the distribution of galaxies and its relation to the distribution of mass in cold dark matter universes. The approach that we have adopted exploits two of the most successful techniques currently used in theoretical cosmological studies: N-body simulations to follow the clustering evolution of dark matter and semi-analytic modelling to follow the physics of galaxy formation. Our main conclusion is that the efficiency of galaxy formation depends in a non-trivial fashion on the mass of the host dark matter halo and, as a result, galaxies, in general, have a markedly different distribution from the mass. This result had been anticipated in early cosmological studies (e.g. Frenk, White & Davis 1983, Davis et al. 1985, Bardeen et al. 1986), but it is only with the development

of techniques such as semi-analytic modelling that realistic calculations have become possible.

The statistics of the spatial distribution of galaxies reflect the interplay between processes that determine the location where dark matter halos form and the manner in which halos are “lit up” by galaxy formation. If the resulting mass-to-light ratio of halos were independent of halo mass, then the distribution of galaxies would be related in a simple manner to the distribution of dark matter in halos. In current theories of galaxy formation, however, the mass-to-light ratio has a complicated dependence on halo mass. On small mass scales, galaxy formation is inhibited by the reheating of cooled gas through feedback processes, whereas in large mass halos it is inhibited by the long cooling times of hot gas. As a result, the mass-to-light ratio has a deep minimum at the halo mass, $\sim 10^{12}M_{\odot}$, associated with L_* galaxies, where galaxy formation is most efficient. Although our calculations assume a specific model of galaxy formation, the dependence of mass-to-light ratio on halo mass displayed in Fig. 3.9 is likely to be generic to this type of cosmological model. The consequence of such a complex behaviour is a scale dependent bias in the distribution of galaxies relative to the distribution of mass.

On scales larger than the typical size of the halos that harbour bright galaxies, the bias in the galaxy distribution is related in a simple way to the bias in the distribution of massive halos. In our $\Omega_0 = 1$ τ CDM model, galaxies end up positively biased on large scales, but in our flat, $\Omega_0 = 0.3$ Λ CDM model, they end up essentially unbiased. On small scales, the situation is more complicated and the correlation function depends on effects such as the spatial exclusion of dark matter halos, dynamical friction, and the number of galaxies per halo. In particular, our simulations show how the statistics of the halo occupation probability influence the amplitude of the galaxy correlation function on sub-megaparsec scales. In these models, the occupation of halos by galaxies is not a Poisson process. Since the amount of gas available for star formation is limited, the mean number of pairs per halo is less than that of a Poisson distribution with the same mean number of galaxies per halo. This property plays an important role in determining the amplitude of small scale correlations.

Remarkably, the correlation function of galaxies in our Λ CDM model closely approximates a power-law over nearly four orders of magnitude in amplitude. This is in spite of the fact that the correlation function of the underlying mass distribution is not a power-law, but has a continuously changing logarithmic slope in the relevant range of scales. Somehow, the various effects just discussed conspire to compensate for these features in

the mass distribution. In particular, on scales smaller than $\sim 3h^{-1}$ Mpc, the galaxy distribution in the Λ CDM model is *antibias* relative to the mass distribution. The apparently scale-free nature of the galaxy correlation function in this model seems to be largely a coincidence (although whether this is also true of the real Universe remains to be seen). The τ CDM model, which has similar physics although a different initial mass fluctuation spectrum, does not end up with a power-law galaxy correlation function.

Colín et al. (1999) have carried out a very high resolution N-body simulation (of a Λ CDM cosmological model similar to ours) which resolves some substructure within dark matter halos. The correlation function of these sub-halos is remarkably similar to the correlation function of the galaxies in our Λ CDM reference model. In a sense, the merger trees of the semi-analytic model keep track of sub-halos within dark matter halos since they follow the galaxies that form within them. (Unlike the simulations, however, the semi-analytic model does not follow the spatial distribution of sub-halos.) Colín et al. select sub-halos by circular velocity, whereas we select galaxies by luminosity. Since, in our models, the luminosity of a galaxy is correlated with the circular velocity of the halo in which it formed, there is some correspondence between the type of objects studied by Colín et al. and ourselves. However, the connection could be complicated by effects such as tidal disruption or stripping of halos within halos but which are not included in our model. Nevertheless, the abundance of sub-halos considered by Colín et al. is similar to the abundance of galaxies in our Λ CDM model and this might account for the similarity between the two correlation functions.

Another noteworthy outcome of our simulations is the close match of the galaxy correlation function in the Λ CDM model to the observed galaxy correlation function, itself also a power-law over a large range of scales (Groth & Peebles 1977, Baugh 1996). This match is particularly interesting because the parameters that specify the semi-analytic galaxy formation model were fixed beforehand by considerations that are completely separate from galaxy clustering (see Chapter 2 and Cole et al. 2000). The procedure for fixing these parameters places special emphasis on obtaining a good match to the observed galaxy luminosity function (c.f. Fig. 3.2), but makes no reference whatsoever to the spatial distribution of the galaxies.

To summarize, the combination of high resolution N-body simulations with semi-analytic modelling of galaxy formation provides a useful means for understanding how the process of galaxy formation interacts with the process of cosmological gravitational evolution to determine the clustering pattern of galaxies. In general, we expect galaxies to

be clustered somewhat differently from the dark matter, and the relation between the two can be quite complex. A flat CDM model with $\Omega_0 = 0.3$ gives an acceptable match to the observed galaxy correlation function over about four orders of magnitude in amplitude (as does an open model with the same value of Ω_0 .) The Λ CDM model is also in reasonable agreement with a number of other known properties of the galaxy distribution.

Chapter 4

The Dependence of Velocity and Clustering Statistics on Galaxy Properties

4.1 Introduction

Measurements of the clustering and peculiar motions of galaxies can, in principle, reveal the way in which galaxies are related to the large-scale distribution of mass in the universe and thus provide strong tests of models of galaxy formation. In this Chapter we exploit the technique developed in Chapter 3 to investigate three aspects of the galaxy distribution that are particularly sensitive to the galaxy-mass connection: the distribution of relative velocities of galaxy pairs, redshift-space distortions in statistical measures of galaxy clustering, and the dependence of clustering strength on intrinsic galaxy properties.

Most current discussion of large-scale structure takes place within the context of the cold dark matter (CDM) theory. It has long been known, however, that cold dark matter models in which the galaxies are assumed to have the same statistical distribution as the dark matter do not account for the relatively low *rms* pairwise velocity dispersion measured in galaxy surveys. Indeed, the first N-body simulations of the standard CDM model (Davis et al. 1985) gave values 2-3 times higher than those that had been measured for galaxies in the Center for Astrophysics (CfA) redshift survey by Davis & Peebles (1983). This conflict stimulated the development of the “high peak” model of galaxy formation (Bardeen et al. 1986), in which galaxies are assumed to be more strongly clustered than, and therefore to be biased tracers of, the dark matter distribution. In this case, mass fluctuations at the present day and the associated peculiar velocities are weaker than in an unbiased model with the same level of galaxy clustering. Davis et al. showed

that galaxies identified with high peaks in the standard $\Omega_0 = 1$ CDM cosmology could approximately match the observed amplitude of galaxy clustering while having a velocity dispersion distribution only slightly higher than that measured by Davis & Peebles (1983) (and similar, in fact, to that in an $\Omega_0 = 0.2$ CDM model).

The observational determination of the pairwise velocity dispersion has been problematic. It is now recognized that this statistic is very sensitive to the contribution from galaxies in rich clusters, which were not fairly represented in early redshift surveys (Mo, Jing & Börner 1993, Zurek et al. 1994, Mo, Jing & Börner 1997, Somerville, Primack & Nolthenius 1997, Somerville, Davis & Primack 1997). More recent estimates from larger surveys, typically sampling volumes on the order of $10^6 h^{-3} \text{Mpc}^3$, are more robust and in much better agreement with one another (Marzke et al. 1995, Jing, Mo & Börner 1998, Ratcliffe et al. 1998a). The current values are around $\sim 500 \text{ km s}^{-1}$, almost twice as large as the original measurement by Davis & Peebles (1983).

A complementary way to explore the connection between the distributions of dark matter and galaxies is to look for a dependence of clustering on galaxy properties such as intrinsic luminosity, morphology or colour. There have been many attempts to measure galaxy clustering as a function of luminosity, often with conflicting conclusions (Phillipps & Shanks 1987, Alimi, Valls-Gabaud & Blanchard 1988, Davis et al. 1988, Hamilton 1988, White, Tully & Davis 1988, Santiago & da Costa 1990, Iovino et al. 1993, Park et al. 1994). A number of recent measurements of the redshift-space correlation function of galaxies selected by absolute magnitude suggest a dependence of clustering strength on luminosity for galaxies brighter than L_* (Benoist et al. 1996, Tadros & Efstathiou 1996, Willmer, da Costa & Pellegrini 1998, Guzzo et al. 1999, but see Loveday et al. 1995 and Hoyle et al. 1999 for counter-examples).

A dependence of clustering on morphological type is, however, well established, with elliptical galaxies found to be more strongly clustered than spiral galaxies as measured both in angular catalogues (Davis & Geller 1976, Giovanelli, Haynes & Chincarini 1986, Iovino et al. 1993) and in three-dimensional surveys (Moore et al. 1994, Loveday et al. 1995). This result can be understood as a reflection of the morphology-density relation, the statement that elliptical galaxies preferentially inhabit high density regions of the Universe (Dressler 1980). The dependence of clustering on the colour of galaxies has also been measured in redshift surveys with the general consensus that red galaxies are more strongly clustered than blue galaxies (Carlberg et al. 1997, Le Fèvre et al. 1996, Willmer, da Costa & Pellegrini 1998). This is, of course, related to the fact that elliptical galaxies

are generally redder than spiral galaxies, and so this effect is due largely to the same physical processes that drive the morphological dependence of clustering. This colour dependence is not as clear cut for studies of the angular correlation function (Infante & Pritchett 1993, Landy, Szalay & Koo 1996), where the red and blue populations may have different redshift distributions.

In order to interpret the significance of these observational results, it is necessary to replace heuristic biasing schemes with physically motivated models that can realistically incorporate the role of the environment in the formation and evolution of galaxies. We applied such a technique in Chapter 3, and found that the semi-analytic model of Cole et al. (2000) produced a galaxy correlation function in a Λ CDM cosmology (i.e. a CDM model with $\Omega_0 = 0.3$ and $\Lambda_0 = 0.7$) that matches remarkably well the real-space correlation function measured from the APM galaxy survey (Baugh 1996). On small scales, the model galaxies are less clustered than the dark matter. A similar conclusion was reached by Pearce et al. (1999) for galaxies formed in cosmological gas dynamics simulations and by Colín et al. (1999) for galactic dark matter halos formed in high-resolution N-body simulations.

In this Chapter, we combine our semi-analytic model of galaxy formation with high-resolution N-body simulations to investigate the clustering properties of galaxies selected according to various criteria. We consider clustering in real and redshift-space and show that galaxy velocity dispersions are significantly lower than those of the dark matter, in reasonable agreement with the data. Our results for this particular statistic differ somewhat from those recently obtained by Kauffmann et al. (1999a) and we explore the reasons for these differences. Kauffmann et al. also investigated the dependence of clustering amplitude on galaxy properties, finding that both early type and red galaxies were more strongly clustered than the galaxy population as a whole, in agreement with the work of Kauffmann, Nusser & Steinmetz (1997) and with the trends seen in the observational data. However, they found no evidence in their models for the luminosity dependent clustering amplitude measured by Willmer, da Costa & Pellegrini (1998). Our results for these properties are quite consistent with those of Kauffmann et al. (1999a), but we compare our models to different datasets.

The rest of this Chapter is laid out as follows. In §4.2, we briefly describe the semi-analytic model of galaxy formation and the N-body simulations used in this study. In §4.3 we describe how we take into account the effects of redshift space distortions and calculate galaxy velocity dispersions. We examine the model correlation functions of

galaxies selected by luminosity, morphology and colour and compare these results with the available observational data in §4.4. Finally, in §4.5 we present our conclusions.

4.2 Description of the Numerical Technique

Here we use the technique developed in Chapter 3 to study the clustering of galaxies in the GIF Λ CDM simulation. Using model parameters identical to those of Table 3.1 the semi-analytic model populates the simulated volume with approximately 12,000 galaxies brighter than $M_B - 5 \log h = -19.5$. In order to investigate the clustering properties of very bright, low abundance galaxies we have used a different simulation of a much larger volume. This was carried out by the VIRGO Consortium and hereafter we will refer to it as the “512³” simulation. It has identical cosmological parameters to the GIF Λ CDM simulation, although the initial conditions were generated using a transfer function calculated from CMBFAST (Seljak & Zaldarriaga 1996) rather than the fitting function of Bond & Efstathiou (1984) which was used for the GIF simulations. The simulated volume is a cube of side $479h^{-1}\text{Mpc}$, populated with 134 million particles, giving a particle mass of $6.8 \times 10^{10}h^{-1}M_\odot$. Resolving halos of 10 or more particles in this simulation results in galaxy catalogues which are complete for galaxies brighter than $M_B - 5 \log h \approx -20.3$. Thus, we will only use this simulation for studying the distribution of the very brightest galaxies.

We have found that the clustering strength of the dark matter in the 512³ simulation is slightly larger than in the GIF simulation at large pair separations. The ratio of the real-space dark matter correlation functions in the two simulations, $\xi_{512^3}/\xi_{\text{GIF}}$, is 1.4 at pair separations of $10h^{-1}\text{Mpc}$, decreasing to 1.1 at $4h^{-1}\text{Mpc}$, and becoming unity at $2h^{-1}\text{Mpc}$. This difference is due primarily to finite volume effects in the GIF simulation, which is missing some of the large-scale power included in the 512³ simulation. This deficit is carried through into the galaxy correlation functions. It should be noted, however, that this effect does not invalidate the conclusions of Chapter 3 and, if anything, it slightly improves the agreement between our models and the APM correlation function on scales greater $\gtrsim 10h^{-1}\text{Mpc}$.

Recall that the technique that we employ differs in two important respects from that used by Kauffmann et al. (1999a). Firstly, Kauffmann et al. extract the full merger history of each dark matter halo directly from the N-body simulation. Although this approach has the advantage that the evolution of any individual galaxy can be followed through the

simulation, the properties of the merger histories generated by the Monte-Carlo method are statistically equivalent to those obtained directly from the N-body simulation (but have superior mass resolution). In particular, the merger histories in the N-body simulation are found not to correlate with environment, in agreement with the assumptions of the Monte-Carlo approach (Somerville et al. 1998, Lemson & Kauffmann 1999). As a result, any statistical properties, such as the correlation function, are not affected by the differences between the two approaches. Secondly, Kauffmann et al. adjusted the parameters of their semi-analytic model to match the zero-point of the Tully-Fisher relation between spiral galaxy luminosity and rotation speed rather than the observed luminosity function as we do. As a result, some of their models give a poor match to the luminosity function. In Chapter 3 we have argued that a good match to the luminosity function is required for a robust estimate of the correlation function which, furthermore, for Λ CDM provides a good match to the data.

Our semi-analytic model also follows the evolution of galaxy morphology, as described in §2.4.2 (see also Baugh 1996, Cole et al. 2000), a feature that we exploit in §4.4.2 to investigate the dependence of clustering strength on morphology. We remind the reader that in our model, virialised gas cools to form a rotationally supported disk and quiescent star formation turns this gas into stars. A galactic bulge can be built up through the accretion of relatively small satellite galaxies or as the result of a violent merger between galaxies of comparable mass. In the latter case, any material in a disk is rearranged into a bulge and a burst of star formation occurs if cold gas is present. The morphology of a galaxy can thus change with time and is related to its environment, which determines both the rate at which gas cools and the merger rate between galaxies.

4.3 Redshift Space Distortions

The clustering of galaxies in real space cannot be measured directly unless a distance indicator is available that does not rely on redshift. However, the real space clustering can be inferred indirectly, by deprojecting either the correlation function measured in terms of projected galaxy separation or the angular correlation function (e.g. Peebles 1980). In the case of noisy projected measurements, both methods require assumptions about the form of the two-point correlation function in real space, whilst the second method also requires knowledge of the redshift distribution of galaxies.

The three dimensional correlation function in redshift-space can be measured directly,

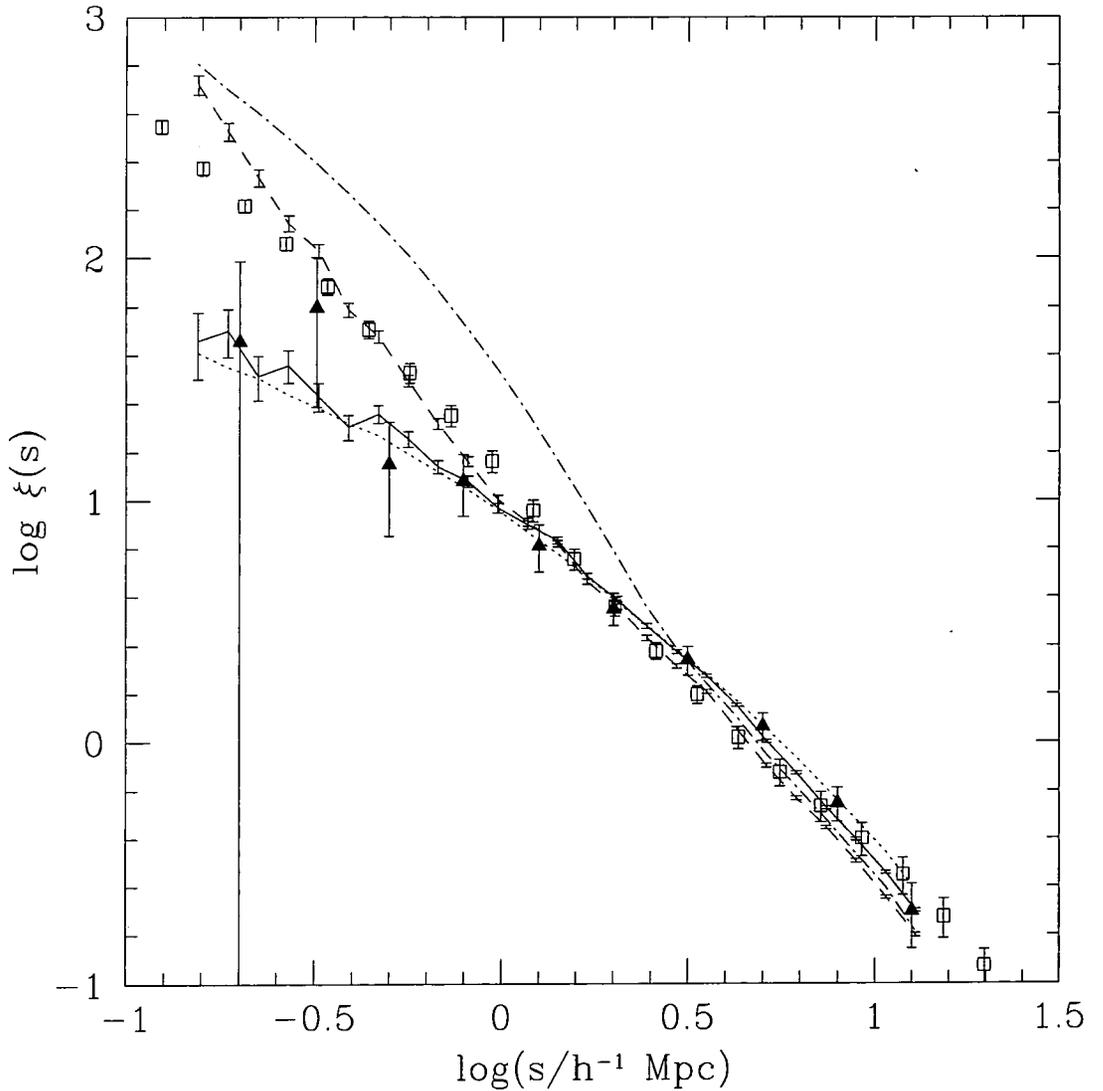


Figure 4.1: Real and redshift-space correlation functions in the Λ CDM model. The redshift-space correlation function of model galaxies brighter than $M_B - 5 \log h = -19.5$ is shown as a solid line, while the redshift-space correlation function of dark matter is shown as a dotted line. The dashed and dot-dashed lines show the real space correlation functions for galaxies and dark matter respectively. Error bars on the galaxy correlation functions are derived from Poisson statistics. The symbols with error bars are observational determinations of the galaxy correlation function: triangles show the redshift-space correlation function of galaxies brighter than $M_B - 5 \log h = -19.5$ from the ESO Slice Project (Guzzo et al. 1999), whilst squares show the real-space galaxy correlation function from the APM survey (Baugh 1996).

by inferring the radial position of each galaxy from its redshift, assuming a uniform Hubble flow. However, galaxies have peculiar motions in addition to the Hubble flow, induced by inhomogeneities in the density field. This leads to a modification of the apparent pattern of galaxy clustering, commonly referred to as “redshift-space distortions”. Generally, these distortions reduce the strength of clustering on small scales, where the motion of galaxies within virialised structures tends to smear out galaxy positions inferred from redshifts (creating ‘Fingers of God’), and boost the strength of clustering on large scales where galaxies experience coherent flows into overdense regions and out of underdense regions (Kaiser 1987, Hamilton 1992). The redshift-space galaxy correlation function in our model therefore depends on both the spatial distribution of the galaxies and their distribution of peculiar velocities. We will comment on uncertainties that this introduces in §4.3.2.

The redshift-space clustering of galaxies in the simulation is straightforward to compute from the position and peculiar velocity of each galaxy. Redshift space positions are calculated by adding v_x/H_0 to the x coordinate of each galaxy, where v_x is the x -component of the peculiar velocity of that galaxy. This gives the redshift space position (at $z = 0$), as viewed by an observer located at infinity. From the resulting catalogue we can compute clustering statistics in the usual way.

In Fig. 4.1 we show the two-point correlation function of galaxies brighter than $M_B - 5 \log h = -19.5$ in redshift space. The figure also shows the real-space correlation function for comparison. The two effects of redshift-space distortion described above are readily apparent in this figure. In real space the galaxy correlation function has an almost power-law form between 0.1 and $10h^{-1}$ Mpc, with a slope of $\gamma \approx -1.8$. In redshift-space, the galaxy correlation function falls below this power-law on small scales, whereas on scales larger than a few megaparsecs, it rises above it.

Kaiser (1987) showed that the amplification in redshift-space of the two-point correlation function on large scales due to coherent inflow onto large-scale structures is approximately a factor of $1 + 2\beta/3 + \beta^2/5$, where $\beta = \Omega_0^{0.6}/b$ and b is the galaxy bias, assumed to be independent of scale. For the model considered here, $\Omega_0 = 0.3$, and the galaxies are essentially unbiased on large scales. The amplification factor is therefore ~ 1.37 . Whilst Kaiser’s expression assumes $\Lambda_0 = 0$, in practice a non-zero cosmological constant makes very little difference, merely increasing the amplification factor to 1.39 (Lahav et al. 1991). The value measured for our galaxies at separations of $10h^{-1}$ Mpc, where linear theory is still a fair approximation, is 1.27 ± 0.04 , in reasonable agreement

with the analytic expectation.

It is remarkable that whilst galaxies and dark matter have very different two-point correlation functions in real space, on scales smaller than $3h^{-1}$ Mpc the redshift-space correlation functions of galaxies and dark matter are almost indistinguishable. This implies that galaxies and dark matter have different velocity dispersions on these scales, and suggests that estimates of galaxy bias obtained from clustering in redshift space should be treated with caution.

4.3.1 Galaxy Peculiar Motions

Pairwise velocity statistics for dark matter and galaxies are plotted in Fig. 4.2. The data points in the figure were obtained from the Las Campanas Redshift Survey (hereafter LCRS, Shectman et al. 1996) by Jing, Mo & Börner (1998). Following Jenkins et al. (1998), a quantity that may be compared to these data is the projected, line-of-sight velocity dispersion given by:

$$\sigma_{\text{los}}^2(r_p) = \frac{\int \xi(r) \sigma_{\text{proj}}^2(r) dl}{\int \xi(r) dl}, \quad (4.1)$$

where r_p denotes projected separation, l distance along the line of sight, $r = \sqrt{r_p^2 + l^2}$, and the integral is taken along the line-of-sight between $\pm\infty$, although in practice convergence is attained with integration limits of $\pm 25h^{-1}$ Mpc. The quantity σ_{proj}^2 is the line-of-sight pairwise dispersion, which is given by:

$$\sigma_{\text{proj}}^2 = \frac{r_p^2 \langle v_{\perp}^2 \rangle / 2 + l^2 (\langle v_{\parallel}^2 \rangle - \langle v_{21} \rangle^2)}{r_p^2 + l^2}, \quad (4.2)$$

where $\langle v_{\perp}^2 \rangle^{1/2}$ is the rms relative pairwise velocity perpendicular to the vector connecting each pair, \mathbf{r}_{12} , $\langle v_{\parallel}^2 \rangle^{1/2}$ is the rms relative pairwise peculiar velocity along \mathbf{r}_{12} and $\langle v_{21} \rangle$ is the mean relative pairwise peculiar velocity along \mathbf{r}_{12} , all of which are functions of $r = |\mathbf{r}_{12}|$.

The line-of-sight velocity dispersion of the *dark matter*, shown in the left-hand panel of Fig. 4.2, is significantly larger than the value for LCRS galaxies on small scales. At $r \sim 0.5h^{-1}$ Mpc, the observed value is $535 \pm 100 \text{ km s}^{-1}$, whereas for the dark matter $\sigma_{\text{los}} = 870 \text{ km s}^{-1}$. For dark matter, we find that the measured σ_{los} from the GIF simulation is in good agreement with that obtained by Jenkins et al. (1998) from a simulation of a volume approximately 5 times larger than the GIF simulation and from that in the 512^3 simulation, differing by less than 50 km s^{-1} on all scales considered here.

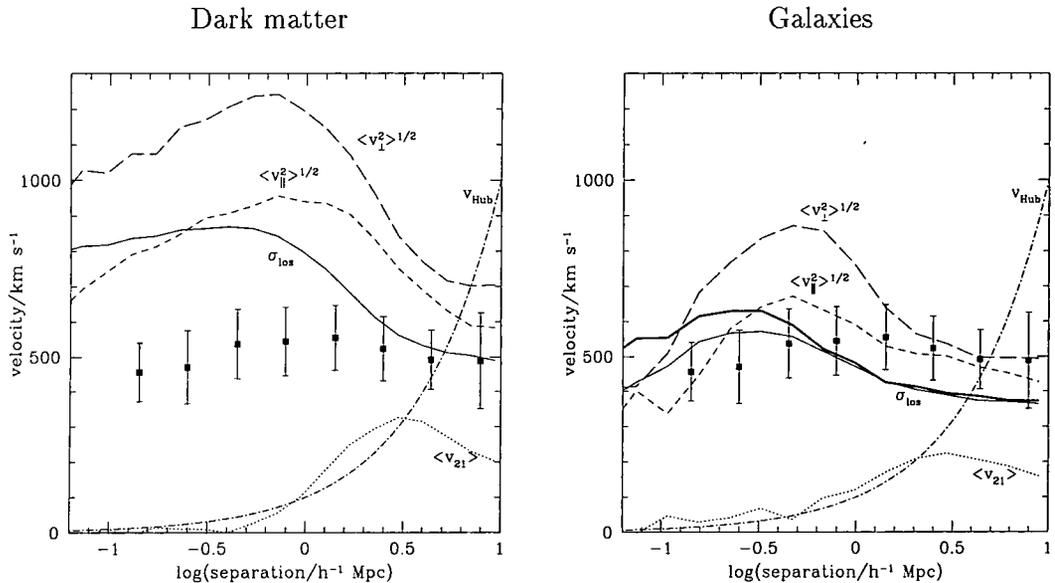


Figure 4.2: Pairwise velocities of dark matter particles (left-hand panel) and galaxies (right-hand panel) measured in the Λ CDM simulation. The dotted line is the mean inward radial peculiar velocity of pairs, $\langle v_{21} \rangle$; the short-dashed line is the rms pairwise radial peculiar velocity, $\langle v_{\parallel}^2 \rangle^{1/2}$; the long-dashed line is the rms pairwise perpendicular peculiar velocity, $\langle v_{\perp}^2 \rangle^{1/2}$; the solid lines are the rms pairwise line-of-sight peculiar velocity, σ_{los} ; and the dot-dash line is the Hubble expansion given by $v_{\text{Hub}} = -H_0 r$, where H_0 is Hubble's constant and r is physical separation. In the right hand panel the thick solid line shows σ_{los} for a sample with $M_R - 5 \log h \leq -21.5$, whilst all other lines are for a sample of galaxies with $M_B - 5 \log h \leq -19.5$. The data points with error bars are taken from Jing, Mo & Börner (1998) and show the pairwise velocity dispersion, σ_{12} , estimated for the Las Campanas redshift survey. The error bars are the square root of the sum in quadrature of the errors derived from the data and 1σ uncertainties inferred from mock catalogues by Jing et al. These points should be compared to the line-of-sight dispersions, σ_{los} , for the models (solid lines). Data points and σ_{los} are plotted against projected separation, r_p , whilst all other curves are plotted against true separation, r .



To compare our results with the LCRS data we have constructed a sample of galaxies brighter than $M_R - 5 \log h = -21.5$ which corresponds approximately to the apparent magnitude limit of the LCRS at its mean depth of $300h^{-1}$ Mpc. The line-of-sight velocity dispersion of these *galaxies*, shown in the right-hand panel by the thick solid line, is in reasonable agreement with the data. A sample selected to have $M_B - 5 \log h \leq -19.5$ has a very similar σ_{los} , as shown by the thin solid line in the right-hand panel.

Fig. 4.2 shows that model galaxies display a markedly different velocity dispersion from the dark matter on all the scales considered. This “velocity bias” is statistical in origin, and is unrelated to dynamical friction in clusters, because each galaxy in our model is assigned the peculiar velocity of a particle belonging to a dark matter halo in the N-body simulation. This effect arises because the galaxy distribution does not constitute a Poisson sampling of the dark matter distribution. Firstly, L_* galaxies are found only in dark matter halos more massive than $\sim 10^{12}h^{-1}M_\odot$, whereas a significant fraction of the dark matter is in smaller mass objects. Secondly, the number of galaxies in each halo is not in direct proportion to the halo mass, as would be required for Poisson sampling (see Chapter 3). The number of galaxies per halo determines how well the velocity structure of the halo is traced by the galaxy distribution. Our results for galaxy velocity dispersions differ somewhat from those found by Kauffmann et al. (1999a) using a similar technique. In §4.3.2 we explore thoroughly the causes of these differences.

The kind of biases seen in the pairwise velocity dispersion function are also present in single galaxy statistics. Fig. 4.3 shows the three-dimensional velocity dispersion of galaxies and dark matter within spheres of radius r , measured relative to the mean velocity within the sphere. At large r , this quantity tends to a constant value, the rms peculiar velocity of individual galaxies. As the figure shows, the dispersion within spheres is lower for galaxies than for dark matter and, furthermore, it depends on galaxy luminosity. For bright galaxies, the velocity dispersion is about 20% lower than for the dark matter. The dependence on luminosity arises because the fainter sample contains relatively more galaxies in the most massive halos which have the largest velocity dispersions. The figure also shows that the sampling variance of the velocity dispersion function is very large, particularly on small scales. For example, for sphere radii of $10h^{-1}$ Mpc, velocities as low as $\sim 100 \text{ km s}^{-1}$ are within $1-\sigma$ of the mean dispersion. This reflects the importance of long wavelength density fluctuations in determining the small-scale velocity field.

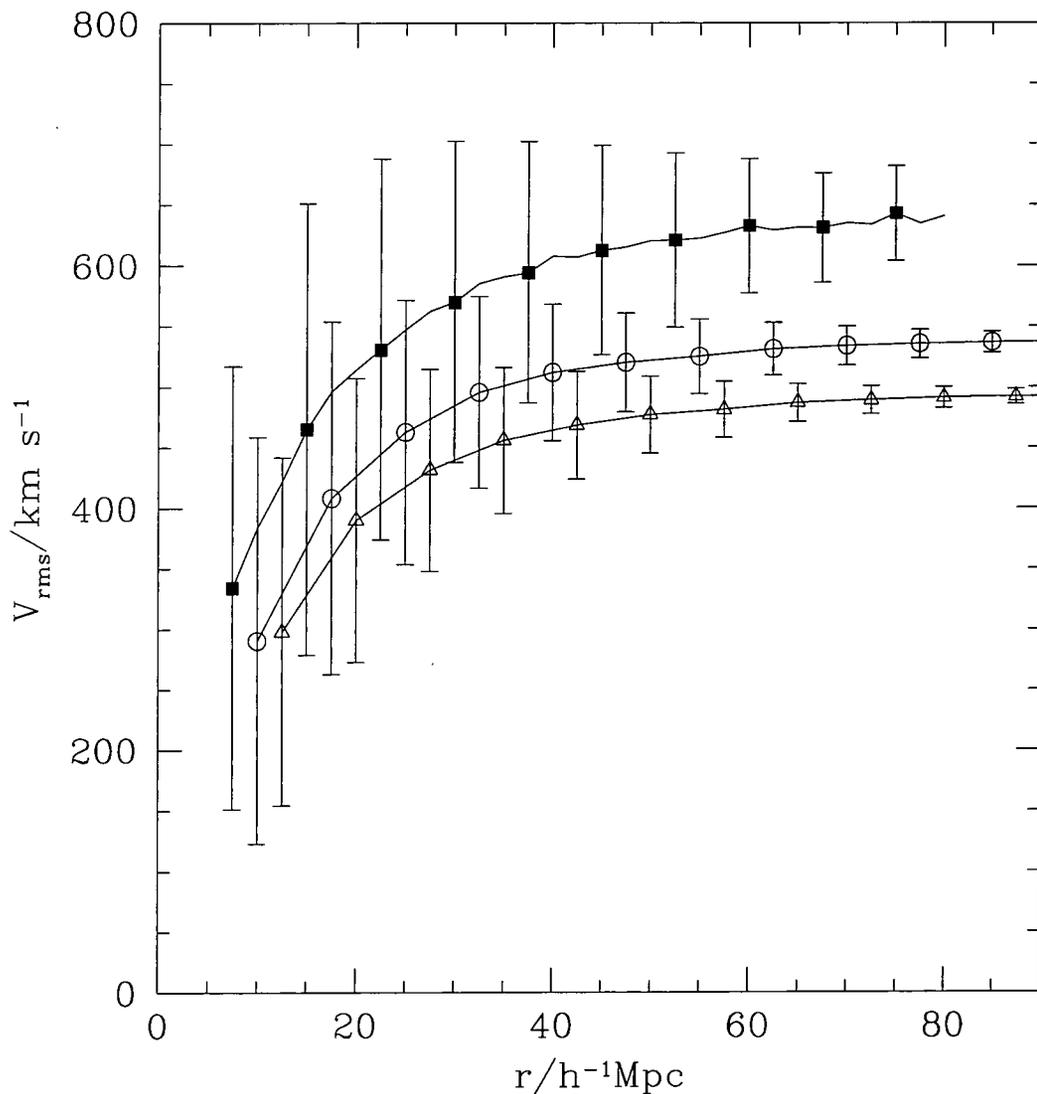


Figure 4.3: The three-dimensional velocity dispersion of individual galaxies and dark matter in spheres of radius r measured relative to the mean velocity within the sphere. Filled squares show the velocity dispersion of the dark matter, whilst open symbols show the velocity dispersion of galaxies brighter than $M_B - 5 \log h = -19.5$ (triangles) and $M_B - 5 \log h = -18.5$ (circles). The symbols indicate the mean velocity dispersion, and the error bars show the *rms* scatter in this quantity.

4.3.2 Comparison to Previous Work

The pairwise galaxy velocity dispersions that we find are somewhat lower than those obtained by Kauffmann et al. (1999a) using exactly the same N-body simulations as us, but a different semi-analytic prescription. Kauffmann et al. found very similar (to within 10%) pairwise velocities for dark matter particles and galaxies on all scales. Thus, unlike ours, their model disagrees with the Jing, Mo & Börner (1998) data. Since the two semi-analytic treatments attempt to model the same physics, it is instructive to understand the reasons for this discrepancy.

The first, and most obvious difference is that Kauffmann et al. (1999a) considered galaxies one magnitude fainter than we do and the pairwise velocities are higher for fainter galaxies (because fainter samples consist of proportionally more satellite galaxies in clusters than brighter cuts). In our model, the line-of-sight velocity dispersion, σ_{los} , of galaxies brighter than $M_{\text{B}} - 5 \log h = -18.5$ is approximately 150 km s^{-1} higher than that for galaxies one magnitude brighter. However, as may be seen in Fig. 4.4, even for galaxies of the same luminosity, there is still a discrepancy between our results and those of Kauffmann et al. There are three effects that could give rise to this discrepancy: (i) different assignments of peculiar velocities and/or positions to galaxies in the same halo; (ii) differences in the frequency with which the two semi-analytic models populate a halo of given mass with a particular number of galaxies; and (iii) statistical variations in the formation histories of galaxies in halos of a given mass. We have carried out several tests of these effects.

The results of the tests are shown in Fig. 4.4. The thick solid line shows σ_{los} for the dark matter in the GIF simulation, whilst the thin solid line shows σ_{los} for galaxies brighter than $M_{\text{B}} - 5 \log h \leq -19.5$ in our model. In order to estimate the variance arising from different galaxy formation histories, we have generated ten realisations of our model using different random number sequences when constructing dark matter merger trees and selecting random halo particles on which to place the galaxies. The variation in σ_{los} between these ten models is an underestimate of the true statistical uncertainty because the population of dark matter halos is, of course, identical in all ten realizations (and to that in the model of Kauffmann et al.). The dashed line shows σ_{los} for galaxies in our model brighter than $M_{\text{B}} - 5 \log h \leq -18.5$, with 1σ error bars derived in this way. Although σ_{los} for these galaxies is larger than for our brighter sample, it is still biased relative to the dark matter and is significantly below the Kauffmann et al. result. Thus,

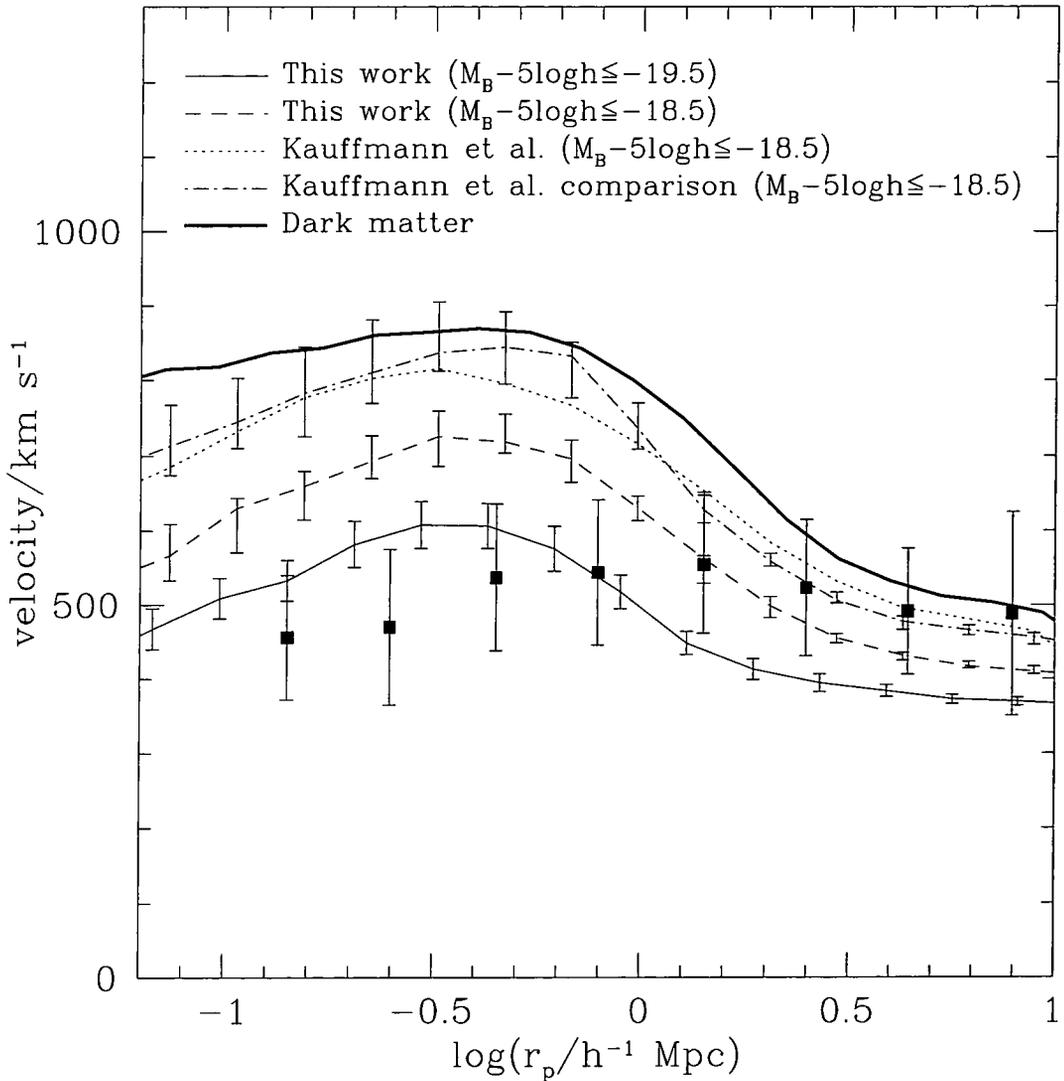


Figure 4.4: Line-of-sight velocity dispersions, σ_{los} for: dark matter (thick solid line); galaxies in our model that are brighter than $M_B - 5 \log h = -19.5$ (thin solid line) and $M_B - 5 \log h = -18.5$ (dashed line); galaxies in the model of Kauffmann et al. (1999a) that are brighter than $M_B - 5 \log h = -18.5$ (dotted line); the Kauffmann et al. galaxies but with positions and velocities assigned using our placement scheme (dot-dashed line). The error bars on these lines reflect sampling uncertainties in the simulated volume as discussed in the text. Squares with error bars are estimated from the Las Campanas Redshift Survey (Jing, Mo & Börner 1998).

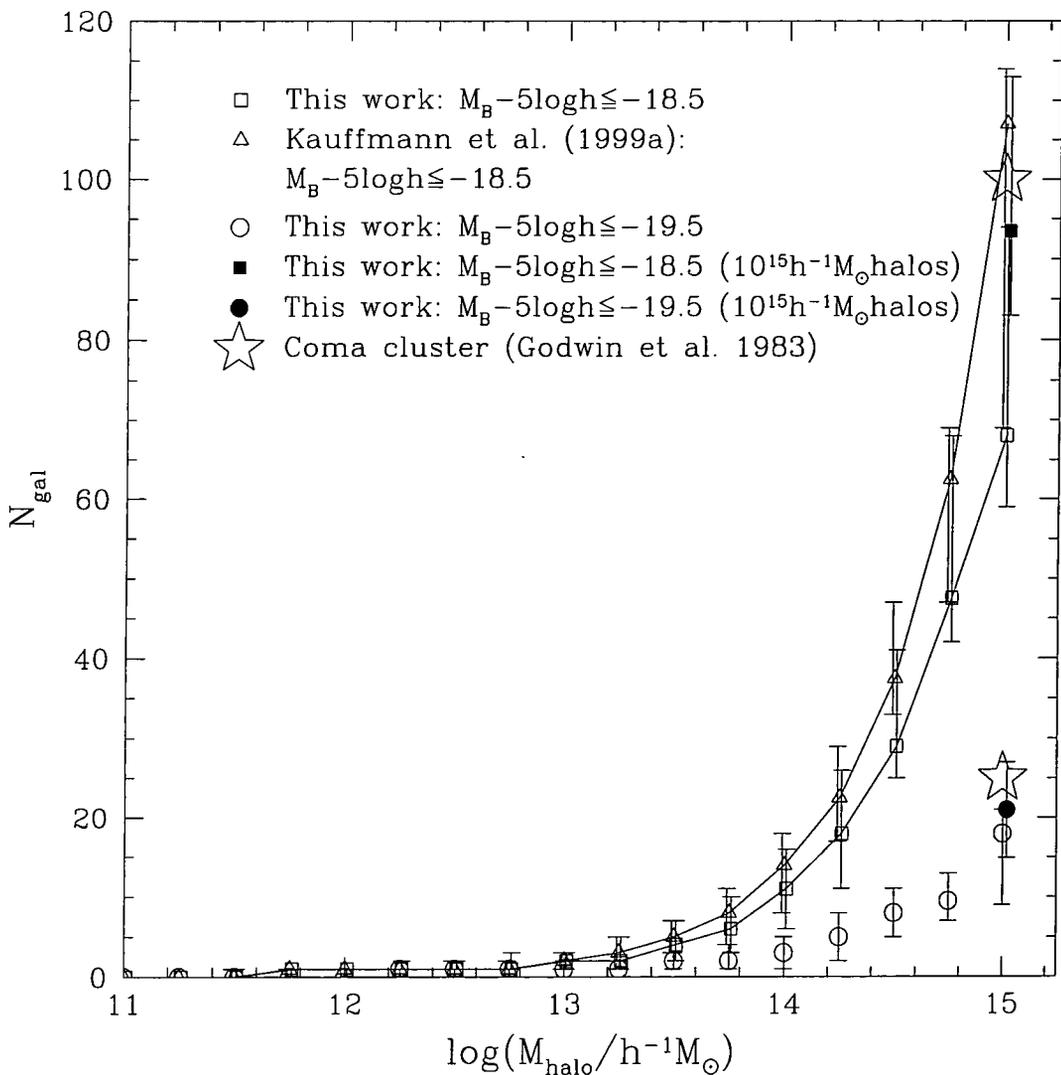


Figure 4.5: The number of galaxies per halo brighter than $M_B - 5 \log h = -18.5$ as a function of halo mass in our model (squares) and that of Kauffmann et al. (1999a) (triangles). Also shown is the number of galaxies per halo brighter than $M_B - 5 \log h = -19.5$ as a function of halo mass in our model (circles). Filled symbols show the corresponding predictions from our model for an independent sample of twenty $10^{15} h^{-1} M_{\odot}$ halos. The symbols show the median of the distribution whilst error bars indicate the 10% and 90% intervals of the distribution. The number of galaxies brighter than $M_B - 5 \log h = -18.5$ and -19.5 within the virial radius of the Coma cluster as found by Godwin, Metcalfe & Peach (1983) are indicated by stars.

statistical uncertainties and varying the magnitude selection, although responsible for relatively large changes in σ_{los} , do not, on their own, explain the difference between the two semi-analytic models.

We now consider differences in the way in which the two models assign positions and velocities to galaxies within a given halo. In our model, each galaxy is assigned the position and velocity of a randomly chosen dark matter halo particle (except the central galaxy which is assigned the position and velocity of the centre of mass of the halo). In the model of Kauffmann et al. (1999a), galaxies initially form on the most bound particle in their halo. If that halo later merges with a larger one, the galaxy becomes a satellite in the new halo and remains attached to the dark matter particle on which it formed. Diaferio et al. (1999) have shown that this placement scheme leads to a dynamical velocity bias, such that blue galaxies in clusters have a higher velocity dispersion than the cluster dark matter because they preferentially occupy particles which are in the process of falling into the cluster for the first time. Such dynamical biases cannot (by construction) arise in our model.

To test the effect of the different placement schemes, we have taken the galaxy catalogue of Kauffmann et al. (limited at $M_{\text{B}} - 5 \log h = -18.5$ and kindly provided by A. Diaferio) and identified the dark matter halo of the GIF simulation to which each of these galaxies belongs. We have then assigned positions and velocities to these galaxies using our own placement scheme. The resulting σ_{los} is shown in Fig. 4.4 as the dot-dashed line (with error bars calculated from ten realisations of the placement scheme). This may be compared to the σ_{los} for the original Kauffmann et al. catalogue shown as the dotted line. The two curves are very similar on small scales and the small differences that there are on scales larger than $1h^{-1}$ Mpc are, in fact, due to our choice of always siting one galaxy at the centre of mass of each halo. (When we place *all* the Kauffmann et al. galaxies on randomly chosen halo particles the two curves agree extremely well on all scales.) This test demonstrates that the differences between our results and those of Kauffmann et al. are not due to the different ways in which galaxies are identified with dark matter halo particles in the two models.

We conclude that the differences between our results and those of Kauffmann et al. must be due to the frequency with which halos of different mass are populated with a particular number of galaxies in the two semi-analytic models. This statistic is shown in Fig. 4.5. Our model and that of Kauffmann et al. agree well for halos below $\sim 10^{13}h^{-1}M_{\odot}$, but for higher masses Kauffmann et al. assign systematically more galaxies

to each halo (typically around 30% more) than we do. Thus, the catalogue of Kauffmann et al. gives greater weight to the most massive, highest velocity dispersion halos to which pairwise statistics are quite sensitive. As a result, they predict significantly higher velocity dispersions than we do. Since the $10^{15}h^{-1}M_{\odot}$ bin contains only three halos in the GIF simulation, we also show our model predictions from an independent sample of twenty halos of mass $10^{15}h^{-1}M_{\odot}$ populated using exactly the same galaxy formation rules (indicated by filled symbols). For comparison, we have indicated in the figure the number of galaxies brighter than $M_B - 5 \log h = -19.5$ found in the Coma cluster and the corresponding number in our model. The virial mass and radius of the Coma cluster are $0.8 \pm 0.1 \times 10^{15}h^{-1}M_{\odot}$ and $1.5h^{-1}\text{Mpc}$ (1.2°) respectively (Geller, Diaferio & Kurtz 1999). Within a circular aperture of this radius centred on NGC4874, Godwin, Metcalfe & Peach (1983) find 100 galaxies brighter than $M_B - 5 \log h = -18.5$ and 25 galaxies brighter than $M_B - 5 \log h = -19.5$ (at the fainter magnitude cut a contamination of approximately 5.5 field galaxies is expected). Although this sample of one cluster cannot be used to confirm our model, it is consistent with our predicted distribution of galaxy numbers but also with that predicted by Kauffmann et al. To differentiate between these two models would require a larger observational sample of rich clusters. Our model has been shown to be in good agreement with the observed total luminosities of groups and clusters determined by Moore, Frenk & White (1993) (Fig. 3.13), indicating that it is populating both groups and clusters with galaxies in a realistic fashion.

Although velocity statistics are sensitive to the numbers of galaxies placed in high mass halos, we have checked that galaxy correlation functions (in both real and redshift-space) are much less affected. For example, the redshift-space correlation function of galaxies brighter than $M_B - 5 \log h \leq -18.5$ differs by only 10% on scales of $10h^{-1}\text{Mpc}$ and by less than 20% on scales of $0.3h^{-1}\text{Mpc}$ when estimated using the number of galaxies per halo predicted by our model and by that of Kauffmann et al. (1999a). This is comparable to the scatter between models with different parameters found in Chapter 3.

Fig. 4.6 shows the effects of varying key model parameters on the pairwise velocity dispersions. In §4.4.1 we describe these models further and explain our criteria for rejecting models as being non-realistic. Of the fifteen variants considered, five would be classed as non-realistic, and are shown as dashed lines. These models are outliers compared with the viable models (solid lines) which have only a small scatter around the fiducial model. This demonstrates that robust predictions for velocity dispersions can be made once non-realistic models are excluded.

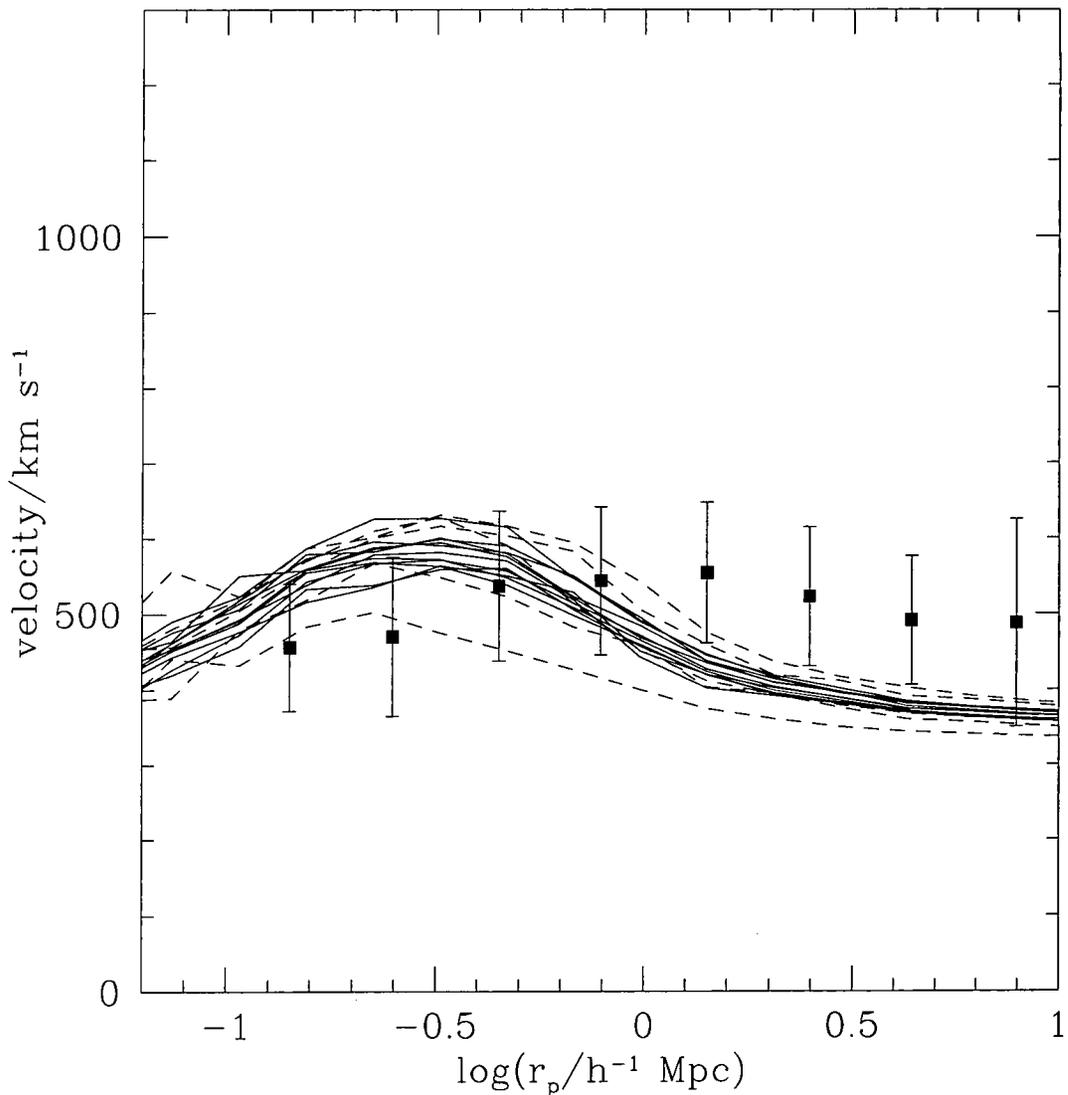


Figure 4.6: The line-of-sight pairwise velocity dispersion, σ_{los} , for galaxies with $M_B - 5 \log h \leq -19.5$. The lines show fifteen variants of our fiducial model in which some of the key parameters have been varied. The five models shown as dashed lines would be rejected as being non-realistic, as discussed in §4.4.1. The data points with error bars are taken from Jing, Mo & Börner (1998) and show the pairwise velocity dispersion, σ_{12} , estimated for the Las Campanas redshift survey.

4.4 The Dependence of Clustering on Galaxy Properties

We now examine the clustering properties of various subsamples extracted from our original galaxy catalogue, selecting by luminosity, morphology or colour.

4.4.1 Dependence of Clustering on Luminosity

In Fig. 4.7 we show our predicted redshift-space correlation functions for four cuts in absolute magnitude, and compare them with the corresponding correlation functions measured in the ESP redshift survey by Guzzo et al. (1999). (The space density of galaxies in our model at each of these magnitude cuts matches closely that found in the ESP survey, since our b_j -band luminosity function is tuned to agree with the ESP survey — c.f. Fig 4.11.) The error bars on the model are estimates of the sample variance, computed in the following manner. For each magnitude cut, we extract fifty randomly placed cubes from our simulation which have a volume equal to that of the corresponding volume-limited samples drawn from the ESP survey. For the ESP survey volume-limited at $M_{b_j} - 5 \log h \leq -18.5$, -19.5 and -20.5 , our simulation volume contains 46, 17 and 6 independent survey volumes respectively. Thus, for the brighter samples our estimate of the sample variance is likely to be an underestimate of the true value. In the figure, we plot the median correlation function of these fifty realisations, and error bars showing the 10% and 90% intervals of the distribution. As can be seen, the model results are in excellent agreement with the data over the whole range of scales and magnitudes shown.

Guzzo et al. (1999) argue that for the brightest magnitude cut they considered, $M_{b_j} - 5 \log h \leq -20.5$, there is evidence for stronger clustering than for galaxies with $M_{b_j} - 5 \log h \leq -19.5$. Unfortunately, this effect is most noticeable on scales somewhat larger than those we can probe reliably with our simulation. At $r \sim 10h^{-1}\text{Mpc}$, the largest scale on which we can reliably measure the correlation function, we find that galaxies with $M_{b_j} - 5 \log h \leq -20.5$ have a clustering amplitude that is only 1.07 ± 0.10 times greater than that of galaxies with $M_{b_j} - 5 \log h \leq -19.5$ (when measured from the full volume of our simulation). Thus, the results from the GIF simulation are inconclusive. We have therefore made use of the 512^3 simulation, described in §4.2, to examine the clustering of very bright galaxies. The correlation function of galaxies brighter than $M_{b_j} - 5 \log h = -20.5$ in this simulation agrees with that of the same galaxies in the GIF simulation, within the rather large errorbars (due to the limited number of such bright galaxies in the GIF volume), except on the largest scales where the different dark

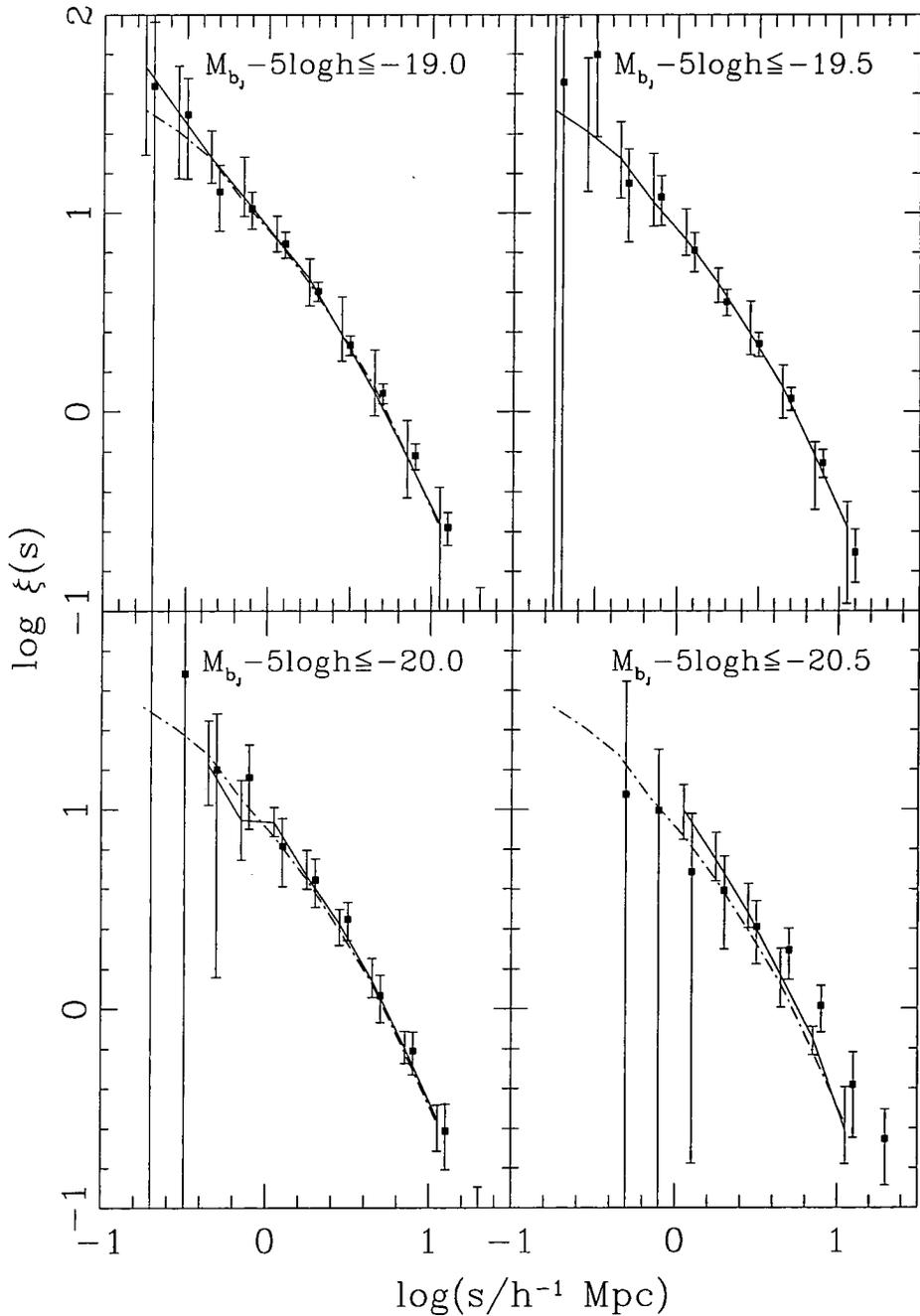


Figure 4.7: Redshift-space correlation functions of galaxies selected by b_J -band absolute magnitude (see panels for magnitude cuts). The lines show the median model correlation function obtained from fifty randomly chosen volumes equal in size to the corresponding ESP sample. The error bars on these lines indicate the 10% and 90% intervals of the distribution of correlation functions. The model result for $M_{b_J} - 5 \log h \leq -19.5$ is reproduced, for reference, in each panel as the dot-dashed line. The filled squares with error bars show the correlation functions estimated from the volume-limited samples of the ESP (Guzzo et al. 1999).

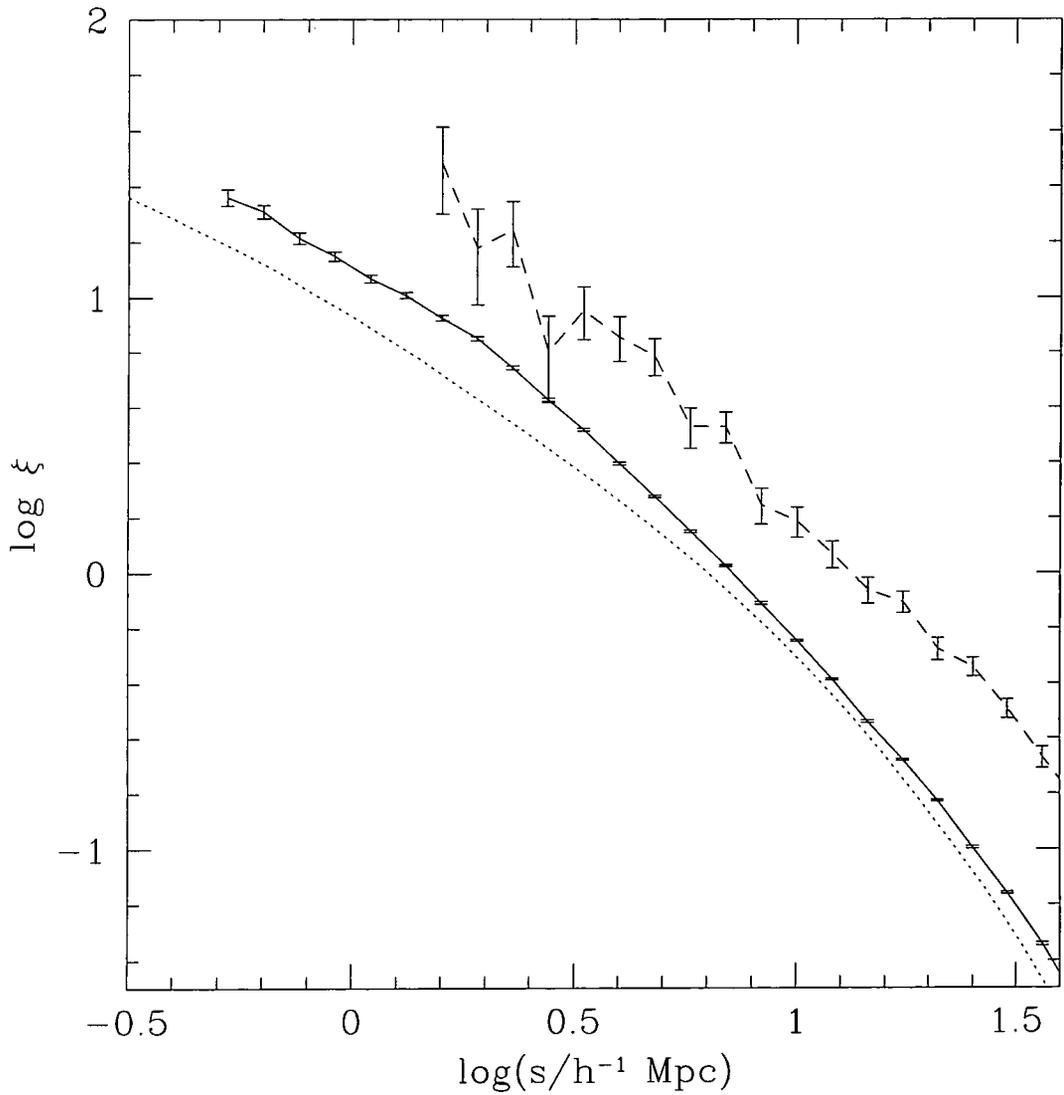


Figure 4.8: Redshift-space correlation functions of galaxies brighter than $M_{b_j} - 5 \log h = -20.5$ (solid line) and -21.5 (dashed line) measured in the 512^3 simulation. The redshift-space correlation function of dark matter in this simulation is shown by the dotted line.

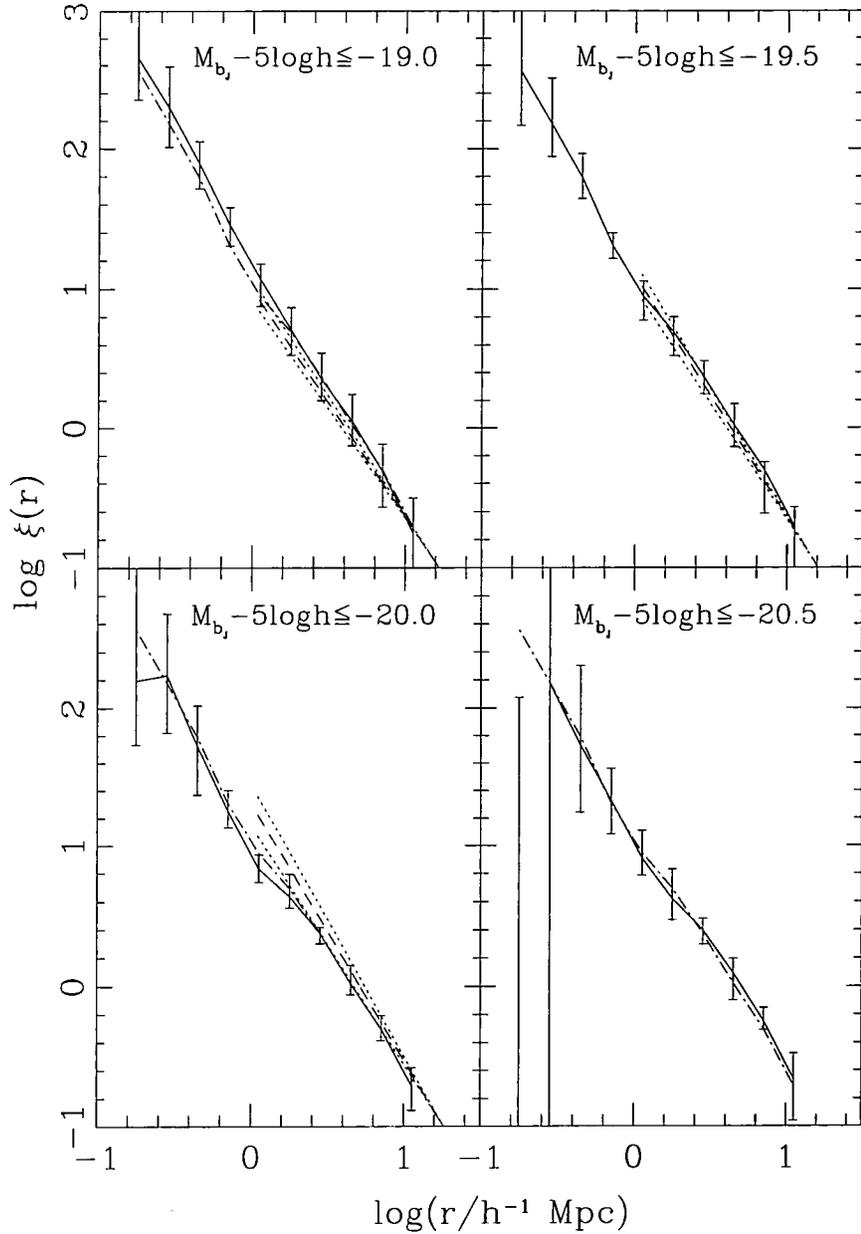


Figure 4.9: Real-space correlation functions of galaxies selected by b_J -band absolute magnitude (see panels for magnitude cuts). The dashed lines show the power-law fits to the corresponding ESP survey correlation functions obtained by Guzzo et al. (1999). (Note that no fit was given for the brightest magnitude cut shown.) Dotted lines track the quoted errors on these fits. The solid line in each panel shows the median model correlation function obtained by splitting our sample into fifty randomly positioned cubes with volume equal to that of the ESP survey at the same magnitude cut. Error bars on this line indicate the 10% and 90% intervals of the distribution from these same samples. The model result for $M_{b_J} - 5 \log h \leq -19.5$ is reproduced, for reference, in each panel as the dot-dashed line.

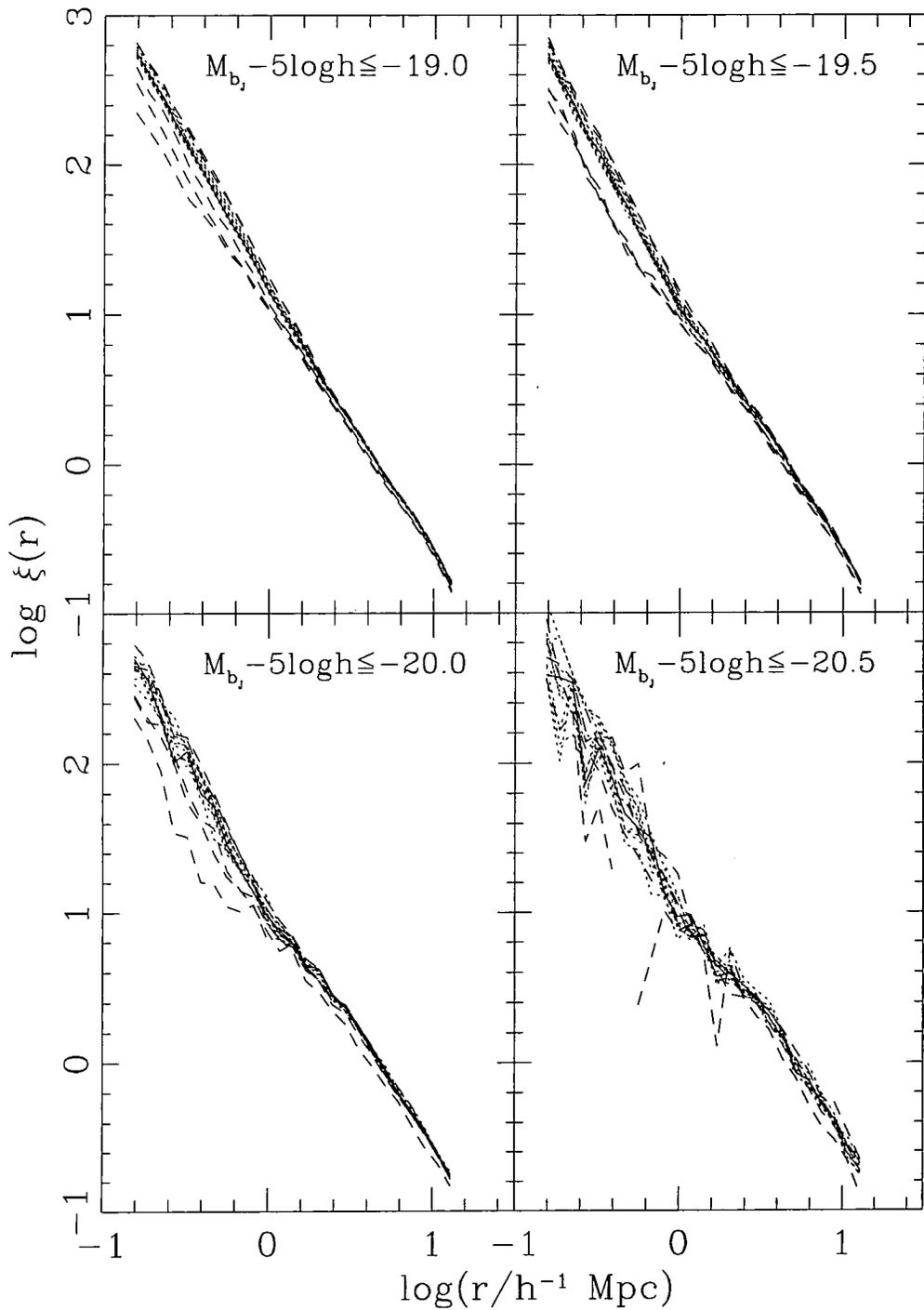


Figure 4.10: Real-space correlation functions of galaxies selected by b_J -band absolute magnitude (see panels for magnitude cuts). The fifteen variant models are plotted either as dotted lines if they produce a correlation function similar to the standard model or as dashed lines if they do not.

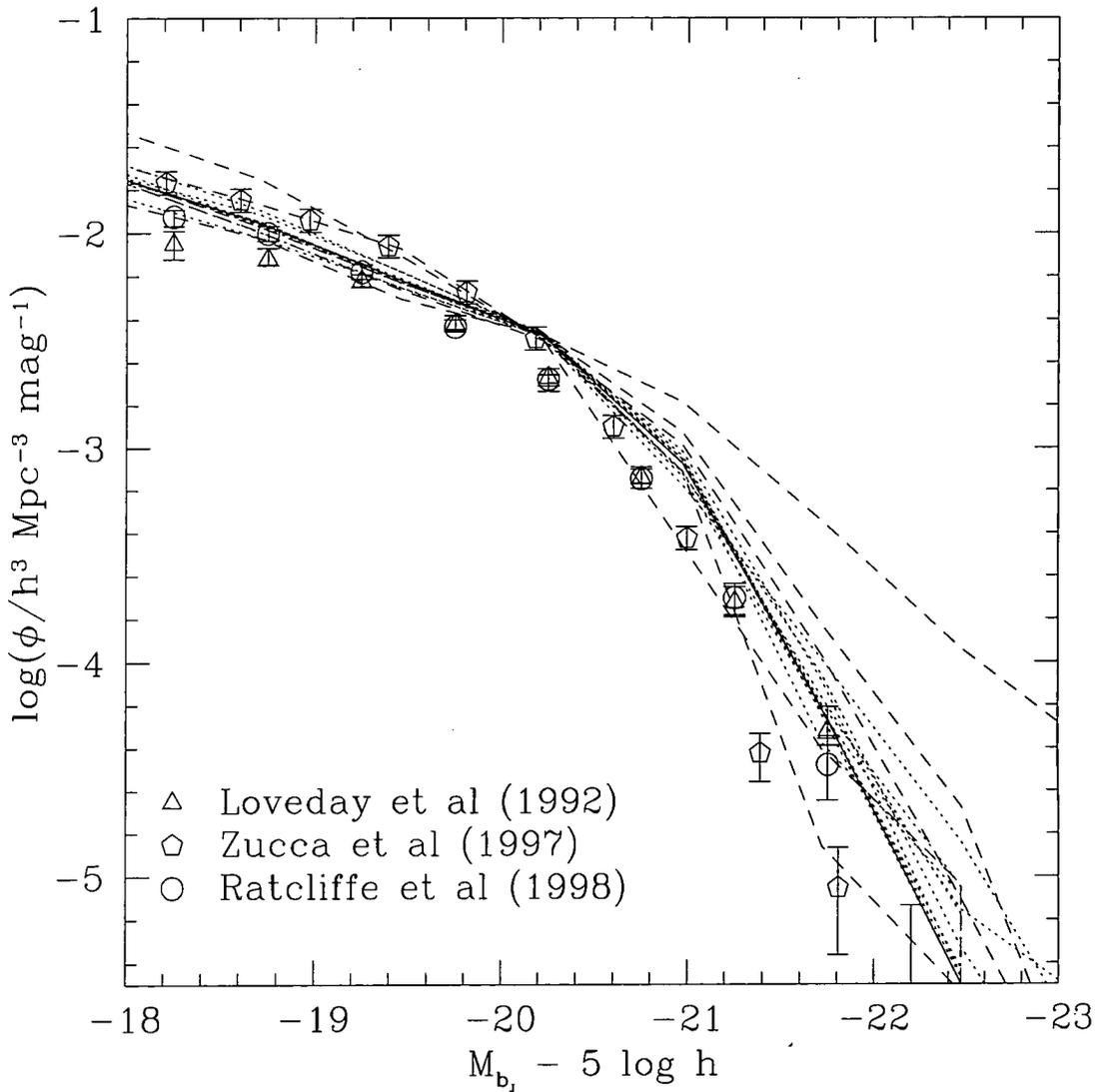


Figure 4.11: The local b_J -band luminosity function. Symbols with error bars show various observational estimates of this quantity, as indicated in the legend. The lines show fifteen variant models along with our standard model (solid line). The five models which show significantly different correlation functions are identified by dashed lines.

matter correlation functions in the simulations (as discussed in §4.2) introduce a similar difference in the galaxy correlation functions. As shown in Fig. 4.8, we find that, in the 512^3 simulation, galaxies with $M_{\text{b}_J} - 5 \log h \leq -21.5$ have a clustering amplitude on scales above $10h^{-1}\text{Mpc}$ which is approximately 4 times higher than that of galaxies with $M_{\text{b}_J} - 5 \log h \leq -20.5$. The space densities of $M_{\text{b}_J} - 5 \log h \leq -20.5$ and -21.5 galaxies are approximately 6 and 200 times lower than that of L_* galaxies respectively. The large increase in clustering amplitude is due to the fact that these very bright galaxies occupy halos with a mean mass of several M_* , for which the halo bias is a rapidly changing function of halo mass (Mo & White 1996).

In Fig. 4.9 we show the real-space correlation function for our models applying the same four absolute magnitude cuts. Guzzo et al. (1999) estimated the real-space correlation function by fitting a power-law to the correlation function measured in terms of projected separation. We plot their results over the range of scales used in their fit, $0.5 \leq r \leq 9.5h^{-1}\text{Mpc}$. Within the errors estimated from sample variance, our models are in reasonable agreement with the measurements. Guzzo et al. claim to find evidence for a weak luminosity dependence in the real-space correlation function, with galaxies in the brightest magnitude cut being the most strongly clustered. This effect is manifest as an increase in the correlation length with intrinsic luminosity obtained from their power-law fits to the correlation function. The slope of the ESP correlation function also increases for intrinsically brighter galaxies. In contrast, we find no significant evidence for a luminosity dependence of either clustering length or the slope of the correlation function in our models. The pair-counting errors that Guzzo et al. estimated for the fitted parameters are largest for the brightest sample, which contains the fewest galaxies. Our model correlation functions, however, indicate that sample variance is particularly significant for the *fainter* catalogues, which sample smaller volumes. Taking this sample variance into account weakens the significance of the effect claimed by Guzzo et al.

To confirm that our model predictions for $\xi(r)$ are robust, we follow Chapter 3 and consider the effect on the correlation function of galaxies of varying key parameters in our semi-analytic model, for example, the galaxy merger rate or the baryon fraction. Each model is constrained to match the ESP b_J -band luminosity function at L_* by adjusting the parameter Υ , which gives the ratio of the total mass in stars (including brown dwarfs) to that in luminous stars. The parameter changes used are similar to those listed in Table 3.4. Note that we do not consider the “No dust” model here.

In Fig. 4.10 we show the real-space correlation functions obtained from the fifteen

models with altered parameter values. Of the fifteen variants considered, five exhibit significantly different clustering properties to the fiducial model. As can be seen in Fig. 4.11, these models are amongst the ones that disagree the most with the observed luminosity function, where these models are shown as dashed lines, and would therefore be rejected as unrealistic models (see end of §4.3.2). This highlights again the point emphasized in Chapter 3 that a good fit to the galaxy luminosity function is a pre-requisite for a robust determination of clustering statistics in semi-analytic models. The same five models were shown as dashed lines in Fig. 4.6, where they show a greater variation in velocity dispersions than models which are a good fit to the luminosity function. This shows that reproducing the bright end of the luminosity function is also important in order to make robust predictions for galaxy velocity statistics.

4.4.2 Dependence of Clustering on Morphology

In Fig. 4.12, we compare the real-space correlation functions for galaxies selected by morphological type in our model with observational data from Loveday et al. (1995). Although we restrict attention to galaxies in the same magnitude range used by Loveday et al., our criteria for morphological classification are inevitably somewhat different. Loveday et al. classified galaxies by eye as either early types (E/SO) or late types (S/Irr). Our classification is based on the bulge-to-total luminosity ratio of each galaxy as measured in dust-extincted b_J -band light (B/T). We classify galaxies with $B/T < 0.4$ as late type galaxies and those with $B/T > 0.4$ as early type galaxies. The observed correlation between bulge-to-total luminosity ratio and morphological type displays a large scatter (see Fig. 1 of Baugh, Cole & Frenk 1996b).

Bearing these caveats in mind, the level of agreement between the observed correlation functions and the model predictions shown in Fig. 4.12 is encouraging. Early type galaxies are the most strongly clustered on all the scales we consider, as found also by Kauffmann et al. (1999a). The greatest difference occurs on scales below $1h^{-1}$ Mpc, where early type galaxies have a clustering amplitude that is more than 4 times higher than that of late type galaxies. On larger scales, the difference in clustering amplitudes persists, but is less pronounced.

In Fig. 4.13, we plot the host halo mass functions of late and early type galaxies in our model (weighted by the number of galaxies per halo). As may be seen, early type galaxies do, on average, reside preferentially in higher mass environments than late type galaxies, as speculated by Loveday et al. (1995). This explains the larger correlation

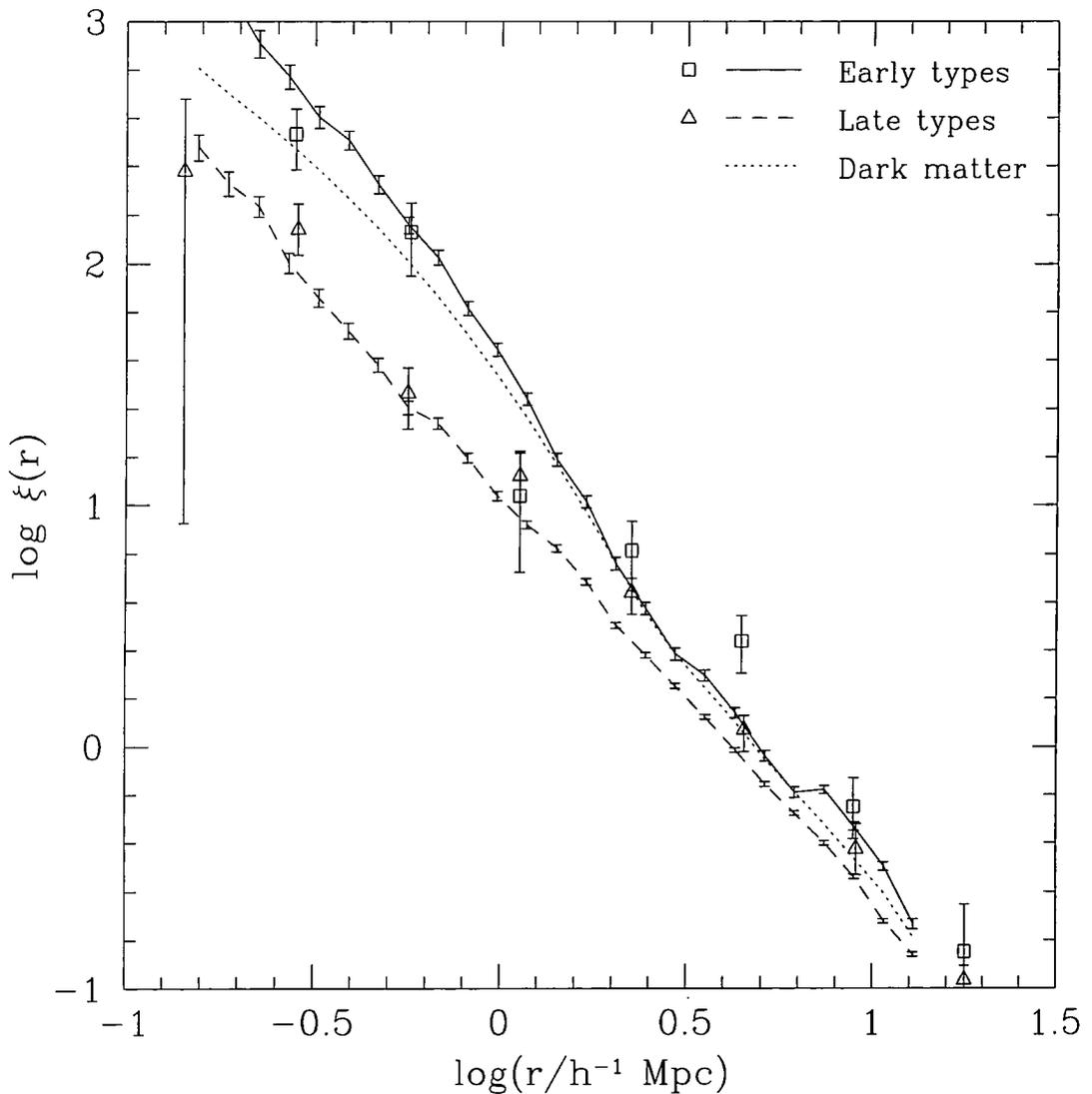


Figure 4.12: Real-space two-point correlation function for galaxy samples selected by morphological type. Squares and triangles with error bars show the observed correlation functions of early and late type galaxies respectively, with $-20.0 < M_{b_J} - 5 \log h < -19.0$, from Loveday et al. (1995). Solid and dashed lines show correlation functions for model galaxies in the same magnitude range. This sample is subdivided by morphological type as described in the text. For the model, the lines indicate the correlation function from the whole simulation volume, which is approximately 1.3 times larger than the $M_{b_J} - 5 \log h < -19.0$ volume-limited Stromlo-APM survey. The dotted line shows the correlation function of the dark matter.

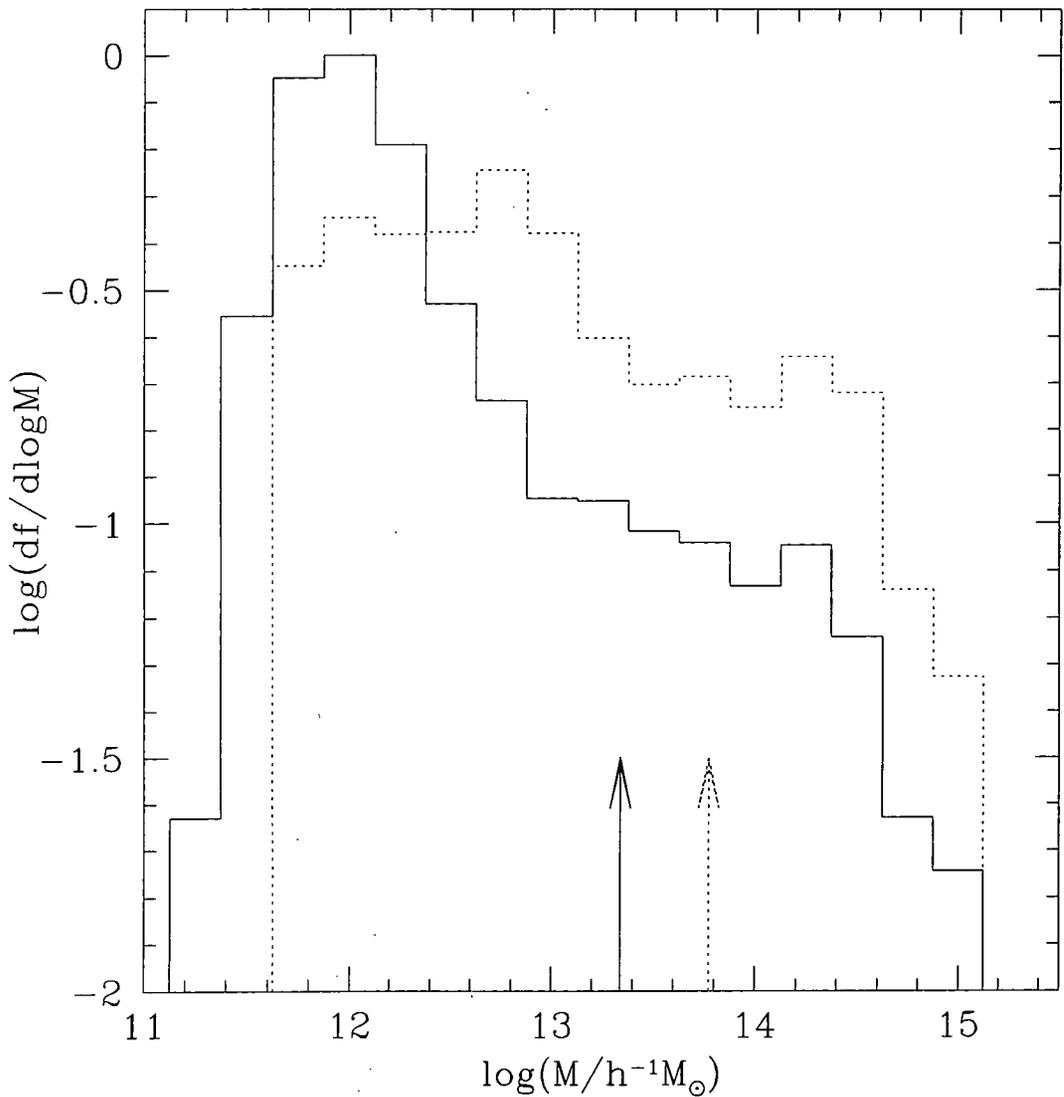


Figure 4.13: The mass function of host dark matter halos, weighted by the number of galaxies with $-20.0 < M_{bJ} - 5 \log h < -19.0$ per halo, for late type galaxies (solid line) and early type galaxies (dotted line) in our model. The mass functions are normalised by the total number of galaxies in each sample. The arrows indicate the mean host halo mass for each population of galaxies.

length of the early types since higher mass halos are intrinsically more strongly biased (Frenk et al. 1988, Mo & White 1996) relative to the dark matter distribution as a whole. It also explains the greater correlation amplitude of early type galaxies on scales below $1h^{-1}$ Mpc, where the correlation function is sensitive to the number of galaxy pairs inside cluster sized halos as shown in Chapter 3. Late type galaxies are rare in these halos and so their pair count on these scales is small, resulting in a lower correlation amplitude.

4.4.3 Dependence of Clustering on Colour

At present, the only available observational constraints on the colour dependence of galaxy clustering come from deep, magnitude-limited samples that cover a significant baseline in redshift. The technique used in this Chapter needs to be extended to include evolution both in the galaxy population and in the clustering of dark matter to produce realistic mock catalogues that can be compared with these observations. However, forthcoming redshift surveys, such as the Two-degree Field (2dF) (Colless 1996) and Sloan Digital Sky Survey (SDSS) (Gunn & Weinberg 1995), will contain large enough numbers of galaxies with photometry in different bands to allow the construction of local volume-limited samples subdivided by galaxy colour.

We present, in Fig. 4.14, our predictions for colour dependent clustering of galaxies in real space, measured in the full volume of our simulation. Red galaxies are defined to have $B-R > 1.3$ and blue galaxies $B-R \leq 1.3$, which matches the cut used by Willmer, da Costa & Pellegrini (1998). The error bars show the Poisson pair counting errors in each bin. The clustering of red galaxies is significantly stronger than that of blue galaxies, which follows from the fact that the red galaxies are predominantly early types, and so are predicted to reside preferentially within higher mass halos than the blue galaxies.

4.5 Discussion and Conclusions

In this Chapter, we have extended the analysis of Chapter 3 to examine the peculiar motions of galaxies, their effect on clustering and the clustering of different types of galaxies.

In Chapter 3, we studied the correlation function of galaxies brighter than L_* in *real-space* (i.e. as a function of true spatial separation; see also Kauffmann et al. 1999a). We found remarkably good agreement between the predictions of the Λ CDM model and measurements of the correlation function of the APM galaxy survey (Baugh 1996), over

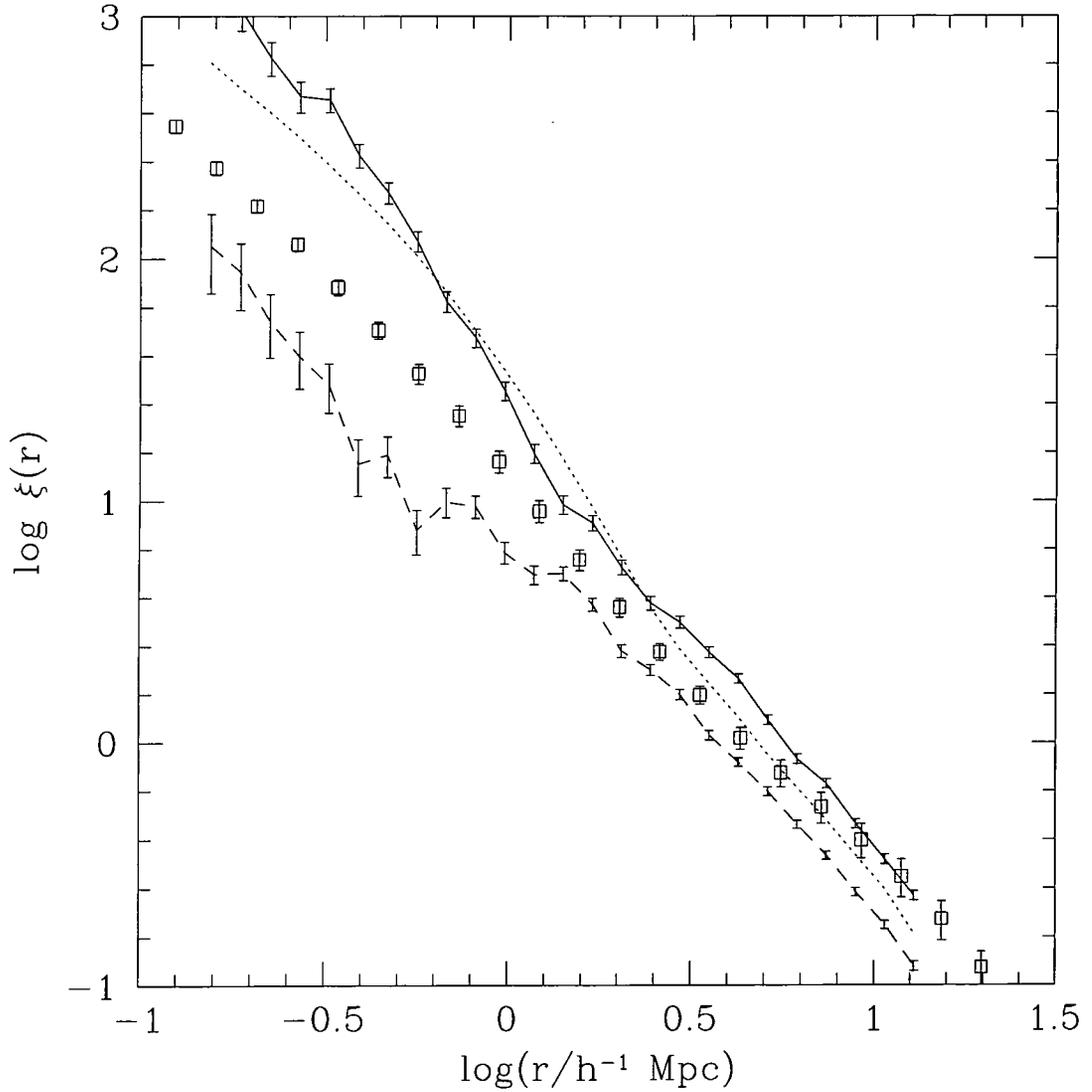


Figure 4.14: The real-space correlation function of galaxies brighter than $M_B - 5 \log h = -19.5$ in our model selected by B-R colour. The solid line is the correlation function of galaxies with $B-R > 1.3$, whilst the dashed line is that of galaxies with $B-R \leq 1.3$. Error bars on these lines show the Poisson pair counting errors. The dotted line shows the correlation function of the dark matter, and the points with error bars are the real-space correlation function measured from the APM survey for all galaxies (Baugh 1996).

four orders of magnitude in correlation amplitude. On small scales, we found the model galaxies to be *less strongly* clustered than (or biased low relative to) the dark matter. We have now seen that the net effect of small-scale peculiar velocities is to cancel out this bias: in redshift space the galaxy and dark matter correlation functions are very similar to one another. This cancellation arises because the pairwise velocity dispersions of galaxies and dark matter are different. Fortuitously, the difference is just sufficient to compensate for the differences in real-space clustering. Thus, although genuinely biased, the distribution of galaxies measured in redshift space appears, to a good approximation, unbiased on small scales. The pairwise velocity dispersion of the model galaxies is $\sim (200 - 300)$ km/s lower than that of the dark matter over almost two decades in spatial separation and is in good agreement with recent observational determinations (Jing, Mo & Börner 1998). However, comparison of our results with those of Kauffmann et al. (1999a) demonstrates the sensitivity of this statistic to the number of galaxies that populate rich clusters. Both models are consistent with the available data for the Coma cluster, but produce line-of-sight velocity dispersions which differ by around 100 km/s.

As discussed in Chapter 3, the physical origin of the offset between the galaxy and dark matter velocity distributions lies in the way in which galaxies sample the velocity field of the dark matter. The mass-to-light ratio of halos in our model is a strong, non-monotonic function of halo mass. Galaxy formation is most efficient in halos of mass $\sim 10^{12}h^{-1}M_{\odot}$, and the mass-to-light ratio increases at higher and lower masses due to long cooling times for the gas and stronger feedback respectively (see Fig. 3.9). As a result, the number of galaxies per halo does not increase as rapidly as the halo mass. Thus, when computing the pairwise velocity dispersion, high velocity dispersion halos are undersampled by galaxies relative to the contribution of these halos to the velocity dispersion of the dark matter itself.

We have explored the sensitivity of our theoretical predictions for peculiar velocities to variations in the galaxy formation parameters. In Chapter 3 we found that predictions for the galaxy two-point correlation function in a particular cosmological model are robust to changes in model parameters, provided that the galaxy luminosity function remains approximately the same. Here, we find a similar result for the peculiar velocities: models with similar luminosity functions produce similar results. The amplitude of the velocity bias depends not only on the luminosity function, but also on the shape of the power spectrum of density fluctuations and on the statistics of the occupation of halos by galaxies. As we showed in Chapter 3, our (cluster-normalized) Λ CDM model gives a

good match to the observed luminosity function of galactic systems. To a large extent, this is the reason why the model also gives a good match to the observed pairwise velocity dispersion function (although our neglect of dynamical biases in the galaxy distribution may affect our results slightly). The differences between our predicted peculiar velocities and those obtained by Kauffmann et al. (1999a) from the same N-body simulations, but using a different semi-analytic model, are simply due to differences in the way in which the two models populate high mass halos. These, in turn, reflect differences in the luminosity functions of galaxies and of galactic systems in the two models.

The dependence of clustering on intrinsic galaxy properties provides, in principle, an interesting test of models of galaxy formation. The most obvious property to consider is galaxy luminosity. Unfortunately, the dependence of clustering on luminosity is difficult to measure from magnitude limited redshift surveys and so the observational situation is inconclusive. A weak effect has been claimed, for example, by Guzzo et al. (1999) for galaxies approximately one magnitude brighter than L_* . For the bulk of the galaxy population, we find no evidence for a dependence of clustering on the intrinsic luminosity of our model galaxies. This is not surprising in view of the fact that galaxies of a given luminosity reside in dark matter halos spanning an appreciable range of masses. We do, however, predict a strong effect for galaxies that are significantly brighter than L_* . Unfortunately, the space density of these galaxies is too low for their clustering to be accurately measured in existing redshift surveys. On the other hand, both models and observations agree that early type or red galaxies are more strongly clustered than late type or blue galaxies, particularly on small scales. This is a reflection of the observed morphology-density relation which arises naturally in hierarchical clustering models (Baugh, Cole & Frenk 1996b, Kauffmann 1996b).

It is worth emphasizing that the galaxy formation model that we have used in this Chapter is essentially the same as that discussed at length in Cole et al. (2000) and adopted in Chapter 3. Cole et al. fixed parameter values by requiring their model to reproduce a variety of properties of the local galaxy population, with emphasis placed on achieving a good match to the local b_J -band galaxy luminosity function. This same Λ CDM model also reproduces the present day clustering of galaxies in real and redshift space, is in reasonable agreement with the observed line-of-sight pairwise velocity dispersion of galaxies, and matches the clustering measured at intermediate and high redshifts (Baugh et al. 1999).

An important conclusion of this work is that the statistical properties of the galaxy

distribution can be quite different from those of the underlying dark matter. A physical approach to modelling the formation and evolution of galaxies, using the kind of techniques discussed in this Chapter and also in Kauffmann et al. (1999a) and Chapter 3, is therefore imperative if we are to understand the biases in the way that different galaxies trace the dark matter. This, in turn, is a pre-requisite for making sense of the unprecedented amount of information about the galaxy distribution in the low and high redshift Universe that will shortly become available from forthcoming redshift surveys.

Chapter 5

Non-Uniform Reionization by Galaxies and its Effect on the CMB

5.1 Introduction

The Gunn-Peterson (GP) effect (Gunn & Peterson 1965) indicates that the smoothly distributed hydrogen in the intergalactic medium (IGM) is already highly ionized by $z = 5$ (Schneider, Schmidt & Gunn 1991, Lanzetta, Wolfe & Tunshek 1996). Barring the possibility of collisional reionization (e.g. Giroux & Shapiro 1994), the GP effect implies the presence of very luminous ionizing sources at high redshifts capable of producing enough Lyman continuum (Lyc) photons to cause photoionization of hydrogen by $z \gtrsim 5$. The two possible sources of these ionizing photons are QSOs and high mass stars.

Models in which QSOs dominate the production of ionizing photons may be able to meet the GP constraint (Miralda-Escudé & Ostriker 1990). However, such models are strongly constrained by the observed decline in the abundance of bright QSOs above $z \approx 3$ (Hartwick & Schade 1990, Warren, Hewett & Osmer 1994, Kennefick, Djorgovski & de Carvalho 1995, Schmidt, Schneider & Gunn 1995). (This result remains somewhat controversial as the number of QSOs upon which it is based is relatively small.) Furthermore, Madau, Haardt & Rees (1999) note that a model in which faint QSOs provide all the required ionizing luminosity can be ruled out on the basis of the number of faint QSOs seen in the HDF.

There is, however, growing evidence for the presence of bright galaxies at redshifts as high as $z \sim 5$ (Spinrad 1998), and perhaps even higher (Yahil, Lanzetta & Fernandez-Soto 1998). Thus the other natural candidate sources of ionizing photons are young, high

mass stars forming in galaxies at redshifts greater than 5 (e.g. Couchman & Rees 1986, Haiman & Loeb 1996, Ciardi et al. 2000). Madau, Haardt & Rees (1999) note that at $z \approx 3$ stars in Lyman-break galaxies will emit more ionizing photons into the IGM than QSOs if more than 30% of such photons can escape from their host galaxy. Whilst such high escape fractions may not be realistic (e.g. local starbursts show escape fractions of only a few percent, Leitherer et al. 1995), this does demonstrate that high-redshift galaxies could provide a significant contribution to (or perhaps even dominate) the production of ionizing photons. In this Chapter we will restrict our attention to ionizing photons produced by stars, although a contribution from QSOs could easily be fitted into the same scheme (see, for example, Kauffmann & Haehnelt 2000).

According to the hierarchical structure formation scenario (e.g. Peebles 1980) perturbations in the gravitationally dominant and dissipationless dark matter grow, by gravitational instability, into virialised clumps, or halos. Galaxies, and later stars, then form by the cooling and condensation of gas inside these halos (e.g. White & Rees 1978, White & Frenk 1991). Dark matter halos continually grow by merging with other halos (e.g. Bower 1991, Bond et al. 1991). In the context of this hierarchical scenario, we present a realistic scheme for studying the reionization of the universe by ionizing photons emitted from massive stars. We focus on the photoionization of the hydrogen component of the IGM. To predict the time dependent luminosity in Ly α photons we use the semi-analytic model of galaxy formation described in Chapter 2. This is modified to take into account Compton cooling by cosmic microwave background (CMB) photons. We use this model to determine the properties of galaxies living in dark matter halos spanning a wide range of masses.

We then estimate the fraction of the ionizing photons which manage to escape each galaxy, and therefore contribute to the photoionization of the intergalactic HI. The fraction of ionizing photons escaping is determined on a galaxy-by-galaxy basis, using physically motivated models. Assuming spherical symmetry, we follow the propagation of the ionization front around each halo to compute the filling factor of intergalactic HII regions, including the effects of clumping in the IGM. Finally, using several alternative models for the spatial distribution of ionized regions within a high resolution N-body simulation of the dark matter distribution, we estimate the anisotropies imprinted on the cosmic microwave background (CMB) by the patchy reionization process, due to the correlations in the ionized gas distribution and velocities (Sunyaev & Zel'dovich 1980, Vishniac 1987). In previous models many simplifications were made in computing both the spatial distribu-

tion of ionized regions and the two-point correlations of gas density and velocity in those regions (Aghanim et al. 1996, Jaffe & Kamionkowski 1998, Gruzinov & Hu 1998, Knox, Scoccimarro & Dodelson 1998, Peebles & Juskiewicz 1998, Haiman & Knox 1999). Our calculations represent a significant improvement over these models as we are able to calculate the two-point correlations between gas density and velocity in ionized regions directly from an N-body simulation.

The rest of this Chapter is arranged as follows. In §5.2 we outline the features of the semi-analytic model relevant to galaxy formation at high redshifts. In §5.3 we describe how we calculate the fraction of ionizing photons escaping from galaxies, and observational constraints on the ionizing luminosities and escape fraction at low and high redshift from $H\alpha$ luminosities and HI masses and column densities. In §5.4 we describe how we calculate the filling factor of photoionized gas in the IGM, including the effects of clumping of this gas. We then present our predictions for reionization, including the effects on the reionization redshift of using different assumptions about escape fractions and clumping factors. In §5.5 we examine the robustness of our results to changes in the other parameters of the semi-analytic galaxy formation model. In §5.6 we describe how the semi-analytic models are combined with N-body simulations to calculate the spatial distribution of the photoionized IGM. We then calculate the spectrum of anisotropies introduced into the CMB by this ionized gas. Finally, in §5.7 we summarize our results and examine their consequences.

5.2 The Semi-Analytic Model of Galaxy Formation

To determine the luminosity in ionizing Ly α photons produced by the galaxy population, we use the semi-analytic model of galaxy formation described in Chapter 2. The parameters of our standard model for this work are the same as those of the Λ CDM model of Chapter 2, except that we adopt the cosmological parameters of the N-body simulation used in §5.6 and adjust ϵ_* and Υ to obtain an improved match to the $z = 0$ b $_J$ -band luminosity function. For reference, these parameters are given in Table 5.1. Definitions of the semi-analytic model parameters can be found in Chapter 2.

As we are employing the semi-analytic model at much higher redshifts than we have previously attempted, we will investigate the effects on our results of changing key model parameters. Of particular interest will be the prescription for feedback from supernovae and stellar winds. Recall that the model assumes that a mass $\beta\Delta M$ of gas is reheated by

Table 5.1: The parameters of our standard model.

Parameter	Value	Parameter	Value
<i>Cosmology</i>		<i>Stellar populations</i>	
Ω_0	0.3	IMF	Kennicutt (1983)
Λ_0	0.7	p	0.02
H_0	70 km/s/Mpc	R	0.31
σ_8	0.90	Υ	1.53
Γ	0.21	<i>Mergers and bursts</i>	
Ω_b	0.02	f_{dt}	1.0
<i>Gas cooling</i>		f_{ellip}	0.3
Gas profile [¶]	CDC	Starbursts	included
r_{core}	$0.33r_{\text{NFW}}$	f_{dyn}	1.0
Recooling*	not allowed	<i>Ionizing luminosities</i>	
<i>Star formation and feedback</i>		S_2^\ddagger	10.0×10^{50} photons/s
α_*	-1.5	$h_z/r_{\text{disk}}^\ddagger$	0.1
ϵ_*	0.01		
α_{hot}	2.0		
V_{hot}	150 km s^{-1}		
Feedback**	standard		

¶ Gas profiles that we consider are CDC (for which the gas density is $\rho(r) \propto (r^2 + r_{\text{core}}^2)^{-1}$ — i.e. an isothermal profile but with a constant density core), and SIS (an isothermal profile with no core).

* Gas ejected from galaxies by supernovae is allowed to cool again in the same halo if ‘Recooling’ is allowed. Otherwise this gas can only cool again once it enters a newly formed halo.

** ‘Standard’ feedback is the form specified in eqn. (5.1). The alternative is ‘modified’ feedback, which is the form specified in eqn. (5.9).

‡ The upper cut off the in luminosity function of OB associations assumed in the DS94 model (see Appendix B.1 and §5.3.2).

† This is the ratio of disk vertical and radial scale lengths used in the DS94 and DSGN98 models for the gas escape fraction (see Appendix B.1 and §5.3.2).

supernovae and ejected from the disk for each mass ΔM of stars formed. The quantity β is allowed to be a function of the galaxy properties, and is parameterised as

$$\beta = (V_{\text{disk}}/V_{\text{hot}})^{-\alpha_{\text{hot}}}, \quad (5.1)$$

where V_{disk} is the circular velocity of the galaxy disk and V_{hot} and α_{hot} are adjustable parameters of the model. As discussed in Chapter 2, Cole et al. (2000) show that α_{hot} and V_{hot} are well constrained by the shape of the B-band luminosity function and the Tully-Fisher relation at $z = 0$. However, since there is very little time available for star formation at the high redshifts which we consider, it is possible that these parameters, or indeed the form of the parameterisation in eqn. (5.1), could be different at high redshift and this would not necessarily affect the model predictions at $z = 0$. In §5.5 we will therefore experiment with different values of these parameters and will also consider a modified functional form for β .

Two other key model inputs are the baryon density parameter, Ω_b , and the stellar initial mass function (IMF). The value of Ω_b determines cooling rates (and so star formation rates) in our model halos. The shape of the IMF determines the number of high mass stars which produce the ionizing photons. For Ω_b our standard value is 0.02, which is consistent with the estimate of Walker et al. (1991) and which allows a good match to the bright end of the observed B-band luminosity function. We will also consider an alternative value of $\Omega_b = 0.04$, which is in better agreement with estimates from the D/H ratio in QSO absorption line systems (Schramm & Turner 1998, Burles & Tytler 1998). For the IMF, we adopt as our standard choice the IMF of Kennicutt (1983), which is close to the “best” IMF proposed by Scalo (1998) on the basis of observations in the Solar neighbourhood and in nearby galaxies. We consider the effects of changing both Ω_b and the IMF in §5.5.

5.2.1 Gas Cooling

The standard semi-analytic model presented in Chapter 2 allows hot halo gas to cool only via collisional radiative processes. At high redshift, Compton cooling due to free electrons in the hot plasma scattering off CMB photons becomes important. The Compton cooling timescale is given by (Peebles 1968)

$$t_{\text{Compton}} = \frac{1161.3(1 + x_{\text{hot}}^{-1})}{(1 + z)^4(1 - T_0^{\text{CMB}}(1 + z)/T_e)} \text{Gyr}, \quad (5.2)$$

where x_{hot} is the ionized fraction of the hot halo gas, T_0^{CMB} is the temperature of the CMB at the present day and T_e is the temperature of electrons in the hot halo gas, which

we set equal to the virial temperature of the halo. At high redshifts this cooling time becomes shorter than the Hubble time and so Compton cooling may be effective at these redshifts. To implement eqn. (5.2) in the semi-analytic model, we assume that the shock-heated halo gas is in collisional ionization equilibrium, and use values for x_{hot} which we interpolate from the tabulated values given by Sutherland & Dopita (1993). In halos with virial temperatures less than around 10^4K collisional ionization is ineffective, and so the ionized fraction in the halo gas will equal the residual ionization fraction left over from recombination. However, since this fraction is small (Seager, Sasselov & Scott 1999), we will simply assume in this Chapter that cooling in halos below 10^4K is negligible.

It should be noted that, unlike the radiative cooling time, the Compton cooling time is independent of the gas density and depends only very weakly on the gas temperature. Whereas with collisional radiative cooling a cooling radius, within which the cooling time is less than the age of the halo, propagates through the halo as more and more gas cools, with Compton cooling the entire halo cools at the same rate. The amount of gas which can reach the centre of the halo is then controlled by the free-fall timescale in the halo.

Including Compton cooling in our model turns out to make little difference to the results. For example, the total mass of stars formed in the universe as a function of redshift differs by less than 5% for $z < 20$ between models with and without Compton cooling. At higher redshifts the differences can become as large as 30% for small intervals of redshift. For example, if a massive halo cools via Compton cooling it will rapidly produce many stars — without Compton cooling it will still form these stars, but not until slightly later when collisional radiative cooling takes place. However, the mass of stars formed at these high redshifts is tiny, and so any differences become entirely negligible at lower redshifts when many more stars have formed. Although at high redshift the Compton cooling time is shorter than the age of the Universe, halos are merging at a high rate and so their gas is being repeatedly shock-heated by successive mergers which, we assume, heat the gas to the virial temperature of the new halo. We find that for the majority of halos, the time between successive major mergers (defined as the time for a doubling in mass of the halo) is less than the Compton cooling time at the redshifts considered here. Therefore, Compton cooling will be ineffective in these halos. In the few cases where the halo does survive long enough that Compton cooling could be important, we find that the collisional radiative cooling time at the virial radius of the halo is often shorter than the Compton cooling time, in which case all of the gas in the halo will cool whether or not we include the effects of Compton cooling. Nevertheless, Compton cooling is included in all models

considered.

In this work we ignore cooling due to molecular hydrogen (H_2). Although molecular hydrogen allows cooling to occur in gas below 10^4K , it is easily dissociated by photons from stars that form from the cooling gas. Previous studies that have included cooling due to H_2 (e.g. Gnedin & Ostriker 1997, Chiu & Ostriker 1999, Ciardi et al. 2000, Valageas & Silk 1999) typically find that it is completely dissociated at very high redshifts. For example, Ciardi et al. (2000) find that molecular hydrogen is fully dissociated by $z \approx 25$. Objects formed by H_2 cooling are therefore not expected to contribute significantly to the reionization of the IGM.

5.2.2 Fraction of Gas in the IGM

At any given redshift, some fraction of the gas in the Universe will have become collisionally ionized as a result of shocks in dark matter halos and some fraction will have cooled to become part of a galaxy. Within the context of our semi-analytic model, we define the IGM as all gas which has *not* been collisionally ionized inside dark matter halos and which has *not* become part of a galaxy (note that we are here only interested in ionization of hydrogen). It is this gas which must be photoionized if the Gunn-Peterson constraint is to be satisfied. The fraction of the total baryon content of the universe which is in the IGM, f_{IGM} , can be estimated by integrating over the mass function of dark matter halos, as follows

$$f_{IGM}(z) = 1 - \left[\int_0^\infty M_{gas} x_H \frac{dn}{dM_{halo}} \frac{dM_{halo}}{\Omega_b \rho_c} \right] - f_{galaxy}(z), \quad (5.3)$$

where $M_{gas}(M_{halo}, z)$ is the mean mass of diffuse gas in halos of mass M_{halo} , $x_H(M_{halo}, z)$ is the fraction of hydrogen which is collisionally ionized at the halo virial temperature (which we take from the calculations of Sutherland & Dopita 1993), $dn/dM_{halo}(M_{halo}, z)$ is the comoving number density of halos (which we approximate by the Press-Schechter mass function), ρ_c is the critical density of the Universe at $z = 0$, and f_{galaxy} is the fraction of the total baryonic mass in the Universe which has been incorporated into galaxies. The quantities M_{gas} and f_{galaxy} can be readily calculated from our model of galaxy formation. Fig. 5.1 shows the evolution of f_{IGM} with redshift.

5.2.3 Observational Constraints & Modelling Considerations

The semi-analytic model provides the spectral energy distribution (SED) of each galaxy, from which we can determine the ionizing luminosity of that galaxy. Summing the contri-

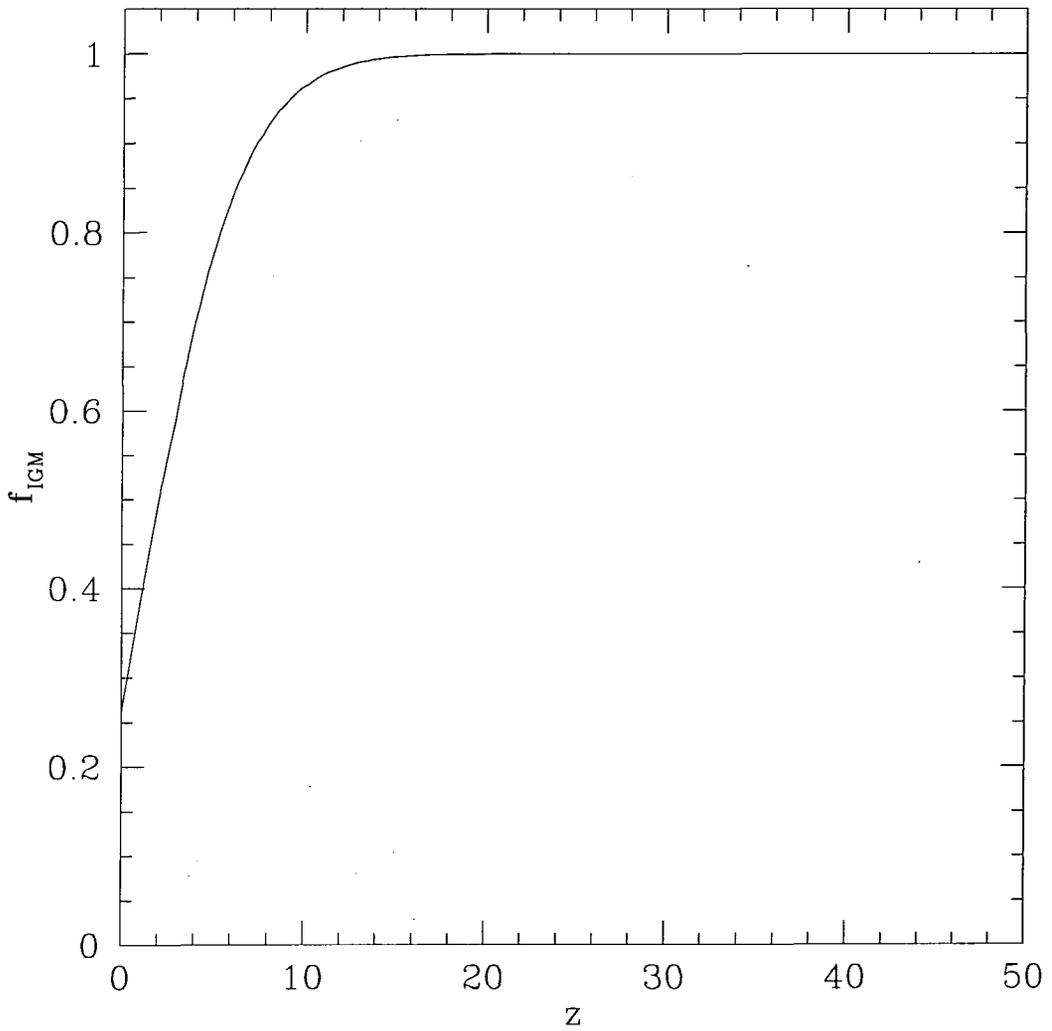


Figure 5.1: The fraction of baryons remaining in the IGM as a function of redshift. Other baryons have either been collisionally ionized in dark matter halos or have cooled to become part of a galaxy.

butions from all galaxies in a given halo yields the total ionizing luminosity produced in that halo. Cojazzi et al. (1999), using the model of Haiman & Loeb (1996), demonstrated that a higher reionization redshift could be obtained if zero-metallicity stars were responsible for reionization, as these produce a greater number of ionizing photons than low (i.e. 10^{-4}) metallicity stars. In our model the very first stars have zero metallicity, but once we include chemical evolution, only a very small fraction of stars have metallicities below 10^{-4} . This is consistent with the results of Tumlinson & Shull (2000) who argue that the epoch of metal-free star formation must end before $z = 3$, as the enhanced emission shortwards of 228\AA from such stars is inconsistent with observations of HeII opacity in the IGM at that redshift. Therefore we cannot appeal to such zero-metallicity stars to increase the redshift of reionization in our model.

As a result of absorption by neutral hydrogen close to the emitting stars and extinction caused by dust, only a small fraction of the ionizing radiation emitted by the stars escapes from each galaxy (Leitherer et al. 1995, Hurwitz, Jelinsky & Dixon 1997, Kunth et al. 1998). We therefore estimate, within the context of the semi-analytic model, the fraction, f_{esc} , of ionizing photons which escape the galaxy to become available for the photoionization of the HI in the IGM. The calculation of f_{esc} is discussed in §5.3.

In Fig. 5.2 we show the redshift evolution of the comoving number density of galaxies with ionizing luminosity \dot{n}_{ion} larger than 10^2 , 10^3 and 10^4 in units of 10^{50} photons per second. These are the unattenuated luminosities produced by massive stars in the galaxies. The abundances of sources of given luminosity rises sharply up to $z = 2 - 4$ (the exact position of the peak depending on luminosity) as more and more dark matter halos form that are capable of hosting bright galaxies. After $z = 2 - 4$ abundances quickly drop towards $z = 0$ as the amount of gas available for star formation declines.

The ragged nature of the curves in the figure (and in other figures in this Chapter) is an artifact of the merger tree construction algorithm used in the semi-analytic model. This algorithm does not accurately reproduce the Press-Schechter mass function when extrapolated over a significant redshift range (see §7.2.2). Since the redshift range we are interested in here is very large this would lead to unacceptably large errors in our calculations. To circumvent this problem we restart our calculations at frequent intervals in redshift, forcing the model to adopt the correct Press-Schechter mass function at each restart. It is this restarting which leads to discontinuities in the curves. It should be noted that this procedure does not completely solve the problem of the merger tree algorithm, since the trees generated still do not match perfectly the extended Press-Schechter

predictions.

The escape fractions in our model will be determined by the mass and radial scale length of the HI gas in galactic disks. It is therefore important to test that our model produces galaxies with reasonable distributions of HI mass and disk scale length. As shown in Chapter 2, our semi-analytic model produces distributions of I-band disk scale lengths in good agreement with the $z = 0$ data of de Jong & Lacey (1999). In Fig. 5.3 we compare our model with observations of damped Lyman- α systems (DLAS) over a range of redshifts and with the HI mass function at $z = 0.0$. Kauffmann (1996a, see also Mo, Mao & White 1998) has shown, in a particular semi-analytic model of galaxy formation, that the observed abundance of DLAS can be accounted for by gas in the rotationally supported disks of high redshift galaxies. However, it should be noted that Haehnelt, Steinmetz & Rauch (1998) have presented evidence that DLAS may in fact be caused by irregular, protogalactic clumps. Using simulations of gas hydrodynamics they find that such clumps of gas can adequately reproduce the observed velocity widths and absorption asymmetries of DLAS.

Under the assumption that DLAS are caused by neutral gas in galactic disks, we compute the DLAS column density distribution in our model, f_{DLAS} , defined such that $f_{\text{DLAS}}(N_{\text{HI}}, t)dN_{\text{HI}}dX$ is the mean number of DLAS at cosmic time t with column densities in the range N_{HI} to $N_{\text{HI}} + dN_{\text{HI}}$ and absorption distance $X(z) = \frac{2}{3} [(1+z)^{3/2} - 1]$ in the interval dX along a line of sight (Lanzetta, Wolfe & Tunshek 1996). Our model is in reasonable agreement with the distribution of DLAS column densities observed by Lanzetta, Wolfe & Tunshek (1996), indicating that both the mass of HI and its radial scale length in our model galaxies are realistic. The $z = 0$ HI mass function from our model is also in reasonable agreement with the data of Zwaan et al. (1997), although it does overpredict the abundance of low HI mass galaxies. Our model predictions assume that all of the hydrogen in galactic disks is in the form of HI. In practice, some of the hydrogen in disks will be in the form of molecules (H_2) or ionized gas (HI), so this over-estimates the HI masses and column densities.

A significant contribution to the ionizing luminosity comes from very low mass halos. We therefore ensure that we resolve all halos that have a virial temperature $\geq 10^4\text{K}$ up to $z = 50$, i.e. all halos down to a mass of $5 \times 10^6 h^{-1} M_{\odot}$. Below this temperature, cooling becomes inefficient (since we are ignoring cooling by molecular hydrogen, and the Compton cooling from the residual free electrons left over after recombination) and so galaxy formation ceases.

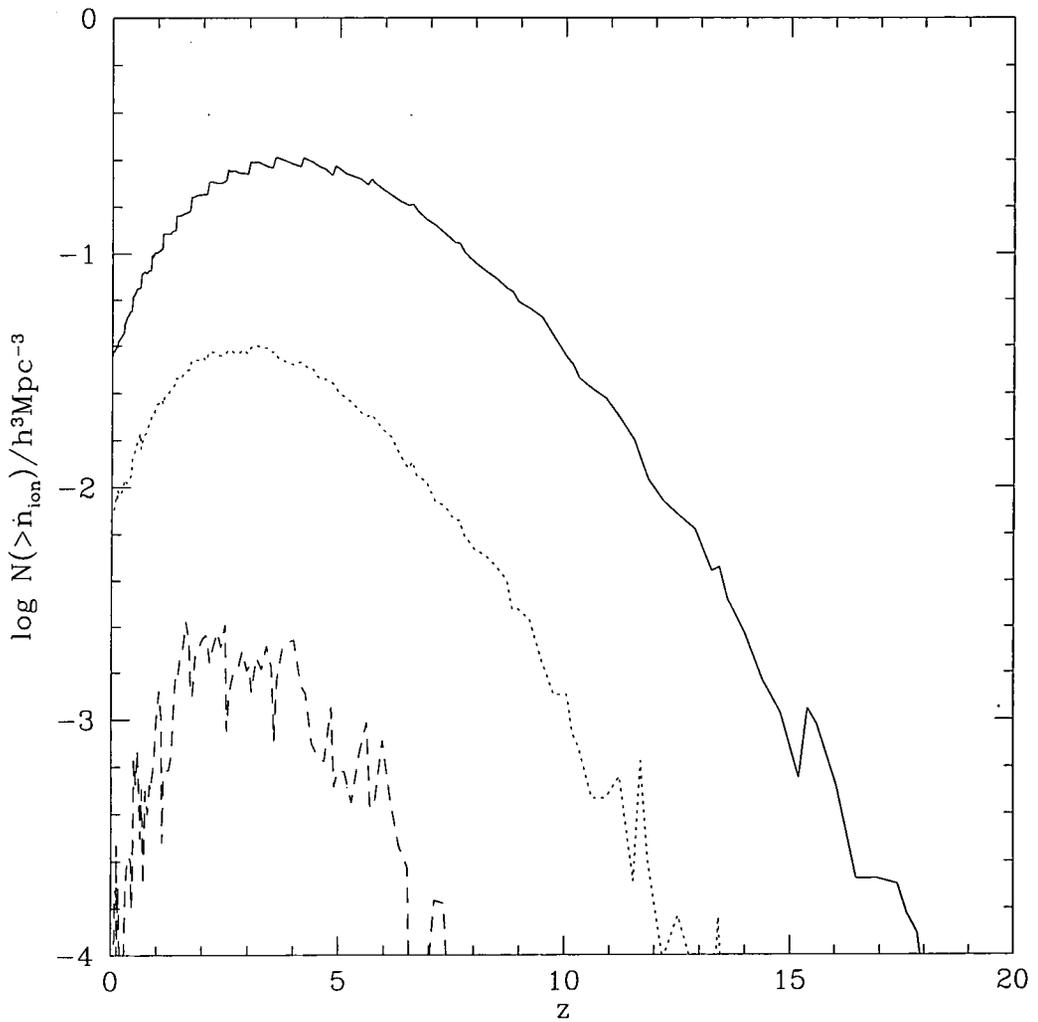


Figure 5.2: The comoving number density of galaxies brighter than a given ionizing luminosity, \dot{n}_{ion} , as a function of redshift. The number is plotted for $\dot{n}_{\text{ion}} = 10^2$ (solid line), 10^3 (dotted line) and 10^4 (dashed line) in units of 10^{50} photons/s.

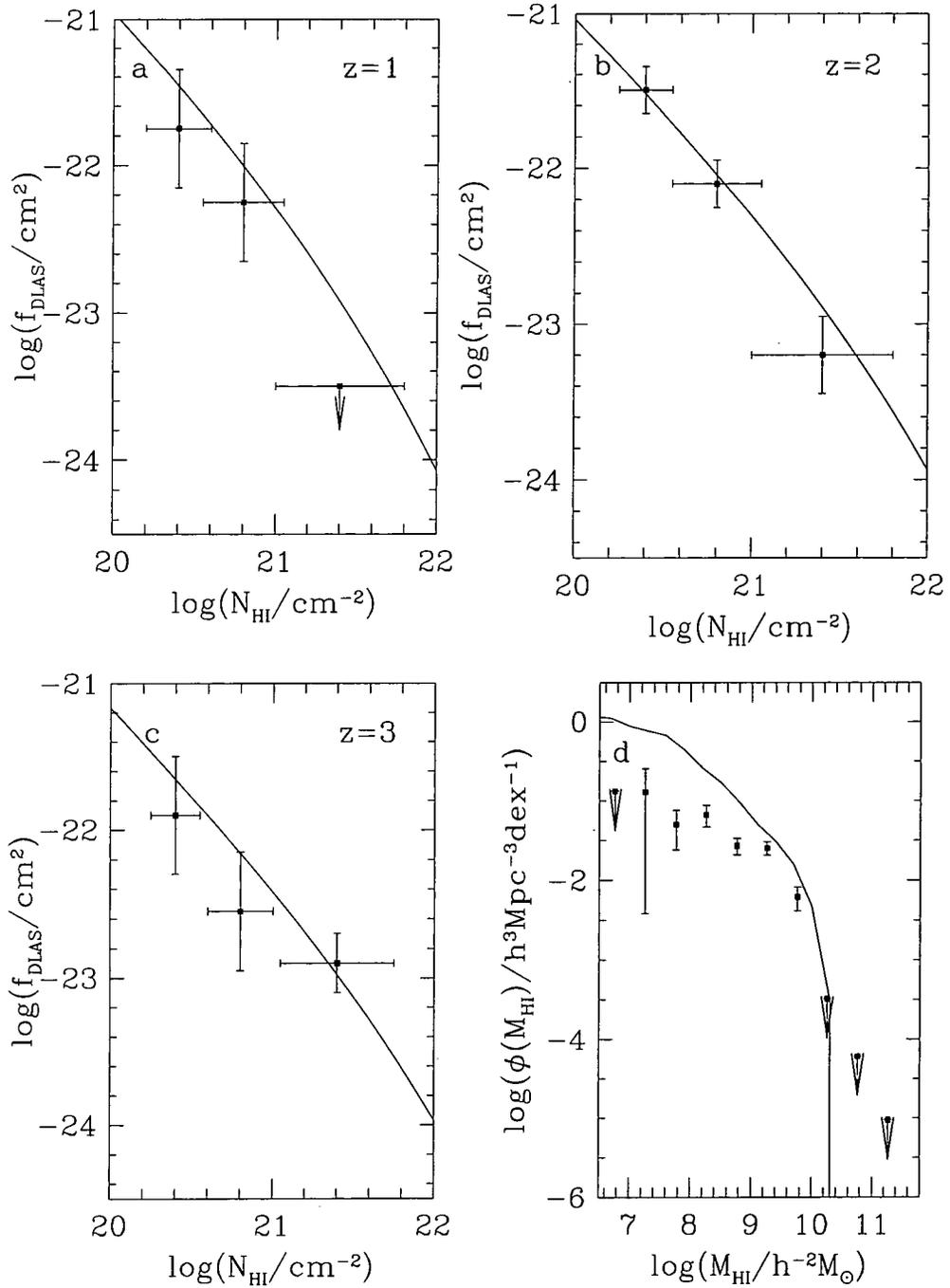


Figure 5.3: Panels (a), (b) and (c) show the distribution of DLAS at $z = 1$, 2 and 3 respectively as a function of their HI column density. Solid lines indicate the distribution determined from our model. Points with error bars are from Lanzetta, Wolfe & Tinschke (1996). Panel (d) shows the HI mass function of galaxies at $z = 0.0$. The solid line is the mass function determined from our model. Points with errorbars are from Zwaan et al. (1997).

The requirement that $5 \times 10^6 h^{-1} M_\odot$ halos be resolved sets an upper limit on the mass of halo that we can simulate due to computer memory limits, since the lower the mass of halo that is resolved, the more progenitors a halo of given mass will have. At $z = 0$, the most massive halos that we are able to simulate make a significant contribution to the total filling factor and ionizing flux. However, for $z \approx 2$ the most massive halos simulated contribute only 1% of the total number of escaping ionizing photons, and this fraction drops extremely rapidly as we look to even higher redshifts. Therefore, at the high redshifts ($z \gtrsim 3$) we will be interested in, ignoring higher mass halos makes no significant difference to our results.

We note that once any halo has begun to ionize the surrounding IGM, it could potentially influence the process of galaxy formation in nearby halos (Dekel & Rees 1987). Ionizing photons from the first halo will act to heat the gas in nearby halos, thereby reducing the effective cooling rate (Efstathiou, Bond & White 1992, Thoul & Weinberg 1996). Since prior to full reionization each halo will see only the flux of ionizing photons from nearby sources, a detailed accounting of this radiative feedback requires a treatment of the radiative transfer of the ionizing radiation through the IGM. This is beyond the scope of the present work. Such radiative feedback is expected to be very efficient at dissociating molecular hydrogen, with Ciardi et al. (2000) finding that H_2 is completely dissociated by $z \approx 25$. Radiative feedback will also inhibit galaxy formation both by reducing the amount of gas that accretes into low mass halos (Gnedin 2000) and by reducing the cooling rate of gas within halos. Thoul & Weinberg (1996) show that radiative feedback may be effective at inhibiting galaxy formation in halos with circular velocities of 50 km/s or less. In our model, the ionizing luminosity becomes dominated by galaxies in halos with circular velocities greater than 50 km/s at redshifts below $z \approx 10$. At higher redshifts we may therefore be overestimating the total ionizing luminosity produced by galaxies, but this should not significantly affect the reionization redshift.

5.3 The Escape Fraction of Ionizing Photons

5.3.1 Global Constraints at Low Redshift

Gas and dust inside galaxies can readily absorb ionizing photons and re-emit the energy at longer wavelengths. Therefore the amount and distribution of these components are the main factors that determine f_{esc} . The model of galaxy formation explicitly provides the mass and metallicity of cold gas present in each galaxy disk and the half-mass radius of

that disk, all as functions of time. The mass of dust is assumed to be proportional to the mass of cold gas and to its metallicity (c.f. eqn. 2.57). We split the escape fraction into contributions from gas, $f_{\text{esc,gas}}$, and dust, $f_{\text{esc,dust}}$, such that the total escaping fraction is given by $f_{\text{esc}} = f_{\text{esc,gas}}f_{\text{esc,dust}}$.

Chapter 2 describes in detail how the effects of dust are included in our model of galaxy formation. This modelling, which uses the calculations of Ferrara et al. (1999), is much more realistic than has been previously included in semi-analytic galaxy formation models, as it includes a fully 3D (though axisymmetric) dust distribution, and the dust optical depths are calculated for each galaxy individually. As noted in Chapter 2, stars are assumed to be distributed in a bulge and in an exponential disk with a vertical scale height equal to 0.0875 times the radial scale length (this ratio was adopted by Ferrara et al. to match the observed values for the old disk population of galaxies like the Milky Way). The dust is assumed to be distributed in the same way as the disk stars. The models give the attenuation of the ionizing radiation as a function of the inclination angle at which a galaxy is viewed, and we average this over angle to find the mean dust extinction for each galaxy. The dust attenuations do not include the effects of clumping of the stars or dust, and also assume that the ionizing stars have the same vertical distribution as the dust. With these two caveats in mind, the dust extinctions we apply should only be considered as approximate.

Some of the emitted Ly α photons are absorbed by neutral hydrogen close to the emitting star, thereby causing H α line emission from the galaxy. Therefore, the H α luminosity function is sensitive to the fraction, $f_{\text{esc,gas}}$, of the ionizing photons which manage to escape through the gas. We will require our models to reproduce the observed H α luminosity function and luminosity density. In Figs. 5.4 and 5.5 we compare the H α emission line properties of galaxies in our model with observational data at low redshift. The observed values are already corrected for dust extinction, so we compare them with the theoretical values before dust attenuation. In order to calculate these properties accurately, we simulate halos of mass up to and including $10^{15}h^{-1}M_{\odot}$. In calculating the H α line luminosity of each galaxy, we assume that a fraction $1 - f_{\text{esc,gas}}$ of the Ly α photons are absorbed by hydrogen atoms, producing H α photons according to case B recombination. The remaining Ly α photons escape, after being further attenuated by dust. The figures show results for $f_{\text{esc,gas}} = 0, 0.05$ and 0.2 , which roughly brackets the likely range of values for typical disk galaxies at the present day, as we discuss below. Both the predicted H α luminosity function and luminosity density are in reasonable agreement with

the observations, demonstrating that our models produce galaxies with realistic total ionizing luminosities (before attenuation by gas and dust). In principle, these observational comparisons provide a constraint on the value of $f_{\text{esc,gas}}$, if the other parameters in the semi-analytical model are assumed to be known. However, in practice it is not possible to reliably distinguish between $f_{\text{esc,gas}} = 0.2$ and $f_{\text{esc,gas}} = 0$ or 0.05 , given the uncertainties in the observational data. The observational results depend on the dust correction factors applied, and there is also some uncertainty in the ionizing luminosities predicted by stellar population synthesis models for a given IMF. With these caveats in mind it would seem that mean escape fractions anywhere between zero and 20-30% are acceptable. Observations of starburst galaxies in the nearby Universe suggest that the escape fraction is actually less than 3% for such galaxies (Leitherer et al. 1995), but starbursts are known to have very high column densities of gas and dust, and so the escape fraction in normal galaxies can probably be significantly higher (e.g. Kennicutt 1998a).

5.3.2 The Dependence of $f_{\text{esc,gas}}$ on Redshift and on Halo Mass

So far, we have assumed that $f_{\text{esc,gas}}$ is a global constant, varying neither with galaxy properties nor redshift. The details of the physical processes which determine $f_{\text{esc,gas}}$ are uncertain, but a constant $f_{\text{esc,gas}}$ seems unrealistic, as the properties of the emitting galaxies depend strongly upon both redshift and the mass of the halos in which they live. Given the complexity of this problem, here we merely aim at establishing the general trend of how $f_{\text{esc,gas}}$ may vary with halo mass and redshift.

We will consider three models for $f_{\text{esc,gas}}$. In the first model, $f_{\text{esc,gas}}$ is assumed to be a universal constant (this will be referred to as the “fixed model”). In the second and third models $f_{\text{esc,gas}}$ is evaluated for each galaxy, based on its physical properties. These two models are described next.

Our first physical $f_{\text{esc,gas}}$ model is based on the approach of Dove & Shull (1994, hereafter DS94) who derived an analytic expression for $f_{\text{esc,gas}}$. In their model, Ly α photons are emitted by OB associations in a galactic disk and escape by ionizing “HII chimneys” in the H I layer. The fraction of photons escaping a disk of given size and gas content can then be calculated. Whilst the original DS94 model assumes that OB associations all lie in the mid-plane of the galaxy disk, we have also considered the case where OB associations are distributed vertically like the gas in the disk.

The HII chimney model of DS94 does not include the effects of finite lifetimes of the OB associations, or of dynamical evolution of the gas distribution around an OB associ-

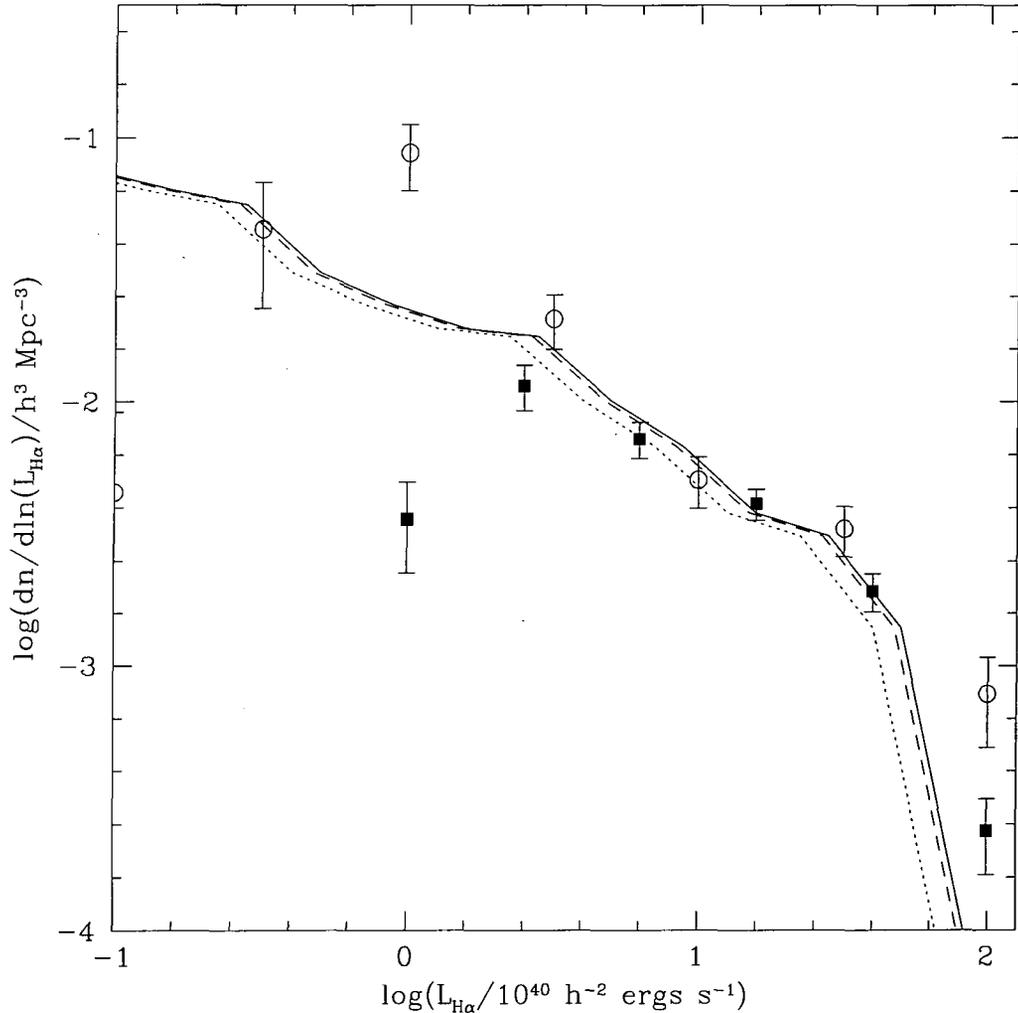


Figure 5.4: The $H\alpha$ luminosity function at $z = 0$. Points with error bars are observational data from Gallego et al. (1995a) (filled squares) and Sullivan et al. (2000) (open circles; note that the median redshift of this survey is $\langle z \rangle \approx 0.15$). The observed $H\alpha$ luminosities are corrected for dust extinction. The solid line is the luminosity function from our model assuming that no ionizing photons escape ($f_{\text{esc,gas}} = 0$). The dashed line is the same function for $f_{\text{esc,gas}} = 0.05$ and the dotted line for $f_{\text{esc,gas}} = 0.2$. The $H\alpha$ luminosities from the model are the values unattenuated by dust.

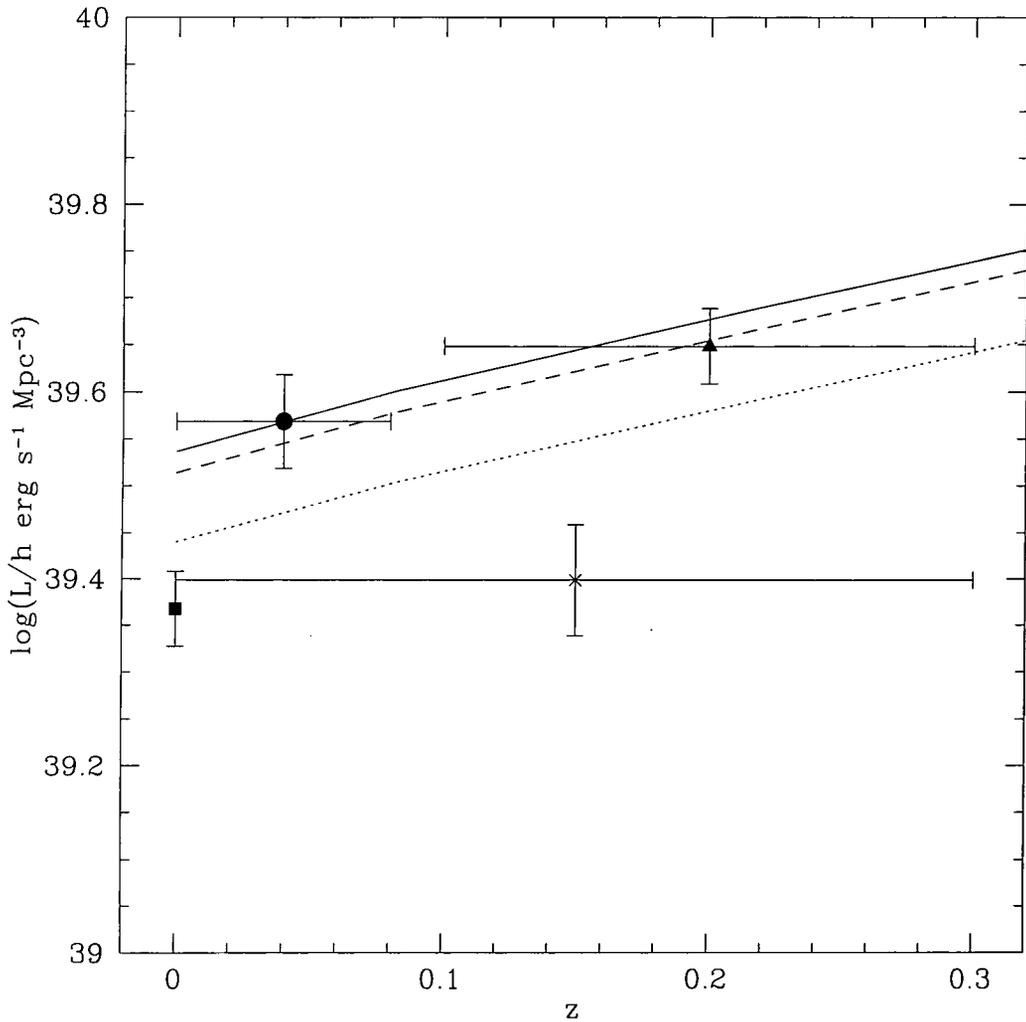


Figure 5.5: The $H\alpha$ luminosity density of the Universe as a function of redshift. Points with errorbars are observational estimates, including corrections for dust extinction: Gallego et al. (1995a) (square); Tresse & Maddox (1998) (triangle); Gronwall (1998) (circle) and Sullivan et al. (2000) (cross). We have converted the data for the effects of differing luminosity distances and volume elements to correspond to the $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$ cosmology assumed in our model. The solid line is the equivalent luminosity density measured in our model assuming that no photons escape ($f_{\text{esc,gas}} = 0$). The dashed line is the same function for $f_{\text{esc,gas}} = 0.05$ and the dotted line for $f_{\text{esc,gas}} = 0.2$.

ation due to energy input by stellar winds and supernova from the OB association itself. Dove, Shull & Ferrara (2000) have calculated the escape of ionizing photons through a dynamically evolving superbubble, which is driven by an OB association at its centre. They find that the resulting escape fractions are slightly lower than those obtained from the DS94 model (since the superbubble shell is able to trap radiation effectively). Numerical solutions of the radiative transfer equations in disk galaxies give results in excellent agreement with the Strömgren sphere approach of DS94 for OB associations at the bright end of the luminosity function, but give somewhat lower escape fractions for the faintest OB associations, the two approaches differing by around 25% for a single OB star (Wood & Loeb 1999).

Our second physical model for $f_{\text{esc,gas}}$ is based on the work of Devriendt et al. (1998, hereafter DSGN98). In this case, the ionizing stars are assumed to be uniformly mixed with the gas in the galaxy, and the gas is assumed to remain neutral. DSGN98 give an approximate analytic expression for the escape fraction in this case, but we have instead calculated the escape fraction exactly by numerical integration, for a specific choice for the gas density profile.

We give details of the calculation of $f_{\text{esc,gas}}$ in the DS94 and DSGN98 models in Appendix B.1. Both models contain one free parameter, h_z/r_{disk} , the ratio of disk scale height to radial scale length. We will consider the effects of varying this parameter in §5.5. For starbursts, we calculate the escape fraction based on a simple spherical geometry, as is also described in Appendix B.1. The contribution to the total ionizing luminosity from bursts of star formation is small ($< 8\%$) at all redshifts.

To summarize, we will show results from three models for $f_{\text{esc,gas}}$ as standard. These are: a model in which $f_{\text{esc,gas}}$ is held constant at 0.1; the DS94 model with OB associations in the disk midplane; and the DSGN98 model using our exact calculation of the escaping fraction. We consider the DS94 model to be the most realistic of our three models for $f_{\text{esc,gas}}$, but also present results from the other models for comparison.

In Fig. 5.6 we show the variation of f_{esc} with dark halo mass at $z = 0$ for the three models. The thin and thick lines show the escape fraction respectively with and without attenuation by dust. When a halo contains more than one galaxy, we plot the mean f_{esc} weighted by ionizing luminosity. At a given halo mass, halos with the lowest f_{esc} tend to have the highest ionizing luminosities, as both the star formation rate and attenuation of photons are increased in galaxies with large gas contents. The three models all show a trend for decreasing escape fraction with increasing halo mass up to $M_{\text{halo}} \sim 10^{12} h^{-1} M_{\odot}$.

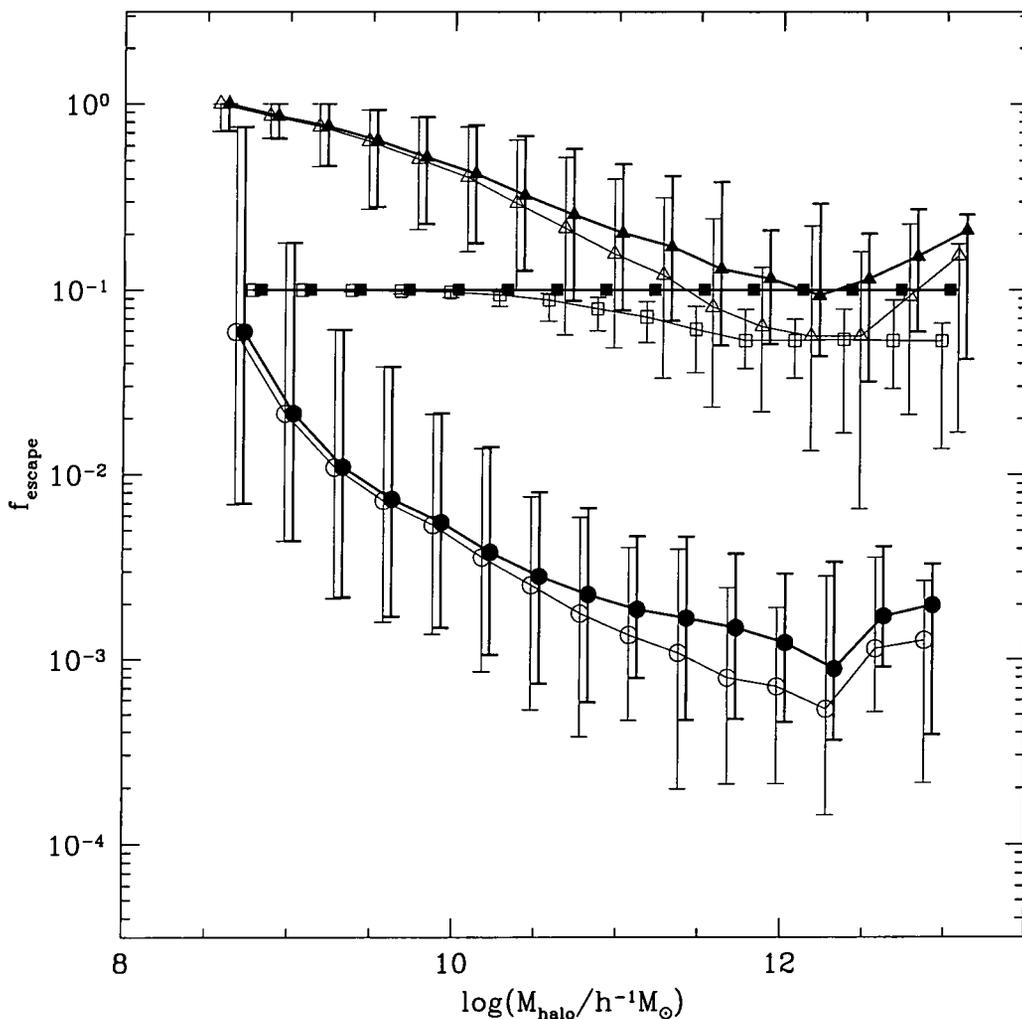


Figure 5.6: The escape fraction, f_{esc} , at $z = 0$, as a function of halo mass. Thick lines with solid symbols show the escape fraction ignoring the effects of dust, whilst thin lines with open symbols include absorption by dust. Three models for the gas escape fraction are plotted: fixed escape fraction of 10% (squares), DS94 (triangles), and DSGN98 (circles). In each case, the symbols indicate the median of the distribution of escape fractions, whilst the errorbars indicate the 10% and 90% intervals of the distribution.

For the fixed gas escape fraction model, the variation in f_{esc} is due entirely to the effects of dust, which can therefore be seen to be negligible in halos less massive than $\sim 10^{10}h^{-1}M_{\odot}$. This decrease in f_{esc} due to dust is enhanced in the other two models by the variation in $f_{\text{esc,gas}}$, which also declines with increasing halo mass. For halos more massive than $10^{12}h^{-1}M_{\odot}$, the escape fractions rise somewhat for the variable $f_{\text{esc,gas}}$ models. Note that the DSGN98 model predicts a much smaller escape fraction than the DS94 model at all masses as in this model only the small fraction of photons emitted in the outer regions of the disk (where the optical depth $\tau \lesssim 1$) typically manage to escape.

In Fig. 5.7 we plot the variation in the (ionizing luminosity-weighted) mean disk scale length and cold gas mass for galaxies in our model as a function of halo mass. Evidently, the decline in f_{esc} with increasing halo mass below $10^{12}h^{-1}M_{\odot}$ seen in Fig. 5.6 is due mainly to the greater masses of gas found in galaxies in these halos. This rapid change in the mass of gas present is due to the effects of feedback, which efficiently ejects gas from galaxies in low mass halos. Above $10^{12}h^{-1}M_{\odot}$, the mass of cold gas in galaxies levels off and then begins to decline as cooling becomes inefficient in more massive halos. This results in an escape fraction increasing with halo mass for the most massive halos simulated. Although galaxy sizes increase with increasing halo mass, thereby reducing gas densities somewhat, this effect is not strong enough to offset the increased cold gas mass in these galaxies.

We find that the DS94 and DSGN98 models applied to our galaxies predict escape fractions (including the effects of dust) for halos of mass $\sim 10^{11}h^{-1}M_{\odot}$ at $z = 0$ of $\approx 20\%$ and $\approx 0.2\%$ respectively. The mean DS94 and DSGN98 luminosity-weighted escape fractions for galaxies at $z = 0$ are lower, being $\approx 6\%$ and $\approx 0.1\%$ respectively. However, we expect some variation in these values with redshift due to the evolution of the galaxy population. In fact, we find a rapid decline in both cold gas content and galaxy disk size with increasing redshift. In Fig. 5.8 we show the evolution of the (ionizing luminosity-weighted) mean f_{esc} between redshifts 0 and 45. All models show an initial rapid decline in f_{esc} with increasing z . After this, in the constant $f_{\text{esc,gas}}$ model, the mean escape fraction increases with redshift since the dust content of galaxies was lower in the past. The DS94 model shows a very gradually rising escape fraction, whilst the DSGN98 model has a more rapid decline.

In our model, the contribution of stellar sources to the UV background is dominated by galaxies at low redshifts ($z \lesssim 1$). We find that immediately shortwards of 912\AA our DS94 model predicts a background due to stellar sources which is very close to that expected

from QSOs (Haardt & Madau 1996), after including the effects of attenuation by the intervening IGM (Madau 1995). At shorter wavelengths the QSO contribution soon becomes dominant. Thus, at $z = 0$ the combined background due to stars (from our DS94 model) and QSOs (from Haardt & Madau 1996) is $J_{912\text{\AA}} \approx 4 \times 10^{-23}$ ergs/s/cm²/Hz/ster. This is consistent with the upper limit of $J_{912\text{\AA}} = 8 \times 10^{-23}$ ergs/s/cm²/Hz/ster found by Vogel et al. (1995), who searched for H α emission from intergalactic HI clouds. The contribution of galaxies to the local ionizing background has also been estimated by Giallongo, Fontana & Madau (1997), based on the luminosity function of galaxies observed in the Canada-France Redshift Survey. They estimated the galactic contribution as $J_{912\text{\AA}} \approx 5 \times 10^{-23}$ ergs/s/cm²/Hz/ster at $z = 0$, assuming an escape fraction of $f_{\text{esc}} = 0.15$. If we assume the same f_{esc} in our model, we obtain $J_{912\text{\AA}} = 5.2 \times 10^{-23}$ ergs/s/cm²/Hz/ster, in excellent agreement with their result.

5.4 The Filling Factor and the Evolution of the Ionization Fronts

We define the filling factor, F_{fill} , as the fraction of hydrogen in the IGM (as defined in §5.2.2) which has been ionized. This is the natural quantity which serves as an indication of the amount of reionization in the IGM. We calculate the growth of the ionized region around each halo, using the ionizing luminosities predicted by the semi-analytic model, and then sum over all halos to find F_{fill} . We make two simplifying assumptions: (1) the radiation from each halo is emitted isotropically, and (2) the distribution of hydrogen is uniform on the scale of the ionization front and larger (but with small-scale clumping). It follows that each halo by itself would produce a spherical ionization front.

The mass of hydrogen ionized within the ionization front, M , in spherical symmetry is given by (Shapiro & Giroux 1987, Haiman & Loeb 1996)

$$\frac{1}{m_{\text{H}}} \frac{dM}{dt} = S(t) - \alpha_{\text{H}}^{(2)} a^{-3} f_{\text{clump}} n_{\text{H}} \frac{M}{m_{\text{H}}}, \quad (5.4)$$

where $\alpha_{\text{H}}^{(2)}$ is the net HII recombination coefficient to all levels but the first, n_{H} is the comoving mean number density of hydrogen atoms (total, HI and HII) in the IGM, $a(t)$ is the scale factor of the universe normalized to unity at $z = 0$, t is time and $S(t)$ is the rate at which ionizing photons are being emitted. The factor $f_{\text{clump}}(t)$, defined by

$$f_{\text{clump}} = \langle \rho_{\text{IGM}}^2 \rangle / \bar{\rho}_{\text{IGM}}^2, \quad (5.5)$$

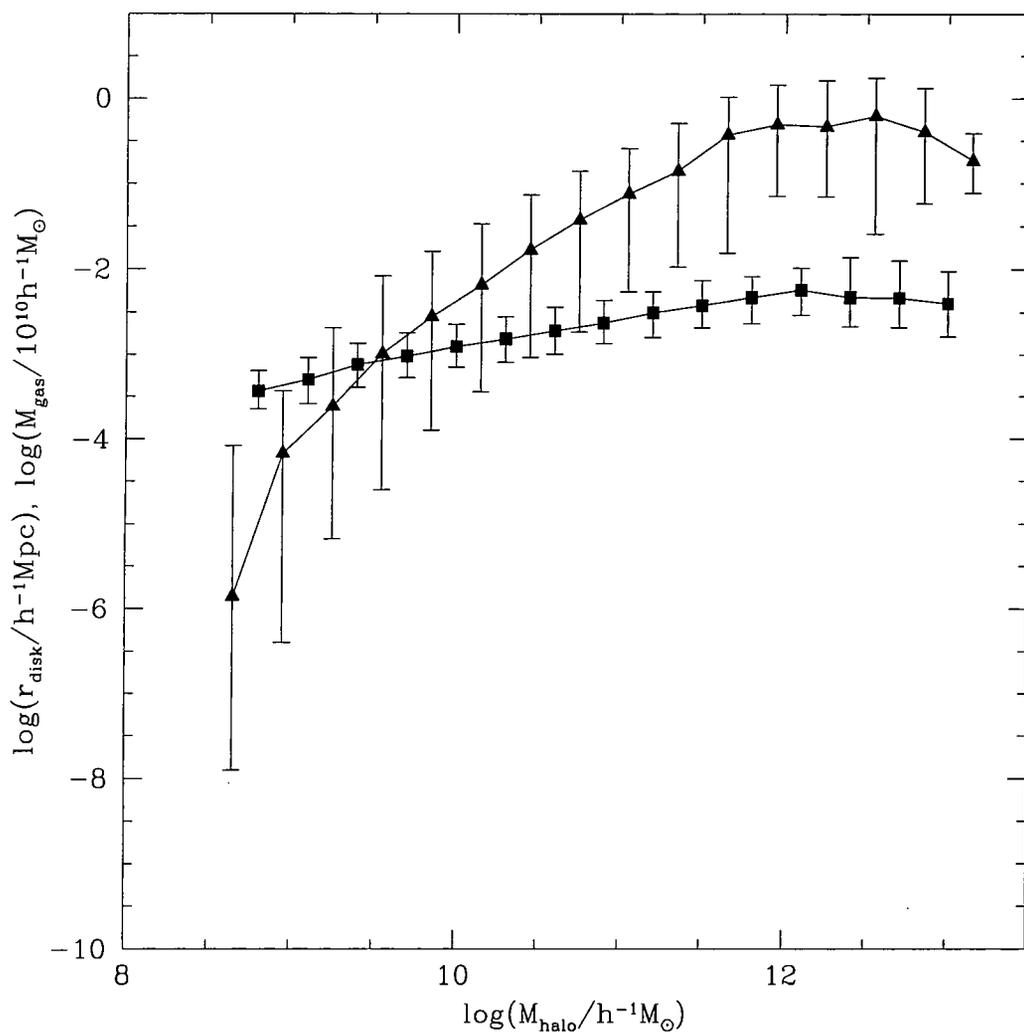


Figure 5.7: The variation in (ionizing luminosity-weighted) mean disk scale length (squares) and gas mass (triangles) for galaxies in our model as a function of halo mass. The symbols show the medians of the distributions, while the errorbars indicate the 10% and 90% intervals.

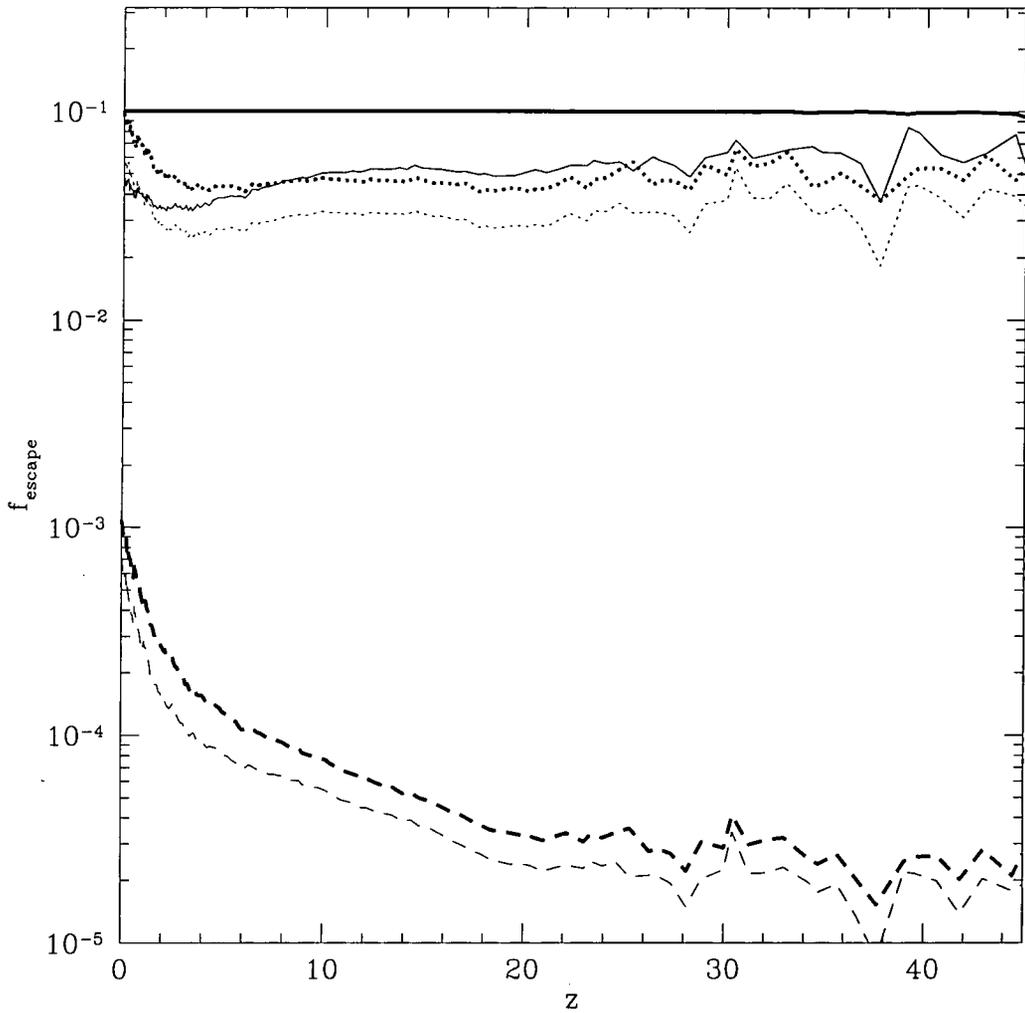


Figure 5.8: The (ionizing luminosity-weighted) mean escape fraction for all galaxies as a function of redshift. Thick lines show the escape fraction ignoring the effects of dust, whilst thin lines include dust. Three models for the gas escape fraction are plotted: constant gas escape fraction of 10% (solid line), DS94 (dotted line) and DSGN98 (dashed line).

is the clumping factor for the ionized gas in the IGM (here ρ_{IGM} is the density of IGM gas at any point and $\bar{\rho}_{\text{IGM}}$ is the mean density of the IGM). Small-scale clumpiness causes the total recombination rate to be larger than for a uniform medium of the same mean density.

The value of $f_{\text{clump}}(t)$ for the ionized gas in the IGM is complicated to calculate analytically. We remind the reader that in our picture, the IGM consists of all gas which has *not* been collisionally ionized in halos nor become part of a galaxy. For a uniform IGM $f_{\text{clump}} = 1$ by definition. If low density regions of the IGM are ionized before high density regions, as suggested by Miralda-Escudé, Haehnelt & Rees (2000), then this would be similar to having $f_{\text{clump}} < 1$ in eqn. (5.4), but for most purposes, $f_{\text{clump}} = 1$ can be considered as an approximate lower bound. We make two different estimates of the possible effects of clumping.

For our first estimate, which we call $f_{\text{clump}}^{(\text{variance})}$, we assume that the photoionized gas basically traces the dark matter, except that gas pressure prevents it from falling into dark matter halos with virial temperatures smaller than 10^4K (the approximate temperature of the photo-ionized gas). Thus, we calculate the clumping factor as $f_{\text{clump}}^{(\text{variance})} = (1 + \sigma^2)$, where σ^2 is the variance of the dark matter density field in spheres of radius equal to the virial radius of a 10^4K halo. σ^2 is calculated from the non-linear dark matter power spectrum, estimated using the procedure of Peacock & Dodds (1996), and smoothed using a top-hat filter in real space.

For our second estimate, which we call $f_{\text{clump}}^{(\text{halos})}$, we include the effects of collisional ionization in halos and of removal of gas by cooling into galaxies in a way consistent with our definition of f_{IGM} given in eqn. (5.3). The diffuse gas in halos with virial temperatures above $\approx 10^4\text{K}$ is assumed to have the density profile of an isothermal sphere with a constant density core. The gas originally associated with smaller halos is assumed to be pushed out of these halos by gas pressure following photoionization, and to be in a uniform density component occupying the remaining volume. As shown in Appendix B.2, the clumping factor is then

$$f_{\text{clump}}^{(\text{halos})} = \frac{f_{\text{m,smooth}}^2}{f_{\text{v,smooth}} f_{\text{IGM}}^2} + \frac{f_{\text{int}} \Delta_{\text{vir}}}{f_{\text{IGM}}^2} \int_{M_{\text{J}}}^{\infty} \langle (1 - f_{\text{gal}})^2 \rangle (1 - x_{\text{H}})^2 \frac{M_{\text{halo}}}{\rho_{\text{c}} \Omega_0} \frac{dn}{dM_{\text{halo}}} dM_{\text{halo}}, \quad (5.6)$$

where M_{J} is the mass of a halo which just retains reionized gas, $f_{\text{m,smooth}}$ is the fraction of the total baryonic mass in the uniform component, and $f_{\text{v,smooth}}$ is the fraction of the volume of the universe occupied by this gas. Here, f_{gal} is the fraction of the baryonic mass in a halo in the form of galaxies, x_{H} is the fraction of hydrogen in the diffuse halo

gas which is collisionally ionized (as in eqn. 5.3), and $\langle \rangle$ indicates an average over all halos of mass M_{halo} . The factor f_{int} is a parameter depending only on the ratio of the size of the core in the gas density profile to the halo virial radius. For a core radius equal to one-tenth of the virial radius $f_{\text{int}} = 3.14$ (see §B.2). This estimate of the clumping factor ignores the possibility of gas in the centres of halos (but not part of a galaxy) becoming self-shielded from the ionizing radiation. Such gas would not become photoionized, and so would not contribute to the recombination rate, resulting in f_{clump} being lower than estimated here. A detailed treatment of the ionization and temperature structure of gas inside halos is beyond the scope of this work.

Of course, before reionization, when the gas is typically much cooler than 10^4K , gas will fall into dark matter halos with virial temperatures below 10^4K . If all gas were in halos, then we would find

$$f_{\text{clump}} \approx \frac{f_{\text{int}} \Delta_{\text{vir}}}{f_{\text{IGM}}^2} \int_0^\infty \langle (1 - f_{\text{gal}})^2 \rangle (1 - x_{\text{H}})^2 \frac{M_{\text{halo}}}{\rho_{\text{c}} \Omega_0} \frac{dn}{dM_{\text{halo}}} dM_{\text{halo}} \quad (5.7)$$

Once reionized, some of the gas in these small halos will flow back out of the halo as the gravitational potential is no longer deep enough to confine the gas.

We consider $f_{\text{clump}}^{(\text{halos})}$ as our best estimate of the IGM clumping factor, at least for lower redshifts, when a significant fraction of the gas is in halos with $M > M_J$, while $f_{\text{clump}}^{(\text{variance})}$ is more in the nature of an upper limit.

The clumping factors $f_{\text{clump}}^{(\text{halos})}$ and $f_{\text{clump}}^{(\text{variance})}$ are plotted as functions of redshift in Fig. 5.9. They show fairly similar behaviour above $z \approx 10$. Below this redshift, $f_{\text{clump}}^{(\text{variance})}$ greatly exceeds $f_{\text{clump}}^{(\text{halos})}$, because it becomes dominated by gas in massive dark matter halos, which, on the other hand, contributes negligibly to $f_{\text{clump}}^{(\text{halos})}$ as it is collisionally ionized. We also plot estimates of the clumping factor from two other works: Valageas & Silk (1999) calculated the clumping factor of the baryons which have been unable to cool (the quantity they call C_n) using their own analytical model. They obtain values of f_{clump} which are comparable to $f_{\text{clump}}^{(\text{halos})}$ at $z \lesssim 10$, but are substantially larger at higher redshifts. Gnedin & Ostriker (1997) performed hydrodynamical simulations in a cosmology similar to that which we consider, from which they measured f_{clump} directly. They calculated two clumping factors: one for all baryons in their simulation, $f_{\text{clump}} = f_{\text{clump}}^{(\text{GO:bb})}$, and the other for baryons in ionized regions only, $f_{\text{clump}} = f_{\text{clump}}^{(\text{GO:HII})}$, which is smaller. $f_{\text{clump}}^{(\text{GO:HII})}$ is more relevant for our purposes, but may still overestimate the clumping of photoionized gas in the IGM, since it includes collisionally ionized gas in galaxy halos. These clumping factors are everywhere lower than $f_{\text{clump}}^{(\text{variance})}$. The quantity $f_{\text{clump}}^{(\text{GO:HII})}$ is close to

our estimate $f_{\text{clump}}^{(\text{halos})}$ at the highest and lowest redshifts, but smaller in the intermediate range.

5.4.1 Model Results

In Fig. 5.10 we show the ionized filling factor of the IGM, F_{fill} , as a function of redshift. Here we compute F_{fill} by summing the volumes of the HII regions formed around each halo, weighted by the number of such halos per unit volume as given by the Press-Schechter theory. (Later we will use the halo mass function measured directly from an N-body simulation to calculate F_{fill} — see Fig. 5.14). Note that since we calculate F_{fill} by simply summing the volumes of ionization fronts around halos it is possible that F_{fill} will exceed 1, contrary to the definition given in §5.4. The case $F_{\text{fill}} > 1$ simply implies that more ionizing photons have been produced than are needed to completely reionize the universe. We show results for our three models for $f_{\text{esc,gas}}$, and for three different assumptions about f_{clump} .

If we ignore the effects of dust, we find that the model with a constant escape fraction of 10% reionizes the Universe by $z = 7.9$ if $f_{\text{clump}} = 1$ but only by $z = 6.6$ if $f_{\text{clump}} = f_{\text{clump}}^{(\text{halos})}$. In order to reionize the Universe by $z = 5$, escape fractions of 1.4% and 3.7% are needed for $f_{\text{clump}} = 1$ and $f_{\text{clump}} = f_{\text{clump}}^{(\text{halos})}$ respectively. When we include the effects of dust, we find that gas escape fractions of 3.3% and 9.3% are needed to reionize by $z = 5$ for these two cases.

If, instead of assuming a constant gas escape fraction, we use the more physically motivated DS94 model, we find reionization occurs at $z = 6.1$ if $f_{\text{clump}} = 1$, but only at $z = 4.5$ if $f_{\text{clump}} = f_{\text{clump}}^{(\text{halos})}$ (both estimates including dust). In the latter case, the ionized filling factor at $z = 5$ is only 76%. If we assume OB associations are distributed as the gas in the DS94 model (as opposed to lying in the disk mid-plane as in our standard model) then a filling factor of 117% (i.e. full reionization) is achieved by $z = 5$. The DSGN98 model, which predicts much lower escape fractions than the DS94 model at all redshifts, is able to reionize only $\approx 2\%$ of the IGM by redshift 3, and full reionization never occurs even if $f_{\text{clump}} = 1$.

We note that in the DS94 model, approximately 90% of the photons required for ionization are produced at $z < 10$. Thus our neglect of radiative feedback effects (which may reduce the number of ionizing photons produced at higher redshifts) is unlikely to seriously effect our determination of the reionization epoch.

As both the DS94 and DSGN98 models predict quite low escape fractions, we have also

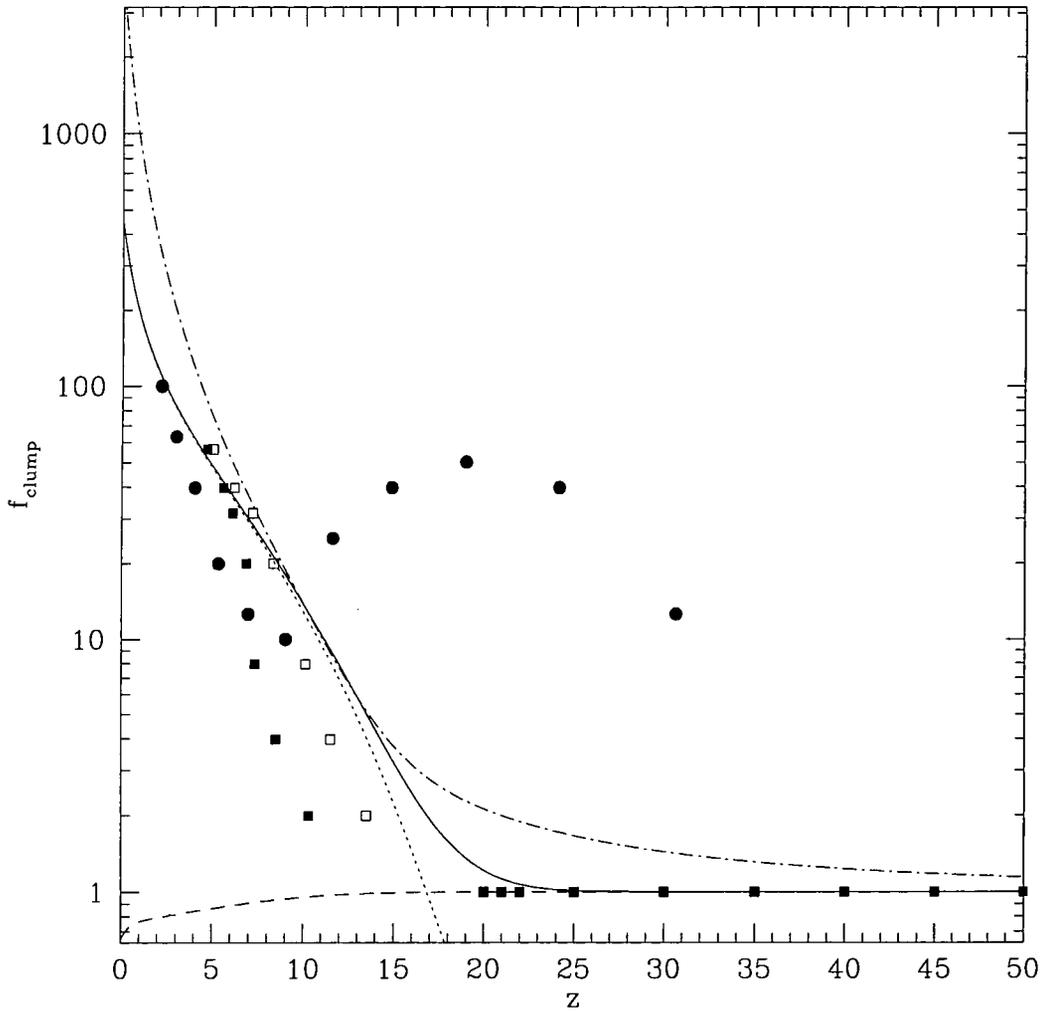


Figure 5.9: The gas clumping factor, f_{clump} , as a function of redshift. The solid line shows $f_{\text{clump}}^{(\text{halos})}$, whilst the dotted and dashed lines show the contributions to this quantity from gas inside and outside halos respectively. The dot-dashed line shows $f_{\text{clump}}^{(\text{variance})}$. Filled circles show the clumping factor calculated by Valageas & Silk (1999). Squares show the clumping factors determined from a simulation by Gnedin & Ostriker (1997) for all baryons (open squares) and baryons in ionized regions only (filled squares).

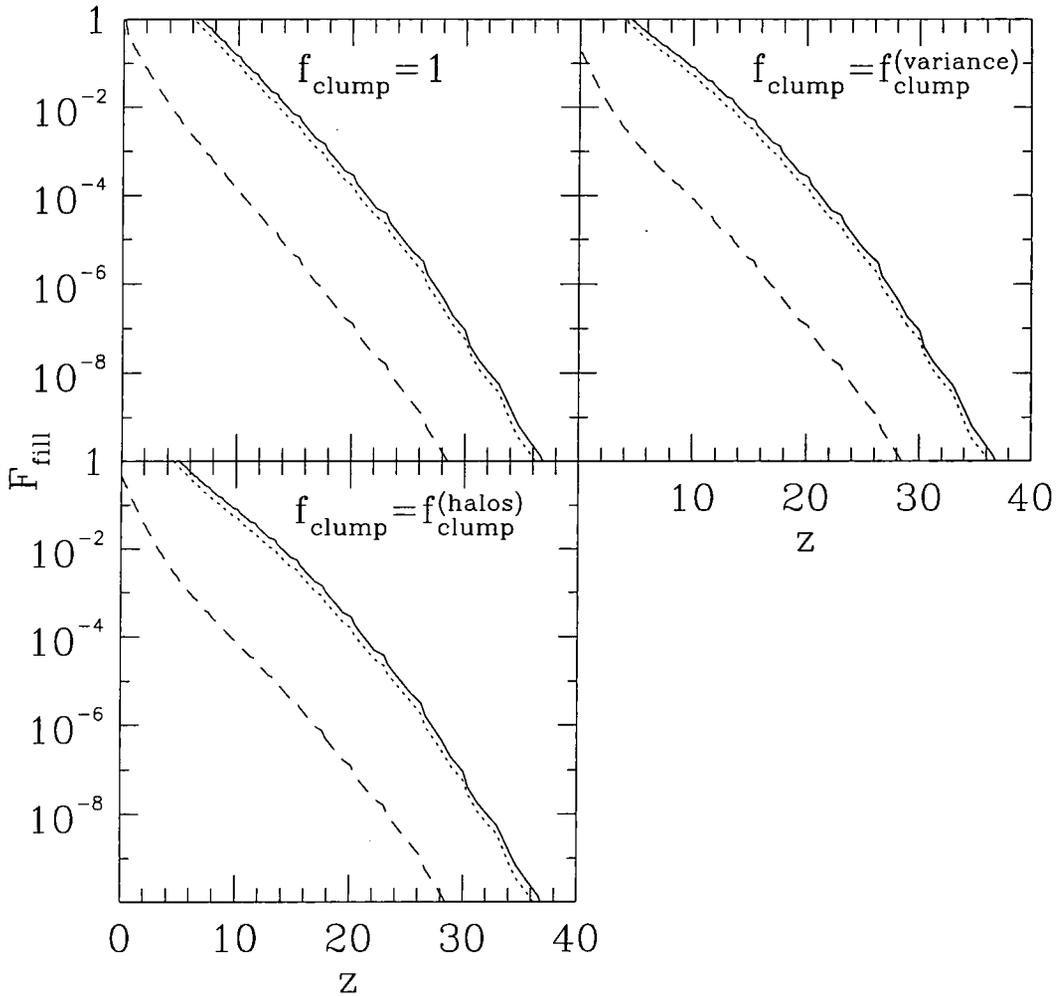


Figure 5.10: The filling factor, F_{fill} , as a function of redshift determined using three different models for the absorption of ionizing photons by gas inside galaxies: constant gas escape fraction of 10% (solid lines), DS94 (dotted lines) and DSGN98 (dashed lines). All models include the effects of dust on the escape fraction. The three panels show the filling factors for three different assumptions about the IGM clumping factor, f_{clump} : no clumping (top left panel), clumping due to virialised halos (bottom left panel) and clumping estimated from the Peacock & Dodds (1996) non-linear power spectrum (top right panel).

considered a much more extreme model which simply assumes that $f_{\text{esc}} = \beta/(1+\beta)$, where β is the feedback efficiency as defined by eqn. (5.1). This toy model, to which we will refer as the “holes scenario”, produces very high escaping fractions for galaxies with low circular speeds, and low escaping fractions for those with high circular speeds. A behaviour for f_{esc} of this general form might result if photons are able to escape through holes in the galaxy disk which have been created by supernovae. Since dust would also be expected to be swept out of these holes we do not include any dust absorption in this model. The holes scenario produces very different results compared to our two physical models for the escape fraction. In this model $f_{\text{esc}} \approx 1$ for $z > 10$, dropping to 45% by $z = 0$. Not surprisingly therefore, this model succeeds in satisfying the Gunn-Peterson constraint, reionizing the Universe by $z = 11.7$ if $f_{\text{clump}} = 1$ and by $z = 10.6$ if $f_{\text{clump}} = f_{\text{clump}}^{(\text{halos})}$. While this model is only a very crude attempt to consider a dynamically disturbed gas distribution in galaxy disks, it clearly demonstrates that such effects may be of great importance for studies of reionization.

We have also computed the filling factor in our model using the clumping factors calculated by Gnedin & Ostriker (1997) and Valageas & Silk (1999) (as given in Fig. 5.9). Of course, this is not strictly self-consistent, as their clumping factors are calculated from their own models for galaxy formation and reionization, which differ from ours. Using either of these with the DS94 model gives a reionization redshift comparable to that obtained using $f_{\text{clump}} = f_{\text{clump}}^{(\text{halos})}$: we find reionization at $z = 3.6$ using the Gnedin & Ostriker (1997) clumping factor, and $z = 4.9$ using that of Valageas & Silk (1999).

In Fig. 5.11 we show the total number of ionizing photons which have escaped into the IGM per unit comoving volume by redshift z , n_γ , divided by the total number of hydrogen nuclei in the IGM per unit comoving volume, n_{H} (which is f_{IGM} times the total number density of hydrogen nuclei). When this number reaches one, just enough photons have been emitted by galaxies to reionize the IGM completely if recombinations are unimportant. This criterion has been used previously to estimate when reionization may occur. Since our model includes the effects of recombinations in the IGM, we can judge how well this simpler criterion performs. If we ignore the effects of absorption by gas and dust on the number of ionizing photons escaping from galaxies, we find that, in this cosmology, our model achieves $n_\gamma/n_{\text{H}} = 1$ by $z \approx 12$. When we account for the effects of dust and gas in galaxies, we find that the redshift at which $n_\gamma/n_{\text{H}} = 1$ is significantly reduced, the exact value depending on the model for the escape fraction. With a fixed gas escape fraction of 10%, $n_\gamma/n_{\text{H}} = 1$ by $z \approx 7$, whilst for the DS94 model $n_\gamma/n_{\text{H}} = 1$ is

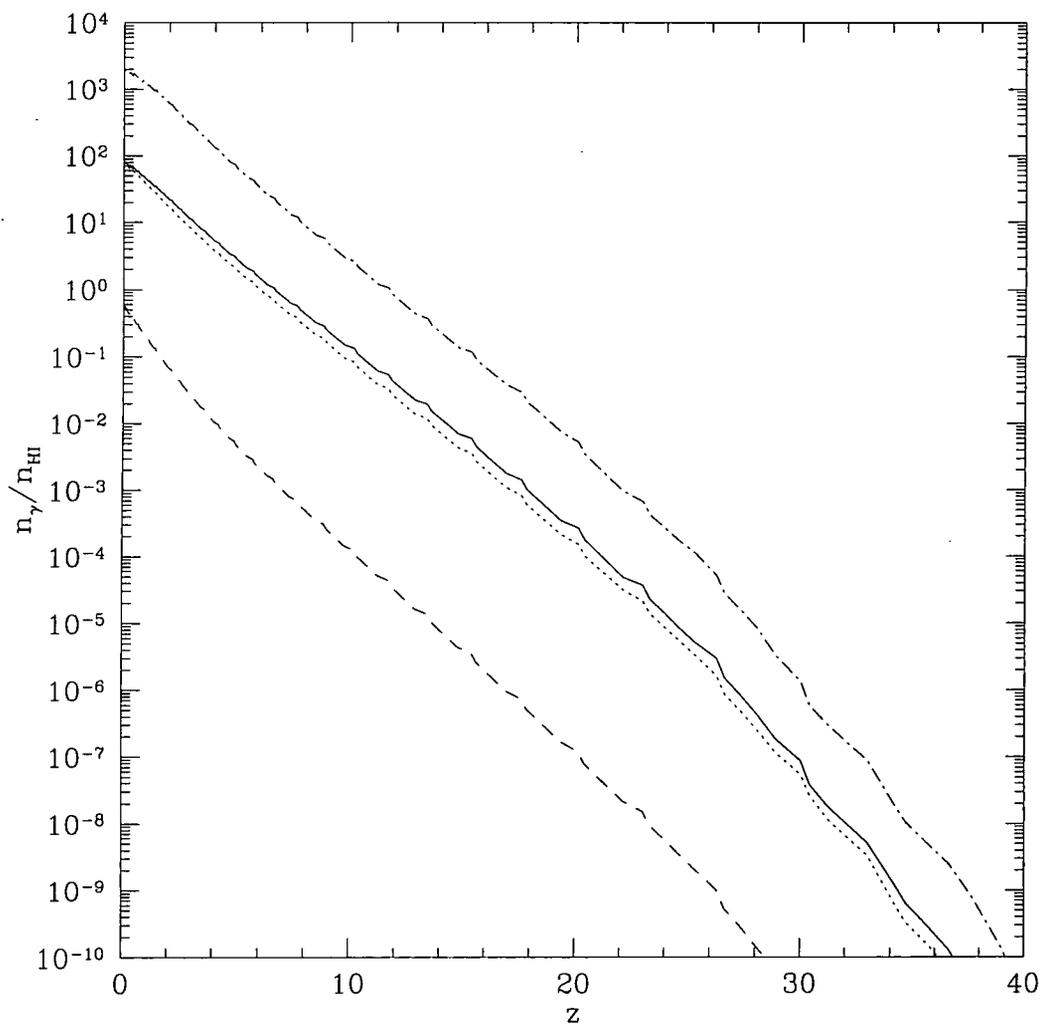


Figure 5.11: The ratio of the total number of ionizing photons which have escaped into the IGM per comoving volume by redshift z to the comoving number density of hydrogen nuclei in the IGM. The dot-dash line assumes that all ionizing photons can escape from their galaxies into the IGM. Three models for the gas escape fraction are also shown (each also includes the effects of dust): fixed escape fraction of 10% (solid line), DS94 (dotted line) and DSGN98 (dashed line).

achieved at $z \approx 6$. In the DSGN98 model $n_\gamma/n_H = 1$ has not been achieved even by $z = 0$. When we include recombinations in the IGM, the model with constant gas escape fraction reaches $F_{\text{fill}} = 1$ only by $z = 6.7$ for $f_{\text{clump}} = 1$, and by $z = 5.1$ for $f_{\text{clump}} = f_{\text{clump}}^{(\text{halos})}$, as illustrated in Fig. 5.10, showing how reionization is delayed.

Note that while the cosmology considered here is similar to Model G of Baugh et al. (1998b), the parameters of the semi-analytic model used are somewhat different. Specifically, Baugh et al. (1998b) used a model in which feedback was much more effective in low mass halos than in our model, since they required their models to produce a B-band luminosity function with a shallow faint end slope. As a result, the epoch at which $n_\gamma/n_H = 1$ was much later in Model G of Baugh et al. (1998b) than in our current model.

In summary, we see that, even for a specific model of galaxy formation, the predicted epoch of reionization is sensitive to the uncertain values of the escape fraction f_{esc} and the clumping factor f_{clump} . If the clumping factor is as large as $f_{\text{clump}}^{(\text{halos})}$, then in the case of a constant gas escape fraction $f_{\text{esc,gas}}$, we need $f_{\text{esc,gas}} \gtrsim 10\%$ in our model to ionize the IGM by $z = 5$, if absorption by dust is included, and $f_{\text{esc,gas}} \gtrsim 4\%$ if dust is ignored. With the more physically-motivated DS94 and DSGN98 models, and the same clumping factor, at most 76% of the IGM is reionized by $z = 5$, which would be inconsistent with observations of the Gunn-Peterson effect. For the extreme case of a uniform IGM, reionization occurs by $z = 6.1$ even with the DS94 model for f_{esc} . Our “best estimate” is based on combining the DS94 model with $f_{\text{clump}}^{(\text{halos})}$ for the IGM clumping factor. As already stated, this model narrowly fails to satisfy the Gunn-Peterson constraint at $z = 5$ (unless we assume that OB associations are distributed as the gas, rather than lying in the disk mid-plane), suggesting that additional sources of ionizing radiation are required at high redshift, either more stars than in our standard model, or non-stellar sources (e.g. quasars). However, given the theoretical uncertainties in f_{esc} and f_{clump} , we consider that this is not yet proven.

5.5 Sensitivity of Results to Model Parameters

We turn now to test the robustness of our results to variations in the parameters of our galaxy formation model. To do this, we have varied key parameters of the models and determined the ionized hydrogen filling factor in each case. We consider several different models. The variant models which we consider are listed in Table 5.2. In each case, we give the value of the parameter which is changed relative to the standard model given in

Table 5.1.

In Chapters 3 & 4 we have shown that normalising models to the $z = 0$ B-band luminosity function allows robust estimates of the $z = 0$ galaxy correlation function to be made. Here we choose a similar constraint, forcing all models to match the $z = 0$ H α luminosity function of Gallego et al. (1995a) at $L_{\text{H}\alpha} = 4 \times 10^{41} h^{-2}$ ergs/s (note that at these luminosities the Gallego et al. luminosity function agrees, within the errorbars, with that of Sullivan et al. 2000). This is achieved by adjusting the value of the parameter Υ (which determines the fraction of brown dwarfs formed in the model). The $z = 0$ H α luminosity functions for all models considered are shown in Fig. 5.12. Dotted lines show those models with H α luminosity functions that are significantly different from that of the standard model (at either the bright or faint ends).

In Table 5.2 we list escape fractions and filling factors in the variant models for the fixed and DS94 models for f_{esc} for the case $f_{\text{clump}} = f_{\text{clump}}^{(\text{halos})}$. Values of F_{fill} may exceed unity, as, in some models, by $z = 5$ more ionizing photons have escaped into the IGM than are required to reionize the universe.

The standard choice for the feedback efficiency, β , makes feedback highly efficient in galaxies with low circular velocities. In this model $\beta = f_{\text{V}}^{-\alpha_{\text{hot}}}$, where $f_{\text{V}} = V_{\text{disk}}/V_{\text{hot}}$. The fraction of cold gas which is reheated by supernovae after infinite time (a quantity with direct physical interpretation) is then

$$\frac{\beta}{1 - R + \beta} = \frac{f_{\text{V}}^{-\alpha_{\text{hot}}}}{1 - R + f_{\text{V}}^{-\alpha_{\text{hot}}}}. \quad (5.8)$$

Thus as $V_{\text{disk}} \rightarrow 0$ all gas is reheated and no stars are formed. For the modified feedback model, we adapt this form such that even in arbitrarily small potential wells not all the gas is reheated by supernovae. We choose

$$\frac{\beta}{1 - R + \beta} = \frac{a_{\text{r}} f_{\text{V}}^{-\alpha_{\text{hot}}}}{1 - R + f_{\text{V}}^{-\alpha_{\text{hot}}}}, \quad (5.9)$$

where a_{r} is an adjustable parameter. Now, as $V_{\text{disk}} \rightarrow 0$ a fraction a_{r} of gas is reheated, whilst a fraction $1 - a_{\text{r}}$ forms stars. The standard feedback model is recovered when $a_{\text{r}} = 1$. For the alternative feedback models considered here a value of $a_{\text{r}} = 0.75$ is used.

It should be noted that the variation having one of the greatest influences on the predicted filling factors is that of Model 20, where we use the alternative form for feedback given above. This model produces an H α luminosity function with a very steep faint end slope, since feedback never becomes highly efficient, even in extremely small dark matter halos. More ionizing photons are produced than in the standard model and higher filling factors are achieved.

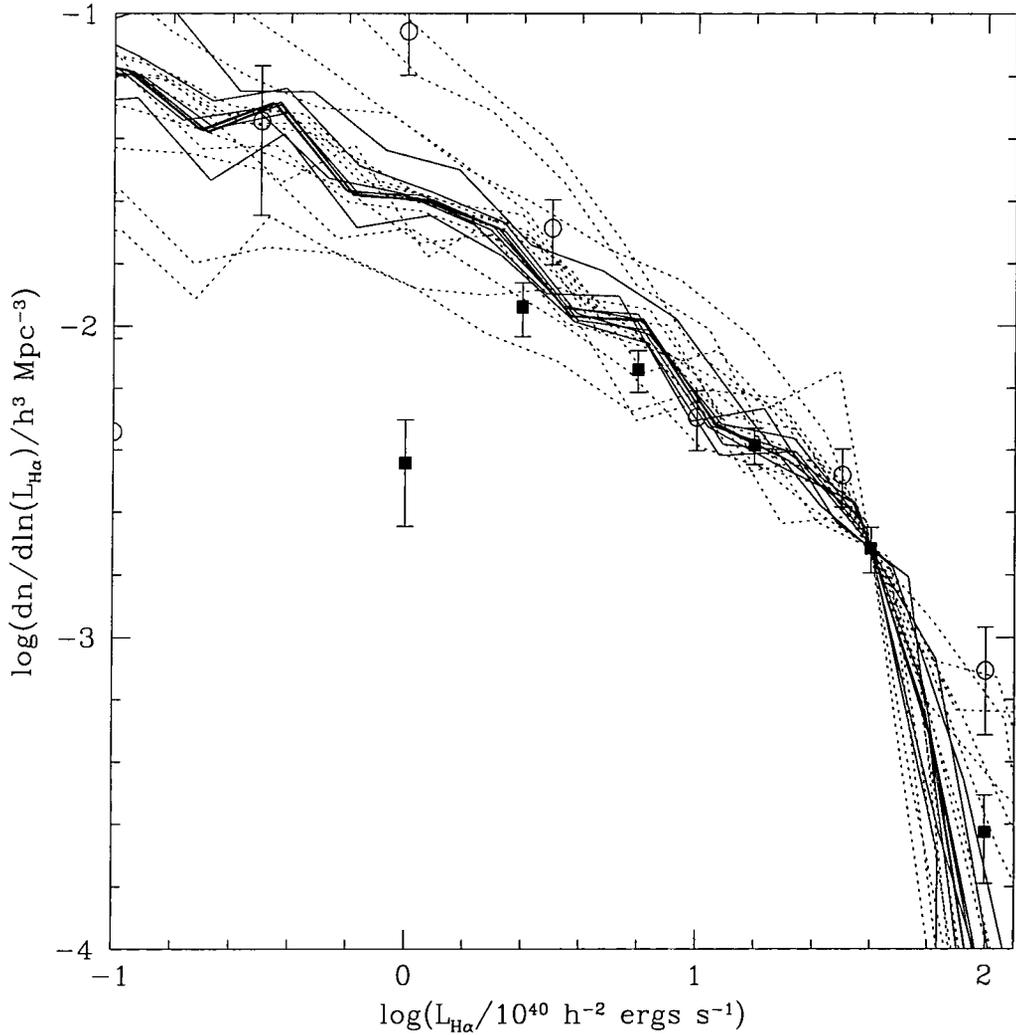


Figure 5.12: The $z = 0$ $H\alpha$ luminosity functions of our variant models. The solid lines show the variant models, whilst the dotted lines indicate the variants models marked with a † in Table 5.2. Points with error bars are observational data from Gallego et al. (1995a) (filled squares) and Sullivan et al. (2000) (open circles). These points include a correction for dust extinction.

Table 5.2: Results of the variant models at $z = 5$. Models marked by a † are those which have very different $H\alpha$ luminosity functions compared to the standard model (see Fig. 5.12). The second column lists those parameters which differ from the standard model. The third column lists the value of Υ required for each model to match the $H\alpha$ luminosity function. Columns 4–5 list the mean escape fraction (including the effects of dust) in the fixed and DS94 models. Finally, columns 6–7 list the filling factors at $z = 0$ in the fixed and DS94 models, for the case $f_{\text{clump}} = f_{\text{clump}}^{(\text{halos})}$. Note that filling factors will exceed 1 if more photons than required to reionize the universe have been produced.

Model	Parameter change(s)	Υ	$f_{\text{escape}}(\%)$		F_{fill}	
			Fixed	($f_{\text{esc,gas}} = 0.1$)	DS94	Fixed
Standard	None	1.53	3.9	2.7	1.08	0.76
1†	$\Omega_b = 0.04$	2.87	1.7	0.8	0.29	0.13
2†	$\Omega_b = 0.01$	0.36	6.4	7.6	6.49	7.60
3†	$\alpha_{\text{hot}} = 0.5$	1.22	2.8	1.6	1.54	0.87
4†	$\alpha_{\text{hot}} = 4.0$	1.66	3.2	2.0	0.49	0.32
5†	$V_{\text{hot}} = 300 \text{ km/s}$	0.95	5.5	4.6	1.22	1.03
6†	$V_{\text{hot}} = 50 \text{ km/s}$	0.91	2.2	1.3	2.01	1.13
7†	$\epsilon_* = 0.020$	1.74	3.9	3.3	1.25	1.07
8	$\epsilon_* = 0.005$	1.19	3.9	2.4	1.02	0.60
9	$f_{\text{d1}} = 5.0$	1.47	4.0	2.9	1.14	0.83
10	$f_{\text{d1}} = 0.2$	1.49	2.6	1.7	1.04	0.68
11†	$f_{\text{ellip}} = 0.05$	0.86	2.8	2.0	1.78	1.27

continued on next page

Table 5.2: (continued)

Model	Parameter change(s)	Υ	$f_{\text{escape}}(\%)$			F_{fill}	
			Fixed ($f_{\text{esc,gas}} = 0.1$)	DS94	Fixed	DS94	Fixed
12	$f_{\text{ellip}} = 0.60$	1.63	4.0	2.9	1.02	0.72	
13	$f_{\text{dyn}} = 5.0$	1.53	4.0	2.8	1.08	0.77	
14	$f_{\text{dyn}} = 0.1$	1.53	4.0	2.9	1.08	0.77	
15 [†]	$p = 0.04$	1.20	3.0	2.4	1.01	0.78	
16 [†]	$p = 0.01$	1.46	5.1	3.3	1.47	0.96	
17	$R = 0.50$	1.86	3.9	2.7	0.89	0.63	
18 [†]	$R = 0.10$	1.23	4.1	3.0	1.34	0.96	
19	$h_z = 0.5$	1.53	4.0	1.6	1.08	0.42	
20 [†]	Feedback: modified ($a_r = 0.75$)	0.58	2.7	1.5	3.25	1.82	
21	$S_2 = 20.0 \times 10^{50}$ photons/s	1.53	4.0	3.4	1.08	0.91	
22	$S_2 = 0.5 \times 10^{50}$ photons/s	1.53	4.0	1.4	1.08	0.36	
23	Starbursts: Not included	1.60	4.0	2.9	1.04	0.73	
24	IMF: Salpeter (1955)	1.45	4.0	2.8	1.06	0.75	
25 [†]	Gas profile: SIS	1.59	2.7	1.7	1.00	0.64	
26 [†]	Recooling: allowed	1.27	3.3	2.2	1.50	0.97	

Other models which alter the strength of feedback (i.e. Models 3, 4, 5 and 6) also cause large changes in the filling factors. Models with weaker feedback (i.e. Models 3 and 6) result in larger filling factors as they allow more star formation to occur in low mass galaxies (these models again producing a steep slope for the faint end of the $H\alpha$ luminosity function). The value of Ω_b also has a strong influence on the filling factors as demonstrated by Models 1 and 2. Finally, in Models 21 and 22 we consider two alternative values of S_2 in the DS94 model. These values span the range of uncertainty for the maximum OB association luminosity in our own Galaxy (Dove & Shull 1994). These models demonstrate that the filling factors predicted by the DS94 model are uncertain by a factor of at least 2 simply because of this uncertainty in the value of S_2 . There is, in fact, further uncertainty introduced as it is not clear if S_2 represents a real cutoff in the luminosity function of OB associations, or merely a turn-over in that function.

All of the models which significantly alter the predicted filling factors are amongst those marked with a † in Table 5.2, indicating that such models do not reproduce well the $z = 0$ $H\alpha$ luminosity function, and can therefore be discarded as being unrealistic. With these models removed, our predictions for F_{fill} are reasonably robust. Considering all the realistic models we find that for the fixed gas escape fraction of 10%, F_{fill} at $z = 5$ is $1.08^{+0.06}_{-0.19}$ (where the value indicates the filling factor in the standard model and the errors show the range found in the realistic variant models). For the DS94 model we find $F_{\text{fill}} = 0.76^{+0.15}_{-0.40}$ (leaving out the models which vary S_2 we find $F_{\text{fill}} = 0.76^{+0.07}_{-0.34}$). Our conclusion that with the DS94 escape fractions reionization cannot happen by $z = 5$ if the clumping factor is as large as $f_{\text{clump}}^{(\text{halos})}$ remains valid under all realistic parameter variations considered here. On the other hand, if the clumping factor is closer to the case of a uniform IGM, then reionization by $z = 5$ is possible in the DS94 model (but not in the DSGN98 model).

So far we have considered a single cosmology, namely Λ CDM. This choice was motivated by the work of Cole et al. (2000) and Chapters 3 & 4, where the semi-analytic model has been shown to reproduce many features of the observed galaxy population for this cosmology. However, in order to explore the effects of cosmological parameters on reionization, we have also considered a τ CDM cosmology, with $\Omega = 1$, in which we use model parameters identical to those of Chapter 3. We note that we were unable to match the galaxy correlation function at $z = 0$ for this cosmology (Chapter 3).

We find that in the τ CDM model, our basic results are unchanged, i.e. with physical models for the escape fraction, the IGM is reionized by $z = 5$ only if it is much less

clumped than in our halo clumping model with $f_{\text{clump}} = f_{\text{clump}}^{(\text{halos})}$. The escape fractions in this cosmology are actually somewhat higher than in the Λ CDM cosmology (due to a lower amount of gas and dust in galaxies). At $z = 5$ the DS94 model predicts a mean escape fraction $\approx 16\%$, whilst the DSGN98 model predicts $\approx 0.1\%$. However, the filling factors are significantly lower (for example, in the DS94 model $F_{\text{fill}} = 0.24$ at $z = 5$ when $f_{\text{clump}}^{(\text{halos})}$ is used as compared to $F_{\text{fill}} = 0.76$ in Λ CDM). This reflects the fact that many fewer ionizing photons are produced in this cosmology (due to the fact that a stronger feedback is required in order that the model fits the properties of galaxies at $z = 0$), and that less of the gas has become collisionally ionized in virialised halos in τ CDM than in Λ CDM. The only factor which works in favour of a higher filling factor in τ CDM is that the clumping factor is somewhat lower. However, this is not enough to offset the two effects described above.

5.6 Spatial Distribution of Ionizing Sources and CMB Fluctuations

5.6.1 Spatial Distribution

We now consider the temperature anisotropies imprinted on the microwave background by the IGM following reionization. These depend on the spatial and velocity correlations of the ionized gas. A fully self-consistent calculation of these correlations on the relevant scales would require very high resolution numerical simulations including both gas dynamics and radiative transfer (e.g. Abel & Haehnelt 1999). No such numerical simulation is yet available with the necessary combination of volume and resolution to calculate the secondary CMB anisotropies on all angular scales of interest. Therefore in this Chapter, we calculate the spatial and velocity distribution of the ionized gas in an approximate way, by combining our semi-analytical galaxy formation model with a high resolution N-body simulation of the dark matter.

We have used the same Λ CDM simulation as in Chapter 3 (and which is described in detail by Jenkins et al. 1998), which has $\Omega_0 = 0.3$, a cosmological constant $\Lambda_0 = 0.7$, a Hubble constant of $h = 0.7$, and which is normalised to produce the observed abundance of rich clusters at $z \approx 0$ (Eke, Cole & Frenk 1996). Using the same semi-analytic model as employed here, we were able to match the observed galaxy two-point correlation function at $z = 0$ in this cosmology (see Chapter 3). The simulation has a box of length 141.3

h^{-1} Mpc and contains 256^3 dark matter particles, each of mass of $1.4 \times 10^{10} h^{-1} M_{\odot}$. We identify halos in this simulation using the friends-of-friends (FOF) algorithm with the standard linking length of 0.2, and then populate them with galaxies according to the semi-analytic model. We consider only groups consisting of 10 particles or more, and so resolve dark halos of mass $1.4 \times 10^{11} h^{-1} M_{\odot}$ or greater. Sources in halos which are unresolved in the simulations can produce a significant fraction of the total ionizing luminosity, according to the semi-analytic models. To take this into account, we add sources in unresolved halos into the simulation in one of two ways. The first method is to place the sources on randomly chosen dark matter particles which do not belong to any resolved halo. An alternative method is to place these sources at random within the simulation volume. This makes the unresolved sources completely unclustered and so is an interesting extreme case. As we will be forced to construct toy models to determine which regions of the simulation are ionized, the exact treatment of these unresolved halos will not be of great importance. The number of unresolved halos added to the simulation volume is determined from the Press-Schechter mass function, multiplied by a correction factor of 0.7 to make it match the low mass end of the N-body mass function in Λ CDM at $z = 3$.

In order to calculate the correlations between ionized regions that are needed to determine the temperature anisotropies induced in the CMB, a simulation with at least the volume of this one is required. Unfortunately, with present computing resources, this excludes the possibility of an exact calculation of the shape and size of the ionized regions, which would require much higher resolution, and also the inclusion of gas dynamics and radiative transfer. Therefore we have used five toy models to determine which regions of the simulation are ionized, for a given distribution of ionizing sources. These models cover a range of possibilities which is likely to bracket the true case, and provide an estimate of the present theoretical uncertainties.

For each model, we divide the simulation volume into 256^3 cubic cells, resulting in a cell size of $0.55 h^{-1}$ Mpc. As the gas distribution is not homogeneous, the volume of gas ionized will depend on the density of gas in the ionized region. We assume that the ionizing luminosity from the galaxies in each halo all originates from the halo centre, and that the total mass M of gas ionized by each halo is the same as it would be for an IGM which is uniform on large scales, but with small-scale clumping f_{clump} , as given by eqn. (5.4). We add to this the mass of any collisionally ionized gas in the halo. We then calculate the volume of the ionized region around each halo using $M = \bar{n}_{\text{H}} m_{\text{H}} V$, where

\bar{n}_H is the mean IGM density within the volume V . We use several different toy models to calculate the spatial distribution of ionized gas in the simulations. In all cases, the total mass of hydrogen ionized is assumed to be the same as for a homogeneous distribution with the specified clumping factor.

Model A (Growing front model) Ionize a spherical volume around each halo with a radius equal to the ionization front radius for that halo assuming a large-scale uniform distribution of HI. Since the HI in the simulation is *not* uniformly distributed, and also because some spheres will overlap, the ionized volume will not contain the correct total mass of HI. We therefore scale the radius of each sphere by a constant factor, f , and repeat the procedure. This process is repeated, with a new value of f each time, until the correct total mass of HI has been ionized.

Model B (High density model) In this model we ignore the positions of halos in the simulation. Instead we simply rank the cells in the simulation volume by their density. We then completely ionize the gas in the densest cell. If this has not ionized enough HI we ionize the second densest cell. This process is repeated until the correct total mass of HI has been ionized.

Model C (Low density model) As model B, but we begin by ionizing the least dense cell, and work our way up to cells of greater and greater density. This model mimics that of Miralda-Escudé, Haehnelt & Rees (2000).

Model D (Random spheres model) As Model A but the spheres are placed in the simulation entirely at random rather than on the dark matter halos. By comparing to Model A this model allows us to estimate the importance of the spatial clustering of dark matter halos.

Model E (Boundary model) Ionize a spherical region around each halo with a radius equal to the ionization front radius for that halo. This may ionize too much or not enough HI depending on the density of gas around each source. We therefore begin adding or removing cells at random from the boundaries of the already ionized regions until the required mass of HI is ionized.

Fig. 5.13 shows six slices through the N-body simulation. The top left slice shows the density of all gas (which is assumed to trace the dark matter), whilst the other slices show only the density of ionized gas. Model A shows particularly well the correlated nature of the ionizing sources (due to the fact that galaxies form in the high density regions of the dark matter), as the densest regions of the simulations are the ones which have become most highly ionized.

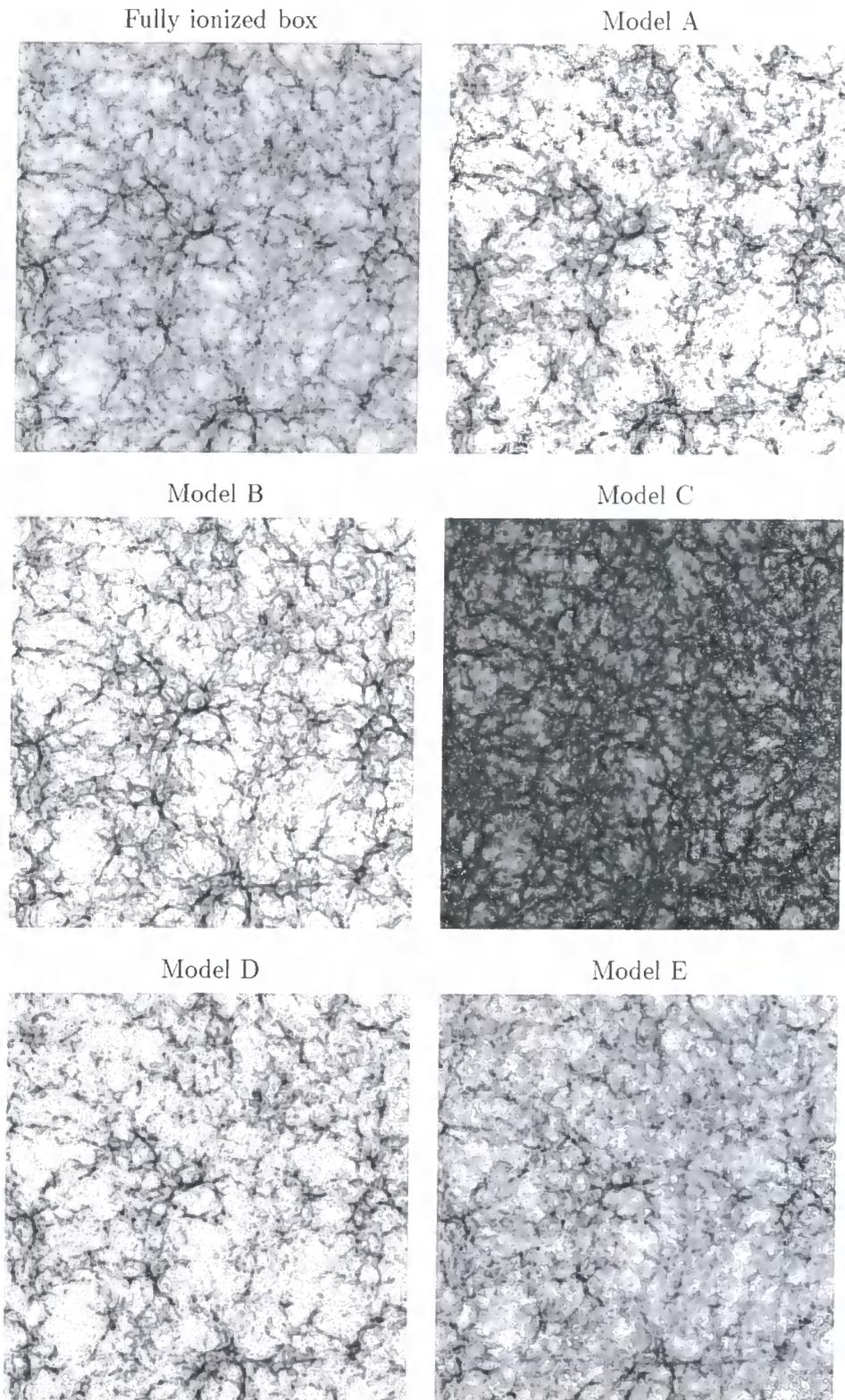


Figure 5.13: The projected density of ionized gas in a slice through the Λ CDM N-body simulation at $z = 3$ shown as a greyscale image, with the densest regions being black. The slice shown has dimensions of $141.3 \times 141.3 \times 8.0h^{-1}$ Mpc. The total (i.e. ionized plus neutral) projected gas density is shown in the upper left hand panel. The remaining panels show the projected density of ionized gas in Models A-E.

In Fig. 5.14 we show the filling factor in the N-body simulation for the fixed $f_{\text{esc,gas}}$ model with different values of $f_{\text{esc,gas}}$, for the case $f_{\text{clump}} = f_{\text{clump}}^{(\text{halos})}$ (we show results for the fixed gas escape fraction rather than our preferred DS94 model as we will use this same model in the following section to explore the effects of varying $f_{\text{esc,gas}}$ on the CMB anisotropies). The filling factors calculated from the simulation are always less (for a given value of $f_{\text{esc,gas}}$) than those calculated in §5.4.1 (see Fig. 5.10). This is because the simulation contains fewer low mass dark matter halos than predicted by the Press-Schechter theory. Hence it contains fewer ionizing sources.

5.6.2 CMB Fluctuations

The reionization of the IGM imprints secondary anisotropies on the CMB through Thomson scattering off free electrons (see, for example, Vishniac 1987, Knox, Scoccimarro & Dodelson 1998, Hu 2000). These anisotropies result from the spatially varying ionized fraction and from density and velocity variations in the ionized IGM. The calculation of these secondary effects involves correlation functions of density fluctuations and velocity fields which are easily determined in our models. To predict the form of these fluctuations, we first calculate the two-point correlations between ionized gas over the redshift range 3 to 18, assuming that gas in the IGM traces the dark matter density and velocity. To do this, we use the 256^3 grid of ionization fractions, x_e , described in §5.6.1. We determine in each grid cell the value of

$$\zeta = \left[\frac{x_e(1+\delta)}{\langle x_e(1+\delta) \rangle} - 1 \right] v_{\text{los}}, \quad (5.10)$$

where δ is the dark matter overdensity in the cell, v_{los} is the component of the mean dark matter velocity in the cell along the line of sight to a distant observer, and the averaging of $x_e(1+\delta)$ is over all cells in the simulation volume. The dark matter density and velocity in each cell are estimated by assigning the mass and velocity of each dark matter particle to the grid using a cloud-in-cell algorithm.

We then compute the correlation function

$$\xi_{\zeta\zeta}(r) = \langle \zeta(\mathbf{x})\zeta(\mathbf{x} + \mathbf{r}) \rangle_{\mathbf{x}}. \quad (5.11)$$

This correlation function is all that is needed to determine the spectrum of fluctuations imprinted in the CMB by the reionization process. A detailed description of how this spectrum is computed is given in Appendix B.3. Code to calculate the CMB anisotropy spectrum from the above correlation functions was kindly provided by Naoshi Sugiyama.

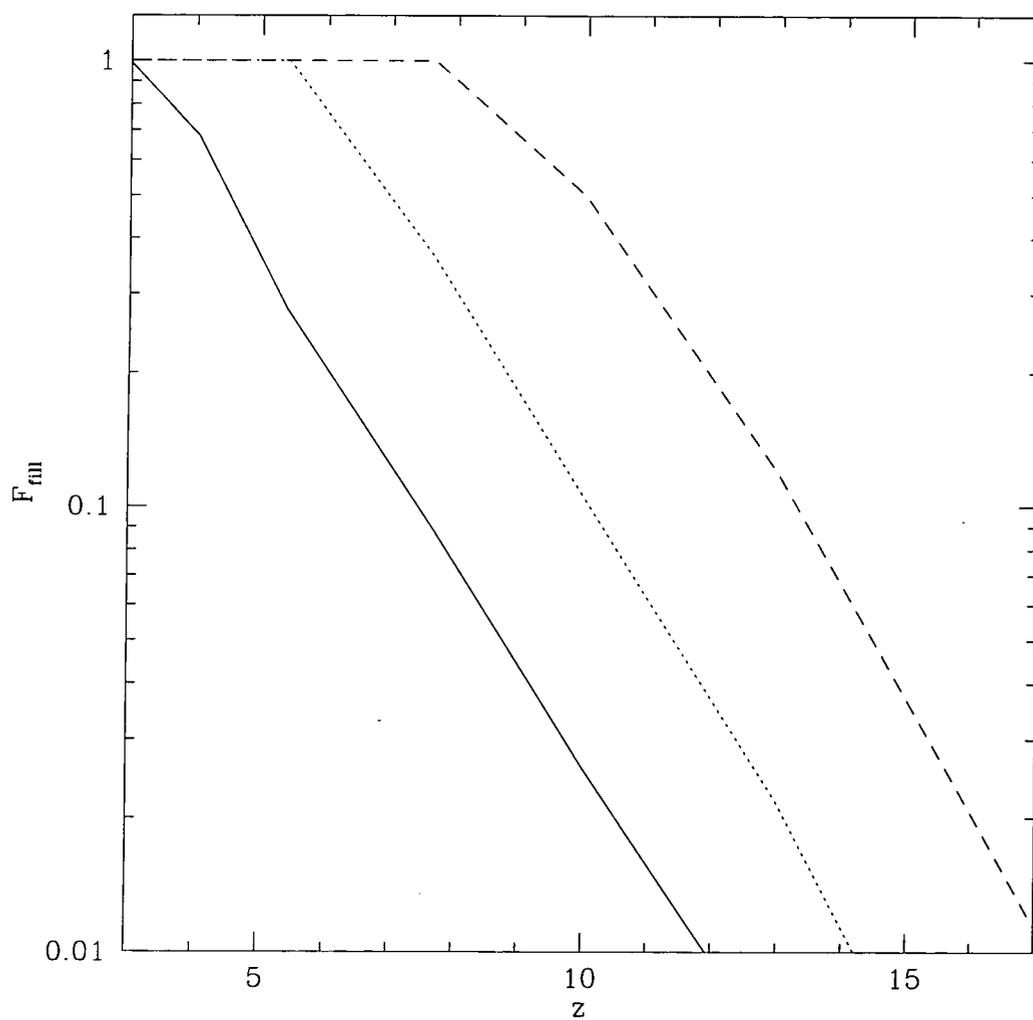


Figure 5.14: Filling factors as a function of redshift in the N-body simulation. Lines are plotted for the ‘fixed’ model with three different values of $f_{\text{esc,gas}}$: 0.05 (solid line), 0.20 (dotted line) and 1.00 (short-dashed line).

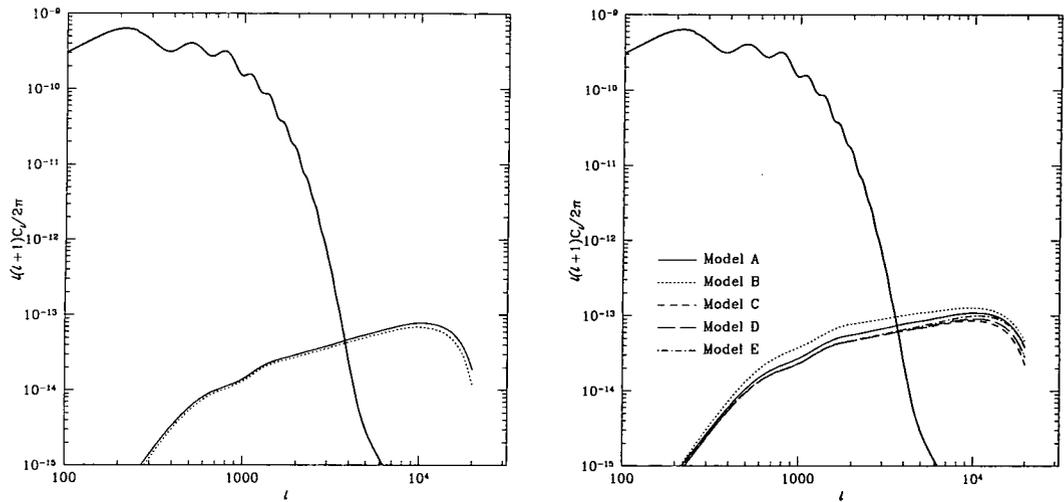


Figure 5.15: The secondary CMB anisotropies measured from the simulation. The left hand panel shows the results for Model E, with a fixed gas escape fraction of 0.10. The solid line indicates a model in which unresolved halos are placed on ungrouped particles, whilst the dotted line shows a model with unresolved halos placed at random within the simulation volume. In the right-hand panel we show the results for $f_{\text{esc,gas}} = 1.00$ and with unresolved halos placed on ungrouped particles. The lines show the results from the five different models as indicated in the figure. In each case, the heavy solid line shows the primary anisotropies in this cosmology.

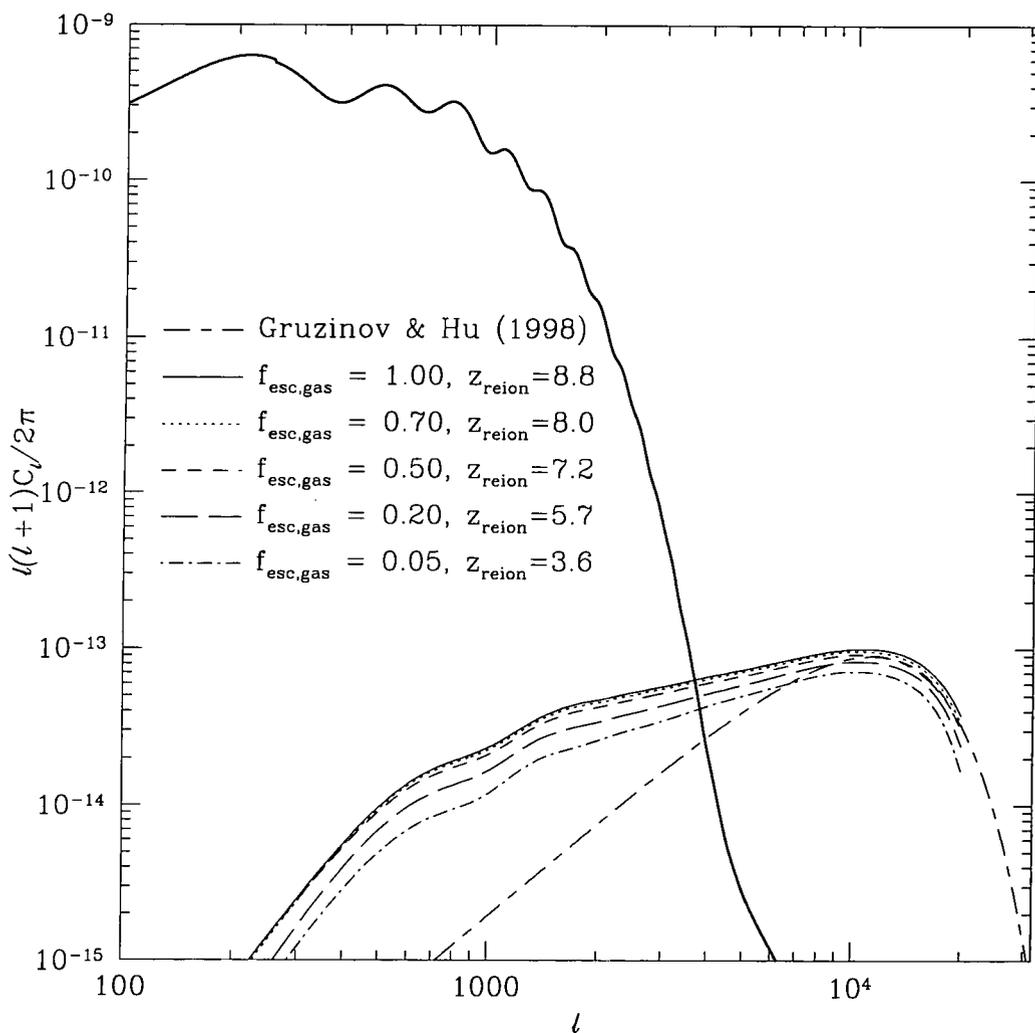


Figure 5.16: The effect of varying the escape fraction, $f_{\text{esc,gas}}$, on the secondary CMB anisotropies. The curves shown are all computed using Model E, with unresolved halos placed on ungrouped particles. Gas escape fractions of 1.00, 0.70, 0.50, 0.20 and 0.05 are shown as indicated in the figure. The redshift of reionization for each model is also indicated. The heavy solid line shows the primary CMB anisotropies.

In Fig. 5.15 we show the secondary CMB anisotropies calculated as described above. The left-hand panel shows the results for Model E with a fixed gas escape fraction of 0.10, for the cases that unresolved halos are placed either on ungrouped particles (solid line) or at random in the simulation volume (dotted line). The clumping factor is $f_{\text{clump}} = f_{\text{clump}}^{(\text{halos})}$, as will be used for all models considered in this section. The particular choice of f_{clump} is not important for our conclusions about the CMB fluctuations, since we will consider different values of $f_{\text{esc,gas}}$. The choice of placement scheme is seen to make little difference to the results, the two curves differing by $\lesssim 10\%$ for $250 < \ell < 5000$, with the difference growing to 40% by $\ell = 20000$. The size of the grid cell used in our calculation of the ionized gas correlation function corresponds to $\ell \approx 25000$ at $z = 3$, and the turnover around $\ell = 10^4$ is simply due to the cell size. Our method is unable to determine the form of the CMB fluctuations at higher ℓ .

The right-hand panel of Fig. 5.15 shows the variations in our estimates of C_ℓ which arise from using the five Models A-E. Here the differences between the curves are larger, with Models B and C differing by a factor of ≈ 2.5 at $\ell = 10^4$. The amplitude of the curves is affected by the strength of the correlations present in each model (e.g. the “high density” model is the most strongly correlated and has the highest amplitude, whilst the “low density” model has the weakest correlations and hence the lowest amplitude). However, the shapes of the curves are all very similar.

Fig. 5.16 examines the effect on the secondary anisotropies of varying the escape fraction $f_{\text{esc,gas}}$. The trend is for increasing amplitude of anisotropy with increasing escape fraction (which results in a higher reionization redshift). If, however, we boost the number of photons produced by increasing $f_{\text{esc,gas}}$ above 1 (this, of course, being an unphysical, but simple way of examining the effects of producing more ionizing photons), little further increase in amplitude is seen.

The form expected for the CMB anisotropies produced by patchy reionization has been calculated for a simple model by Gruzinov & Hu (1998). In this model, reionization of the universe is assumed to begin at some redshift, z_i , and is completed (i.e. the filling factor reaches unity) after a redshift interval δz . Sources are assumed to appear at random positions in space and to each ionize a spherical region of comoving radius R . Once such an ionized region has appeared it remains forever. In this model, the power $\ell^2 C_\ell / 2\pi$ is predicted to have the form of white noise at small ℓ , since the ionized regions are uncorrelated. Gruzinov & Hu (1998) show that in such a model the secondary

anisotropies have the form

$$\ell^2 C_\ell / 2\pi = A \ell^2 \theta_0^2 \exp\left(-\theta_0^2 \ell^2 / 2\right), \quad (5.12)$$

where

$$A = \frac{\sqrt{2\pi}}{36} \tau_0^2 \langle v^2 \rangle \frac{R}{\eta_0} \delta z (1 + z_i)^{5/2} D_V^2, \quad (5.13)$$

and $\theta_0 = R/D_A$. Here η_0 is the horizon distance at $z = 0$, $\tau_0 = \sigma_T n_0 \eta_0$ (where n_0 is the number density of free and bound electrons and σ_T is the Thomson cross section) is the optical depth to Thomson scattering across the present particle horizon, $\langle v^2 \rangle^{1/2}$ is the rms velocity of dark matter today (which we take to be 648 km/s as given by Jenkins et al. 1998), D_A is the comoving angular diameter distance to redshift z_i and D_V is the linear growth factor for velocity to the same redshift. Thus, the amplitude of the C_ℓ spectrum is determined by the optical depth τ_0 , the velocity dispersion $\langle v^2 \rangle$ and the characteristic angular size of ionized regions, R/D_A . The position of the peak is determined entirely by the angular size of the ionized regions.

We compare the simple model of Gruzinov & Hu (1998) with our own results in Fig. 5.16. Since in our model reionization has no well-defined starting redshift, and ionized regions span a range of sizes, we simply choose values of R , z_i and δz in order to match the two models at the peak in the spectrum (even though the position of this peak in our results is an artifact of our simulation resolution, we simply wish to demonstrate here the difference in small ℓ slopes between our model and that of Gruzinov & Hu 1998). The chosen values of $R = 0.85h^{-1}\text{Mpc}$, $z_i = 11$ and $\delta z = 5$ are all plausible for the ionization history and sizes of ionized regions seen in our model (the mean comoving size of regions ranging from $1.4h^{-1}\text{Mpc}$ at $z = 3$ to $0.2h^{-1}\text{Mpc}$ at $z = 18$). Note that Knox, Scoccimarro & Dodelson (1998) calculate a somewhat different form for the anisotropy spectrum for this same model, in which the amplitude, A , is roughly half that found by Gruzinov & Hu, and the peak in the spectrum occurs at slightly higher ℓ . Despite these differences both Gruzinov & Hu (1998) and Knox, Scoccimarro & Dodelson (1998) agree upon the general form of the spectrum (sharp peak plus white noise at small ℓ), and this is all we are interested in here.

The C_ℓ declines much more rapidly as $\ell \rightarrow 0$ in the Gruzinov & Hu (1998) model than in ours. Note that Model D, the random sphere model, also shows the same behaviour as our other models, indicating that it is not the correlated positions of the ionizing sources in our model which produce the excess power at small ℓ . If we force all halos in our model to have equal ionized volumes surrounding them, whilst retaining the same total filling

factor, we find that the excess power above the white noise spectrum at small ℓ remains, so neither is the excess due to the range of ionizing front radii, R , present in our model. This excess power can therefore be seen to be due to the correlations in gas density and velocity induced by gravity. In fact, if we repeat our calculations but ignore correlations in the gas density field (i.e. we set $\delta = 0$ everywhere) we find a CMB spectrum which has a slope for small ℓ which is much closer to the Gruzinov & Hu (1998) white-noise slope, and which has an amplitude over five times lower than when density correlations are included. The remaining differences between our model and that of Gruzinov & Hu (1998) in this case are due to the correlated nature of ionized regions in our model.

The amplitude of the secondary anisotropies also depends upon the assumed value of Ω_b (as this determines the optical depth for electron scattering). The preferred value for our galaxy formation model of $\Omega_b = 0.02$ is relatively low compared to estimates based on light element abundances and big bang nucleosynthesis. Fig. 5.17 shows the effect of increasing Ω_b to 0.04. If the evolution of the ionized regions were the same in both models, we would expect the amplitude to increase by a factor of four (since it is proportional to the square of the baryon density). The evolution of ionized regions is actually quite similar in the two cases, and so the factor of four increase is seen. With this higher value of Ω_b , the secondary anisotropies due to patchy reionization would be potentially detectable above $\ell \sim 3000$.

We note that a similar approach to computing the spectrum of secondary anisotropies due to patchy reionization has been taken by Bruscoli et al. (2000). Using a different model of galaxy formation, Bruscoli et al. (2000) grow spherical ionization fronts around dark matter halos identified in an N-body simulation, and from these they estimate the spectrum of CMB anisotropies produced. The simulations employed by Bruscoli et al. (2000) have higher resolution (but much smaller volume) than the GIF simulations used in our work. Bruscoli et al. (2000) therefore do not have the problem of locating unresolved halos in their simulation, but their calculation of secondary anisotropies is restricted to smaller angular scales (roughly $5 \times 10^3 \lesssim \ell \lesssim 2 \times 10^5$) compared to ours. Furthermore, Bruscoli et al. make some approximations in calculating the anisotropies which we do not, ignoring variations in the total IGM density, and assuming that the ionized fraction is completely uncorrelated with the velocity field. Bruscoli et al. consider only a single model for reionizing the simulation volume. As we have shown, our five toy models for the distribution of ionized regions lead to factors of 2–3 difference in the secondary anisotropy amplitudes, indicating that the results are not very sensitive to the model adopted for the

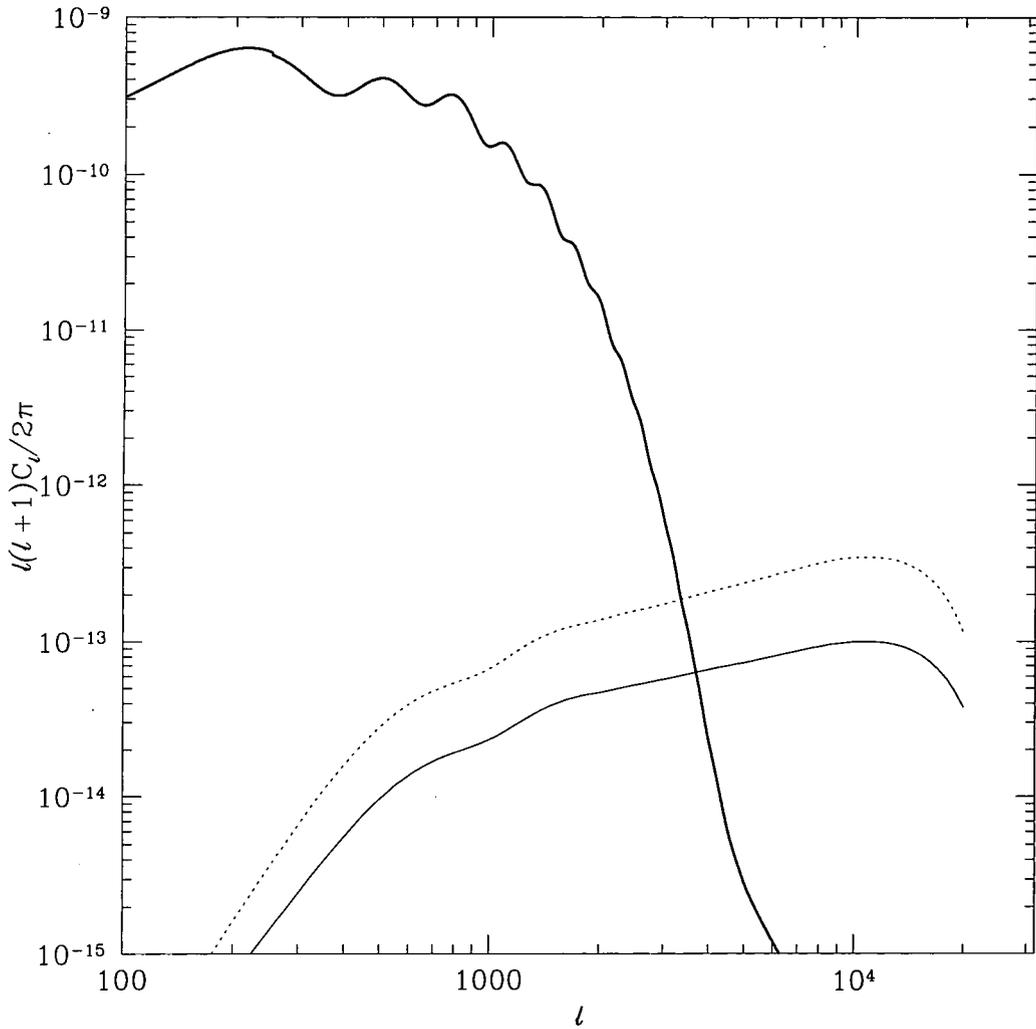


Figure 5.17: The effect of varying Ω_b on the secondary CMB anisotropies. The curves shown are all computed using Model E, with unresolved halos placed on ungrouped particles and a gas escape fraction of 1.0. The solid line shows $\Omega_b = 0.02$ whilst the dotted line shows $\Omega_b = 0.04$ (Model 1 in Table 5.2). The heavy solid line shows the primary CMB anisotropies.

distribution of ionized regions or to the treatment of unresolved halos in the simulation. Bruscoli et al. carried out their calculations in a different cosmology to ours, also with a different value of Ω_b , but once these differences are taken into account, their results seem reasonably consistent with ours.

5.7 Discussion and Conclusions

We have outlined an approach to studying the reionization of the universe by the radiation from stars in high redshift galaxies. We have focussed on the reionization of hydrogen, but the approach can be generalised to study helium reionization (e.g. Giroux & Shapiro 1994), and also to include radiation from quasars. Our main conclusions are:

(i) Using a model of galaxy formation constrained by several observations of the local galaxy population, enough ionizing photons are produced to reionize the universe by $z = 11.7$. This assumes that all ionizing photons escape from the galaxies they originate in, and that the density of the IGM is uniform. Reionization is delayed until $z \approx 10.9$ in the case of a clumped IGM, in which gas falls into halos with virial temperatures exceeding 10^4K . Galaxies can reionize such a clumped IGM by $z = 5$ providing that, on average, at least 4% of ionizing photons can escape from the galaxies where they are produced. In the case of a uniform IGM, an escape fraction of only 1.4% is sufficient to reionize by $z = 5$.

Using a physical model for the escape of ionizing radiation from galaxies, in which photons escape through “HII chimneys” ionized in the gas layers in galaxy disks (Dove & Shull 1994, hereafter DS94), we predict reionization by $z = 6.1$ for a uniform IGM or by $z = 4.5$ for a clumped IGM. Models which assume that all the gas in galaxy disks remains neutral are unable to reionize even a uniform IGM by $z = 0$. Using alternative estimates of the IGM clumping factor from Gnedin & Ostriker (1997) or Valageas & Silk (1999), we find reionization redshifts comparable with those found using our own clumping model, i.e. in the range $z = 4.5\text{--}5.0$ with the DS94 model for the escape fraction.

(ii) Once the ionizing escape fraction and IGM clumping factor have been specified, our estimates for the filling factor of ionized gas in the IGM are reasonably robust, providing that we consider only models which are successful in matching the $\text{H}\alpha$ luminosity function of galaxies at $z = 0$. By far the greatest remaining influences on the ionized filling factor come from the value of the baryon fraction Ω_b and the prescription for feedback from supernovae. However, we have shown that altering these parameters also produces large

changes in the $z = 0$ $H\alpha$ luminosity function.

(iii) We combined our model for reionization with N-body simulations of the dark matter distribution in order to predict the spectrum of secondary anisotropies imprinted on the CMB by the process of reionization. The *shape* of this spectrum is almost independent of the assumptions about reionization, but the *amplitude* depends on the spatial distribution of the ionized regions, the redshift at which reionization occurs and the baryon fraction. We find considerably more power in the anisotropy spectrum at small ℓ than predicted by models which do not account for the large-scale correlations in the gas density and velocity produced by gravity. Despite the uncertainty in the spatial distribution of ionized regions, we are able to determine the amplitude of this spectrum to within a factor of three for a given Ω_b (the amplitude being proportional to Ω_b^2). The results found by Bruscoli et al. (2000) using a similar technique are reasonably consistent with ours, once differences in Ω_b and other cosmological parameters are allowed for.

Detection of these secondary anisotropies, which would constrain the reionization history of the Universe, would require fractional temperature fluctuations of $\sim 10^{-7}$ to be measured on angular scales smaller than several arcminutes. Although the Planck and MAP space missions are unlikely to have sufficient sensitivity to observe such anisotropies, the Atacama Large Millimeter Array is expected to be able to measure temperature fluctuations of the level predicted at $\ell \sim 10^4$ in a ten hour integration (or less if Ω_b is greater than we have assumed).

Previous studies of reionization have either used an approach similar to our own, i.e. employing some type of analytical or semi-analytical model (e.g. Haiman & Loeb 1996, Valageas & Silk 1999, Chiu & Ostriker 1999, Ciardi et al. 2000), or else have used direct hydrodynamical simulations (e.g. Gnedin & Ostriker 1997). While the latter technique can in principle follow the detailed processes of galaxy formation, gas dynamics and radiative transfer, in practice the resolutions attainable at present do not allow such simulations to resolve the small scales relevant to this problem. Furthermore, the implementation of star formation and feedback in such models is far from straightforward.

There are two main uncertainties in our approach, as in most others: the fraction f_{esc} of ionizing photons that escape from galaxies, and the clumping factor f_{clump} of gas in the IGM. Future progress depends on improving estimates of the effects of clumping using larger gas dynamical simulations, on better modelling of the escape of ionizing photons from galaxies, and on better understanding of star formation and supernova feedback in high redshift objects.

Chapter 6

Diffuse X-ray Emission from Late-Type Galaxy Halos

6.1 X-rays from Galaxy Formation

In the model of galaxy formation by hierarchical clustering described in Chapter 2, the disks of spiral galaxies form by late accretion of gas which cools from an extended reservoir around the galaxy. In this and other, similar models disks are still expected to be growing at the present day. Such models have been very successful in accounting for properties of galaxies at optical and infrared wavelengths as shown in §2.7 (see also Kauffmann, White & Guiderdoni 1993, Cole et al. 1994, Baugh, Cole & Frenk 1996b, Kauffmann et al. 1999a). The proposed disk formation mechanism should produce a signature at X-ray wavelengths which, in principle, is detectable with the present generation of satellites.

At the virial temperature of galaxy halos, the dominant cooling mechanism is X-ray Bremsstrahlung. If the cooling rate is significant, this flux may be visible as a component of diffuse X-ray emission extending well beyond the galaxy's optical radius. In this Chapter, we set out to test this disk formation model by looking for X-ray emission from the hypothesized reservoir of cooling gas around three large, late-type galaxies, NGC 2841, NGC 4594 and NGC 5529.

If we assume that the mass of the present disk has been built up by a constant rate of accretion over the age of the universe (t_0), a simple estimate of the X-ray luminosity can be made from the binding energy of the disk.

$$L_X \approx \frac{\frac{1}{2} M_{\text{disk}} v_{\text{esc}}^2}{t_0} \quad (6.1)$$

where M_{disk} is the present mass of the disk and v_{esc} is the velocity required by material at the edge of the disk to escape the halo. The escape velocity is found by equating $v_{\text{esc}}^2/2$

with the potential energy needed to move unit mass from the edge of the disk (which we take to correspond approximately to the optical radius) to infinity:

$$\frac{v_{\text{esc}}^2}{2} = \int_{r_{\text{disk}}}^{\infty} \frac{GM(r)}{r^2} dr, \quad (6.2)$$

where $M(r)$ is the total mass of the halo within radius r . Assuming an isothermal profile truncated at the virial radius, r_{vir} , in which $GM(r)/r = V_C^2$ is constant with radius (here V_C is the measured circular velocity of the disk) the above becomes

$$\frac{v_{\text{esc}}^2}{2} = \int_{r_{\text{disk}}}^{r_{\text{vir}}} \frac{V_C^2}{r} dr + \int_{r_{\text{vir}}}^{\infty} \frac{V_C^2 r_{\text{vir}}}{r^2} dr = V_C^2 [\ln(r_{\text{vir}}/r_{\text{disk}}) + 1], \quad (6.3)$$

or, $v_{\text{esc}} = V_c [2 \ln(r_{\text{vir}}/r_{\text{disk}}) + 2]^{1/2}$. Adopting parameters suitable for a large spiral galaxy such as NGC 2841 (ie., $M_{\text{disk}} = 5 \times 10^{10} M_{\odot}$, and $V_c = 317$ km/s), this suggests that the halo should emit with an average bolometric luminosity of at least $\sim 8 \times 10^{41}$ ergs s^{-1} *. If supernovae and other feedback processes are effective in heating halo gas then even more energy must be radiated to assemble the observed disk. Note that if we assume the mass of the disk is proportional to that of its dark matter halo then $M_{\text{disk}} \propto V_C^3$ so eqn. (6.1) predicts $L_X \propto V_c^5$. Furthermore, the temperature of the cooling gas should scale as the halo virial temperature, $T_X \propto V_c^2$. Thus, it is important to search for emission in the most massive systems available.

The picture of hierarchical galaxy formation developed by White & Frenk (1991, WF91; see also Kauffmann et al. 1993, Cole et al. 1994) and laid out in detail in Chapter 2 predicts halo X-ray luminosities more directly by computing the cooling rate of gas as a function of time. At the present epoch, such models (which are tuned to reproduce the observed optical properties of galaxies) predict an X-ray luminosity of about 10×10^{41} ergs s^{-1} for galaxies like NGC 2841.

To date, studies of X-ray emission from nearby galaxies (e.g. Read, Ponman & Strickland 1997) have concentrated on X-ray emission on scales comparable to the optical image of the galaxy. While extended emission has been detected in some groups of galaxies (e.g. Trinchieri, Kim & Fabbiano 1994, Mulchaey et al. 1996, Trinchieri, Fabbiano & Kim 1997), this emission is thought to be associated with the group potential. Our aim here is to test for the equivalent emission in *isolated* spiral galaxies, which we would identify with a cooling flow supplying the galaxy disk with gas. A few studies have reported diffuse emission from isolated elliptical galaxies. However, this has been interpreted either as

*Unless otherwise stated we have assumed $h = 0.5$, $\Omega_0 = 1$, $\Lambda_0 = 0$, $\Omega_b = 0.06$ and $\sigma_8 = 0.67$ in this Chapter.

Table 6.1: Observed properties of the targeted galaxies. Shown for each galaxy are its heliocentric radial velocity, distance, W_{20} (the width of the 21cm line at 20% of its peak intensity), N_H (the column density of neutral hydrogen in the Milky Way in the direction of the galaxy), r_{optical} (the 25 B-magnitudes per square arcsecond isophotal radius of each galaxy measured along the major axis), M_B (the absolute B-band magnitude) and morphological type.

Heliocentric							
Name	radial velocity (km/s) ^a	Distance (Mpc)	W_{20} (km/s) ^d	N_H (10^{20} cm ⁻²) ^e	r_{optical} (arcmin) ^d	M_B	Hubble type
NGC 2841	638 ± 3.0	13.8 ± 5.2^b	674 ± 4	1.45	4.1 ± 0.04	-20.6 ± 0.2	Sb
NGC 4594	1091 ± 5.1	24.54 ± 2.5^c	753 ± 7	3.67	4.4 ± 0.04	-23.0 ± 0.2	Sa
NGC 5529	2885 ± 5.1	57.72 ± 2.5^c	592 ± 5	1.09	3.3 ± 0.07	-21.1 ± 0.2	Sc

^a From the NASA Extragalactic Database (NED).

^b de Vaucouleurs (1979).

^c From heliocentric radial velocity plus peculiar velocity correction assuming $H_0 = 50$ km s⁻¹ Mpc⁻¹.

^d From the RC3 catalogue (available from NED), corrected for inclination.

^e Dickey & Lockman (1990).

resulting from the expulsion of gas from the galaxy (e.g. Read & Ponman 1998, Mathews & Brighenti 1998) or as a relic of a collapsed group (Ponman et al. 1994). Furthermore, since in hierarchical models of galaxy formation ellipticals are not necessarily expected to be accreting gas at the present day, the existence or otherwise of a diffuse X-ray component associated with them does not provide a strong test of such theories. For this reason, we have restricted our study to isolated spiral galaxies.

Previous X-ray observations of spirals (see, for example, Fabbiano & Juda 1997, Burstein et al. 1997, Pelligrini & Ciotti 1998, Brighenti & Matthews 1999) do not place strong constraints on the models we aim to test. These studies have focussed on measuring X-ray fluxes in regions comparable to the optical radius of the galaxy, where the intensity is strongest. X-ray emission from galactic sources (such as X-ray binaries) and uncertainties in the spatial X-ray surface brightness profile make it difficult to determine what fraction, if any, of the observed X-ray flux could be due to a low surface brightness cooling flow.

Since the cooling gas must, at some point, settle into the galaxy's disk, some emission is expected from this region as the gas cools from the virial temperature of the halo to much lower temperatures. For example, in the case of NGC 4594 models of the type considered here predict a luminosity comparable to that measured by Fabbiano & Juda (1997), $8.9 \pm 1.4 \times 10^{40}$ ergs/s in the 0.1-2.4keV band within the optical radius of the galaxy. However, likely contamination by galactic sources complicates the interpretation of this comparison.

In order to test the models in a manner free from uncertainties of the type just described it is essential to look for low surface brightness emission at large distances from the optical disk of the galaxy. It is also necessary to study the optically brightest galaxies which have the highest circular velocities since the expected X-ray luminosity depends so strongly on the circular velocity of the galaxy's halo. For this reason, we have chosen for this study nearby galaxies with particularly high circular velocities. Most previous work has studied relatively low mass galaxies from which only relatively weak emission is predicted.

If diffuse X-ray emission from gas in the halos of isolated spiral galaxies proves undetectable, this need not imply that the mass of halo gas is small since the gas may simply be too diffuse to emit X-rays efficiently. However, it would then be unable to cool and flow to the centre of the halo. We are thus testing whether the disks of spiral galaxies are being built up by the accretion of halo gas at the present day. The alternative is that

spiral disks were assembled at higher redshift and evolved as essentially closed systems thereafter.

The structure of this Chapter is as follows. In §6.2, we search for large-scale diffuse emission from massive spiral galaxies. We are not able to make any convincing detection; the upper limits we establish are almost an order of magnitude below the predictions of the WF91 model. In §6.3 we discuss the uncertainties in the models and describe how the models might be modified to agree with the data. The prospects for future investigations are outlined in §6.4.

6.2 Observations

6.2.1 Target Selection

The primary aim of this Chapter is to search for diffuse X-ray emission from nearby isolated spiral galaxies. The galaxy must be nearby for the signal to be detectable. Furthermore, the galaxy must have a high circular velocity in order that the energy released by the cooling gas be large. A high circular velocity also favours detectability because it implies a high characteristic temperature for the emission and so less absorption by neutral hydrogen in the Milky Way. In addition the efficiency of supernovae feedback is expected to decline with increasing circular velocity.

We created a suitable target list by searching for all spiral galaxies with recession velocities below 3000 km s^{-1} and with HI 20% line widths greater than 580 km s^{-1} . Of the sixteen galaxies satisfying these criteria three have been observed by the ROSAT PSPC: NGC 2841, NGC 4594 (the Sombrero galaxy) and NGC 5529. The PSPC data for these fields were obtained from the Leicester Database and Archive Service (LEDAS).

- NGC 2841. The Sb galaxy NGC 2841 is the nearest of our target galaxies. It has a flat rotation curve extending to the limit of HI observations, and is an archetype of galaxies with massive dark matter halos (Kent 1987). Its star formation properties have been studied extensively by Young et al. (1996) using both $H\alpha$ and far infrared indicators.
- NGC 4594. Also known as M104 or the “Sombrero Galaxy,” this is a highly luminous Sa seen almost edge on. Since this galaxy is bulge dominated, diffuse X-ray emission similar to that seen around E or S0 galaxies (Forman, Jones & Tucker 1985) might be expected. However, X-ray observations by the *Einstein* satellite revealed

Table 6.2: Measured rotation velocity and predicted gas temperature for the galaxies in our sample.

Name	V_c (km s ⁻¹) ^a	T_{vir} (keV) ^b
NGC 2841	317 ± 2	0.287 ± 0.004
NGC 4594	358 ± 4	0.366 ± 0.009
NGC 5529	277 ± 3	0.219 ± 0.005

^a Calculated from HI line widths using the expression given in Tully & Fouqué (1985).

^b Determined from the measured rotation velocity of the galaxy extrapolated to the halo virial radius assuming a flat rotation curve and by then applying the relation given by White & Frenk (1991).

spectral properties and core X-ray colours typical of less massive spiral galaxies (Kim, Fabbiano & Trinchieri 1992).

- NGC 5529. This is an Sc galaxy seen almost edge-on. Because of its relatively large distance, this galaxy is likely to provide weaker constraints than the previous two.

Distances, line widths, optical radii, and other relevant properties of these galaxies are listed in Table 6.1. The distances to two of them (NGC 4594 and NGC 5529) had to be determined from their observed heliocentric radial velocity; a model of the local peculiar velocity field (Branchini et al. 1999) was used to correct the observed velocity to the true Hubble velocity using an iterative procedure. Table 6.2 lists the circular velocity of each galaxy as calculated from its HI line-width and also the virial temperature of the halo estimated, as in WF91.

6.2.2 Data Reduction

Ideally, we would like to compare the X-ray flux from an annulus centered on the galaxy (but excluding the optically bright region to avoid contamination from stellar emission) with similar annuli centred on blank fields. This is not possible, however, because different ROSAT fields have different particle backgrounds. We determine the background for each field from a large radius annulus in the outer part of the PSPC detector as described in Snowden et al. (1994). This is then subtracted from the measured count rate in each field. Background corrected ‘on source’ and ‘off source’ fields are then compared

to determine the flux from the diffuse halo of the target galaxies. We find that the background subtraction algorithm of Snowden et al. tends to produce negative flux in the off-source images because of the different point-source detection thresholds in the centre and outer parts of the ROSAT field. Our analysis takes this effect into account.

The analysis of the data was carried out using the Asterix package[†]. The procedure for analysing each image was as follows. Firstly, point sources were detected down to a 3σ threshold and excluded by cutting out a circle centred on each source. The radius of this circle was set equal to the radius of the PSF enclosing 95% of the energy at the position of each source. For each image a background annulus of inner radius 0.5° and outer radius 0.7° was then used to estimate the vignetting corrected background count-rate, and to create two background-subtracted spectra from source-centred annuli with inner radius of 5 arcminutes and outer radii of 9 and 18 arcminutes. Choosing a fixed angular radius ensures that the blank field analysis is the same for each image. The PSPC field of view is obstructed by the structure which supports the mirrors. The structure takes the form of a ring connected to a set of spokes. The larger outer radius for our source-centered annuli was chosen to include the entire region inside the ring of this structure, where the field of view is unobstructed. This is comparable to the predicted X-ray emitting region of the closer of our target galaxies (see §6.3). The $9'$ ring is matched to the X-ray emitting region of the more distant galaxy NGC 5529. The innermost aperture was excluded from the analysis in order to avoid contamination by the diffuse and unresolved stellar X-ray emission of the galaxy itself (c.f. Read, Ponman & Strickland 1997, for example). The X-ray image of each galaxy is compared with the optical image in Figure 6.1.

The resulting spectra were analysed with the XSpec package (Arnaud 1996) to determine the source flux using the *mekal* X-ray emission model (Mewe, Gronenschild & van den Oord 1985, Mewe, Lemen & van den Oord 1986, Kaastra 1992) and *wabs* neutral hydrogen absorption model (based on the photoelectric cross-sections of Morrison & McCammon 1983). For each halo we adopted a virial temperature of 0.2keV. This is slightly below the virial temperatures estimated in Table 6.2, in order to obtain conservative estimates of the bolometric luminosity. The hot halo gas is assumed to have a metallicity of $0.3Z_\odot$ (where Z_\odot is the Solar metallicity). Neutral absorption column densities varied between the fields and were taken from Dickey & Lockman (1990). The only remaining free parameter in these models is the overall normalization. This procedure appropriately weights the contributions of the different energy channels and corrects for

[†]See the Asterix homepage at <http://www.sr.bham.ac.uk/asterix/>.

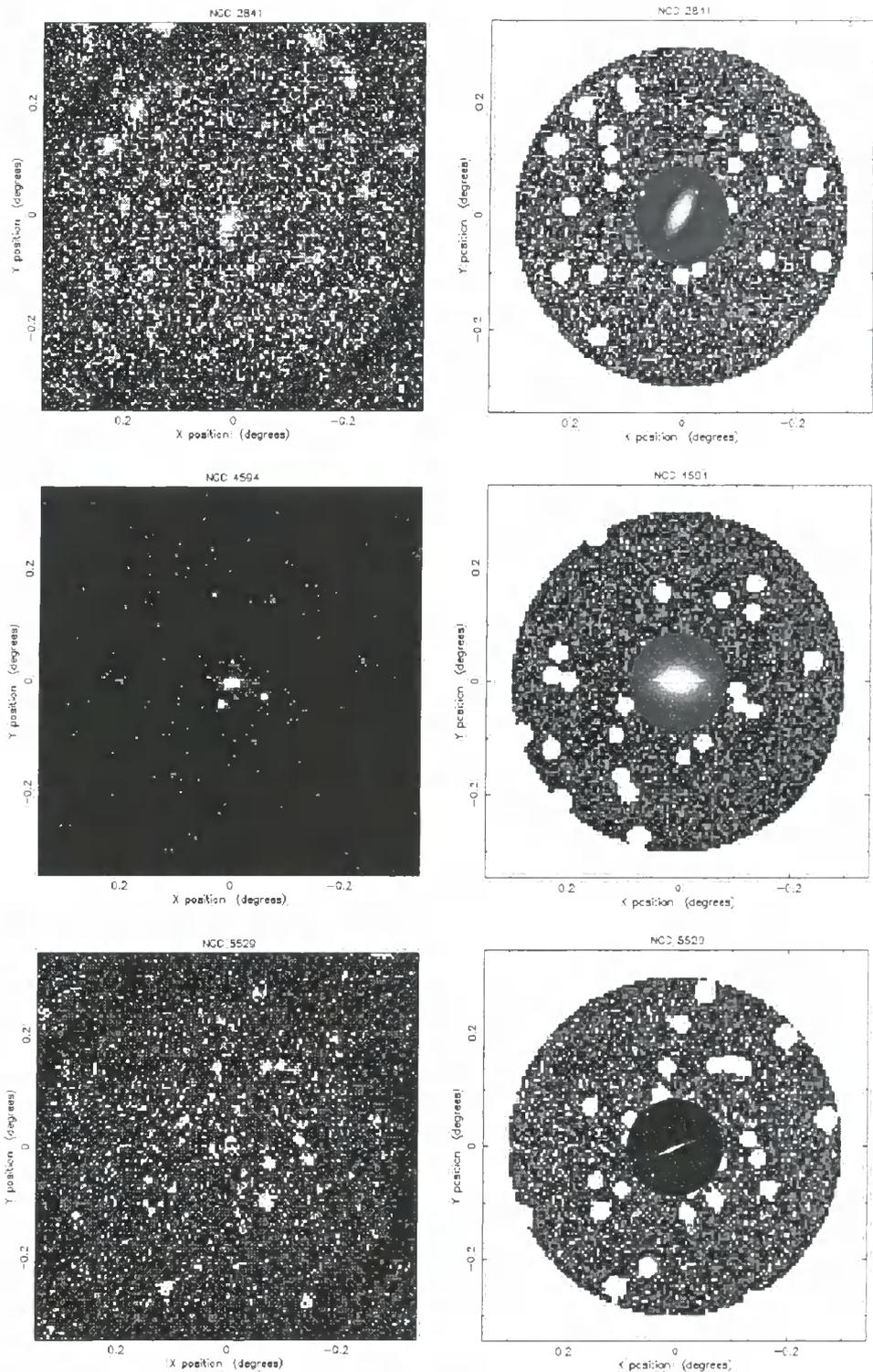


Figure 6.1: Source images for each galaxy in the sample. The left hand column shows the raw image and the right hand column the image after background subtraction, point source removal and selection of the region from 5 to 18 arcminutes. An optical image of each galaxy is superimposed on the X-ray image to give an appreciation of the size of the halo region.

the different neutral hydrogen cross-sections of the fields.

The fitted spectrum is then used to calculate the absorption corrected flux from each halo. These fluxes are determined within the range 0.2 to 2.0 keV in order to minimise dependence on the assumed temperature. Note that it is possible for the flux in the outer (5–18 arcminutes) ring to be lower than that in the inner (5–9 arcminutes) ring since the flux from some regions of the image is negative. These are converted to pseudo-bolometric luminosities (0.001 to 10.0 keV) using the conservative temperature of 0.2 keV and the distances given in Table 6.1.

For comparison, six off-source fields (Table 6.4) with point source targets were selected at random from the ROSAT International X-ray/Optical Survey (RIXOS) (Mittaz et al. 1999). As we have previously noted, the analysis procedure tends to produce a small negative count-rate when applied to the off-source fields. This is due (at least in part) to the different flux limits at which point sources are removed from the target and background annuli. The mean flux deficit in the off-source fields was added to the measured fluxes of the target galaxies to correct for this effect. Alternatively one could add the off-source field count rates to the on-source fields and only then fit a model spectrum to the resulting counts to determine a corrected flux. We investigated performing the analysis in both ways and found that each procedure gives the same result to within 10%. Since the off-source count rates vary from field to field it is possible that the true correction required for each target field is somewhat different from the mean correction from the off-source fields. We therefore measure the variance in the off-source field correction and add this quantity in quadrature to the errors on the fluxes measured in our target fields. The variations from field to field are our dominant source of uncertainty.

The fluxes and luminosities of the target galaxies corrected in this way are presented in Table 6.3. For each galaxy we expect a luminosity around 10×10^{41} ergs s^{-1} . However none of the galaxies is detected at a 3σ confidence level in either of the annuli considered. Instead, our data set 95% upper confidence limits of 0.37, 1.2 and 3.4×10^{41} ergs s^{-1} in the larger aperture for NGC 2841, 4594 and 5529 respectively. For NGC 5529, the relatively weak upper limit is due to the galaxy's larger distance. For the two more nearby galaxies, the upper limit lies an order of magnitude below the initial theoretical prediction.

6.2.3 Uncertainties in the Observations

In §6.3, we consider possible modifications to the model of WF91 that might alleviate the discrepancy with our data. First, we take a closer look at the assumptions we made in

Table 6.3: X-ray fluxes and bolometric luminosities for the halos of three spiral galaxies. The fluxes and luminosities are determined for two annuli, with inner radius of 5 arcminutes and outer radii of 9 (ring 1) and 18 (ring 2) arcminutes. The first column gives the name of the galaxy. The flux corrected for blank field offsets is given in the following two columns, and the final two columns list the bolometric luminosity of each halo corrected for blank field offsets.

Galaxy	Bg. Corr.		Bg. Corr.		Bolometric	
	Flux (ring 1) (10^{-14} ergs/cm ² /s)	Flux (ring 2) (10^{-14} ergs/cm ² /s)	Flux (ring 1) (10^{-14} ergs/cm ² /s)	Flux (ring 2) (10^{-14} ergs/cm ² /s)	L _x (ring 1) (10^{41} ergs/s)	L _x (ring 2) (10^{41} ergs/s)
NGC 2841	5.62 ± 10.33	22.90 ± 19.87	0.04 ± 0.07	0.16 ± 0.14		
NGC 4594	17.96 ± 11.33	14.21 ± 22.40	0.43 ± 0.27	0.34 ± 0.54		
NGC 5529	5.91 ± 10.03	-2.28 ± 19.22	0.75 ± 1.28	-0.29 ± 2.44		

Table 6.4: The total counts per second in six blank fields. Background and point sources were removed in each case and an annulus from 5 to 9 arcminutes (ring 1) and from 5 to 18 arcminutes (ring 2) was selected.

Field	Ring1	Ring 2
rp201558	-0.017 ± 0.007	-0.063 ± 0.015
rp700255	-0.004 ± 0.006	-0.040 ± 0.013
rp700319	-0.021 ± 0.005	-0.024 ± 0.011
rp700376	-0.007 ± 0.005	-0.003 ± 0.011
rp700387	$+0.007 \pm 0.004$	-0.013 ± 0.008
rp700389	-0.005 ± 0.006	-0.023 ± 0.013

interpreting the observational data. There are a number of uncertain quantities that could affect our conclusions. Throughout the analysis we have consistently assumed conservative values for these quantities (i.e. those that would minimize the conflict between model and data). It is nevertheless instructive to consider how these choices might affect our results.

The dominant source of uncertainty in our analysis is the process of background subtraction. Although an accurate estimation of the background from a mosaic of surrounding pointings would be preferable, our off-source fields indicate the magnitude of the uncertainty. Even using the ‘worst case’ field (rp201558), the predicted and observed count-rates still disagree by a factor of 2–3.

Another important source of uncertainty is the extrapolation from ROSAT count-rate to bolometric flux. At the low temperatures appropriate to galactic halo gas, neutral hydrogen absorption in our own Galaxy has a strong effect. If our adopted values for the neutral hydrogen column density were inaccurate, this could have a substantial effect on our results. However these galaxies are at high Galactic latitude and we find that increasing the assumed column density by a factor of 3 over the values given by Dickey & Lockman (1990) only changes the inferred bolometric luminosity by 40%.

The effects of background correction and absorption alone cannot reconcile the observed and predicted luminosities. Altering the assumed temperature of the halo plasma, however, is a more promising alternative. This sensitivity arises because, at the relevant temperatures, most of the emission lies outside the 0.2–2.0 keV band of the observations, forcing us to extrapolate the model fit to lower energies. For example, at a temperature

of 0.2 keV, 90% of the bolometric luminosity is emitted in the ROSAT 0.2–2.0 keV band, while at 0.1 keV, only 75% of the emission lies in this band. At even lower temperatures, the effect becomes even more pronounced. Applying a lower temperature model to the observed spectra, we find that a decrease in temperature by a factor of 4.0 (for NGC 2841) to 4.5 (for NGC 4594) is required to bring the observations into line with the predicted luminosity. It is not clear, however, how these lower temperatures can be accommodated within the theoretical model. We now turn to a detailed examination of this model.

6.3 Discussion

The model presented by WF91 predicts that the X-ray luminosity due to gas cooling within a dark matter halo should be:

$$L_X = \dot{M}_{\text{cool}}(r_{\text{cool}}) V_c^2 \ln \left(\frac{r_{\text{cool}}}{r_{\text{optical}}} \right), \quad (6.4)$$

where $\dot{M}_{\text{cool}}(r_{\text{cool}})$ is the rate at which mass is cooling in the dark matter halo, r_{cool} is the cooling radius (which grows with time), and r_{optical} is the optical radius of the galaxy, which we take to be the extent of the 25 B-magnitudes per square arcsecond isophote along the major axis. This model assumes that the halo has an isothermal potential with circular velocity V_c . Allowing the gas to adopt a profile with a constant density core (as described in Appendix C) makes little difference to the X-ray luminosity provided that the core radius is significantly smaller than the cooling radius. As shown in Appendix C, the model can be readily extended to the case where the gravitational potential follows the functional form favoured by Navarro, Frenk & White (1996, NFW96), which is a good fit to the density profile of halos in N-body simulations.

We have used the formulae and parameter choices in Appendix C to predict the bolometric X-ray luminosity for each of the three galaxies (using the circular velocities of Table 6.2), as well as the region in which cooling is expected to occur. Table 6.5 lists the cooling radii and luminosities of the halos predicted by WF91's standard model and by a model with an NFW potential. In column 5, we list the luminosity that is expected within the 5-18' annulus used in the observational measurements. This geometric correction is made under the assumption of an isothermal flow in quasi-hydrostatic equilibrium as described in Appendix C. The figure given includes a correction for the flux from the target galaxy that is expected in the outer background annulus by extrapolating beyond the cooling radius with an r^{-2} density profile. The corrected flux is about 50% and

70% of the total for NGC 2841 and NGC 4594 respectively. The larger correction for NGC 2841 arises because the emission from the target almost fills the PSPC field of view. The geometric correction for NGC 5529 is greater because a large proportion of the luminosity is projected into the 5 arcminute ring which is excised in order to exclude the optically bright regions of the galaxy. In the case of this galaxy, the large observational uncertainties mean that the emission model is only weakly constrained.

For both NGC 2841 and NGC 4594 the bolometric X-ray luminosities predicted by the model are strongly inconsistent with the observations. The WF91 model predicts cooling rates of 7.4 and $9.0M_{\odot}\text{yr}^{-1}$ for NGC 2841 and 4594 respectively. Since the mass cooling rate is proportional to the X-ray luminosity in these theories, the simplest interpretation is that the cooling rates in these galaxies are also much less than those predicted by WF91. If we assume that the geometric factors would be unchanged by reducing the cooling rate, we arrive at upper limits to the present-day accretion rate of 0.5 and $1.44M_{\odot}\text{yr}^{-1}$.

Further insight can be gained by considering the total mass of gas within the cooling radius and the mass of the galaxy disk. In the case of NGC 2841 the disk mass can be determined by fitting the galaxy rotation curve. Kent (1987) derives a disk mass of $9.3 \times 10^{10}M_{\odot}$ allowing a free-fit of disk and isothermal halo parameters. Adopting an NFW model for the halo, Navarro (2000, N00) derives a best-fit disc mass of $5 \times 10^{10}M_{\odot}$; a significantly smaller disk mass would be inconsistent with the blue luminosity of the disk (Young et al. 1996). The WF91 model prediction for the mass of gas within r_{cool} , which is assumed to have formed the disk, is $9.8 \times 10^{10}M_{\odot}$. Thus the model allows enough gas to cool to make the observed disk. Consequently X-ray observations imply that the present-day accretion rate does not significantly contribute to the total disk mass. Clearly, this is an important result, since (if these galaxies are representative) it implies that the masses of spiral disks have remained almost constant over recent look-back times. Finally, we can compare the gas accretion rate with the star formation rate in NGC 2841 of $0.8M_{\odot}\text{yr}^{-1}$ inferred from $\text{H}\alpha$ emission (Young et al. 1996). Thus, although our data exclude an infall rate sufficient to build the disk over a Hubble time, they may allow enough infall to replenish the gas being used up by star formation.

Before reaching wide-ranging conclusions, however, we must ask whether it is possible to modify the emission model to bring it more into line with the observational data. Below we consider possible modifications to the halo model that might reduce the discrepancy with the observational results.

Feedback. We have chosen galaxies with particularly deep potential wells in order to minimise the uncertainties arising from supernovae feedback effects on our predictions. According to the models of Kauffmann, White & Guiderdoni (1993) and Cole et al. (1994) feedback in ~ 300 km/s halos reduces the predicted star formation rate by only $\sim 5\%$. Any gas ejected from the galaxy by feedback should begin cooling once more, and may actually increase the X-ray luminosity above our theoretical predictions. Feedback in these models cannot, therefore, reconcile theory and observations. However, in other models of galaxy formation, such as that of Nulsen & Fabian (1995, 1997) and Wu, Nulsen & Fabian (1999), in which feedback in earlier generations of dark matter halos alters the gas density profile (thereby reducing the cooling rate), the expected X-ray luminosities of spiral halos are likely to be substantially less than in the models we have tested. Although these models do reduce the luminosity of X-ray halos to levels consistent with our data, the cooling rate at late times appears much too small to allow the growth of disks as large and as massive as those observed. Further work is needed to investigate this point further.

Cosmological Parameters. The predicted X-ray luminosities have a relatively weak dependence on the values of the cosmological parameters (here we adopted $H_0 = 50$ km s $^{-1}$ Mpc $^{-1}$, $\Omega_0 = 1$, $\Lambda_0 = 0$, $\Omega_b = 0.06$ and $\sigma_8 = 0.67$). The strongest dependence is on the baryon fraction since the cooling rate is proportional to the square of the density of the X-ray emitting plasma. Using $\Omega_b = 0.04$ (a 50% reduction) reduces the luminosities by a factor ~ 2 , but at the expense of lowering the mass cooling rate also by a factor ~ 2 . Reasonable changes in cosmological parameters are unable to lower the predicted X-ray luminosity by more than a factor ~ 3 . In particular, lowering Ω_0 tends to increase the cooling rate (and consequently the X-ray luminosity) because the baryon fraction of the halo is then increased.

X-ray Emission Temperature. An assumption of the model is that the X-ray emission occurs at a characteristic temperature close to that of the galactic halo. If the plasma had a multiphase structure, then the emission weighted temperature could be lower. However, a 25% decrease in temperature reduces the expected flux by only 20%. Much larger changes in temperature are required in order to account for the discrepancy.

Spatial Distribution of X-ray Emission. Because the cooling radius estimated for NGC 2841 is so large, the correction for the spatial distribution of the X-ray emis-

sion is important. In our standard model, approximately 60% of the total luminosity is emitted within the 5–18 arcminute annulus which we observe. If, for example, we considered a very different spatial distribution for the luminosity (whilst keeping the total luminosity the same), such as one in which the X-ray emissivity remains constant within the cooling radius, then the fraction of the total luminosity emitted within the annulus would drop to around 30%. The luminosity (corrected for geometric effects) listed in Table 6.5 would therefore be reduced by a factor of 2. These corrections are much less important for NGC 4594 where the cooling radius is better matched to the PSPC field of view.

Halo Circular Velocity. The luminosity of the flow is based on the premise that the high circular velocities of the disks of these galaxies reflect the circular velocities of the halo that confines the hot plasma. Recently, several authors (Zaritsky et al. 1997, N00) have suggested that high circular velocity galaxies could have formed in significantly cooler halos. N00 has analyzed the rotation curve of NGC 2841, and found a best fitting model with $V_{200} = 167_{-13}^{+134}$ km s⁻¹ (where V_{200} is the circular velocity at the virial radius). Using the corresponding potential in the calculation of the X-ray luminosity we predict a substantially reduced X-ray emission of 0.9×10^{41} ergs s⁻¹ (including geometric factors), in much better agreement with the observational results. The ratio of the predicted value to the observational upper limit is 0.62. Unfortunately such a model underpredicts the disk mass inferred by N00 by a factor of 4. Increasing the gas fraction to remedy this problem restores the predicted X-ray luminosity to an unacceptably high value.

Metal Abundance. We have assumed that the hot gas in the halos has a metal abundance similar to that measured in galaxy clusters, namely $Z = 0.3Z_{\odot}$. Using $Z = 0.5Z_{\odot}$ increases the predicted X-ray luminosity by approximately 30%, since gas with a higher metal content cools more efficiently. Reducing the abundance to $0.1Z_{\odot}$ would, for the same reason, reduce the predicted luminosities (and gas accretion rates), by a factor of ≈ 3 , but would also reduce disk masses by the same factor.

Remaining Gas Fraction. The predicted luminosity also depends upon the value of f_g , the fraction of the baryonic material in the halo that is in gaseous form at the time when the halo forms. This parameter is unlikely to be much less than the value of unity that we have assumed unless a significant fraction of the baryons in galaxies

are in some dark form. Baugh et al. (1998b) find that $f_g > 0.7$ for halos of circular velocity comparable to those of the galaxies considered here. If f_g were lowered enough so that the predicted X-ray luminosity agreed with the observed upper limits, then the gas cooling rate would be significantly lower, making it impossible to build up a disk of the observed mass by the present. Therefore, although we do not know the exact value of f_g , a lower value cannot reconcile the model with both the observed X-ray luminosity and the inferred disk mass.

Multiphase Flow. We have assumed that gas flows into the galaxy at the centre of the halo at a fixed temperature and only cools once it has settled in the galaxy disk. However, it is conceivable that the gas could instead cool “in-situ” (i.e. at the location in the dark matter halo where it originated) and then fall into the galaxy as cold clouds (Thomas, Nulsen & Fabian 1987). If this were indeed the case, then the bolometric X-ray luminosity emitted by the cooling gas would be $2.5\dot{M}_{\text{cool}}V_C^2$ (assuming that the gas cools at constant pressure). For typical values of r_{cool} and r_{optical} this is about equal to the luminosity predicted by eqn. (6.4). Furthermore, this luminosity would be emitted at a lower mean temperature than the virial temperature of the halo. Using the cooling flow model of Mushotzky & Szymkowiak (1988) we estimate that, for gas cooling from 0.2 keV, 11% of the bolometric luminosity would be emitted in the 0.2-2.0 keV band after accounting for absorption. This is approximately half of the 24% that is emitted in this band for gas at a fixed temperature of 0.2 keV. Thus, in the multiphase model the predicted flux in the 0.2-2.0 keV band is lower by a factor of approximately 2 compared to our standard model. These models are thus less discrepant with the observational upper limits.

It should be noted, however, that even in this scenario the infalling gas must still release both its thermal and gravitational energy somewhere. The discussion above accounts only for the thermal component, and this model only succeeds if the gravitational component is radiated outside the ROSAT energy band. This not readily accomplished: if the cold clouds fall ballistically towards the centre, the gravitational energy would be radiated as post-shock X-rays when they collide with the galaxy’s gas disk. It does not help if the cold clouds are coupled to the hot gas by magnetic fields since although they could flow in more slowly by transferring their gravitational energy to the hot component, this energy will ultimately be radiated

in the X-ray band.

Support for a multiphase model may be provided by the high-velocity clouds recently detected by Blitz et al. (1999) if these are indeed infalling onto the Milky Way. However, Blitz et al. estimate that the present day infall rate of such clouds onto the Milky Way is around $0.8M_{\odot} \text{ yr}^{-1}$, and this is significantly lower than the rate predicted by the models we consider ($\sim 5\text{--}6 M_{\odot} \text{ yr}^{-1}$). Thus, unless the observed clouds represent only a small fraction of the total cloud population, the simple multiphase model that we have considered seems to conflict with the inferred infall rate of cold gas onto the Milky Way.

6.4 Conclusions

We have searched for large-scale diffuse X-ray emission from three nearby, high rotation velocity spiral galaxies. In each case, we find that the observed emission is almost *an order of magnitude* below that required to build the disk at a constant rate within the age of the universe. The accretion rate is similarly discordant with theoretical estimates of the cooling rate of gas in an isothermal or NFW-profile halo.

Although extremely suggestive, our results are not definitive. The reasons for this are:

- Strong constraints are derived only for the two nearer galaxies. We may, by chance, have selected two systems that have anomalous cooling rates. This is plausible for the bulge dominated NGC 4594 galaxy (‘the Sombrero’). NGC 2841 appears to be more typical of spiral galaxies, but its proximity makes accurate background subtraction difficult, and our results are dependent on the profile assumed for the X-ray emission. By contrast, the third galaxy meeting our selection criteria is too distant for strong limits on the X-ray flux to be set. We note, however, that the X-ray limits established for these galaxies are compatible with the emission that is thought to occur in the halo of our own galaxy (e.g. Moore & Davis 1994, Wolfire et al. 1995).
- There are a number of parameters that have been assumed in order to predict the X-ray flux from the properties of the halo. No parameter, or combination of parameters, can simultaneously give a low X-ray luminosity and build the observed disk. To reconcile the predictions with the observed luminosities we must assume that either: (a) disks were built at high redshift; or, (b) accretion onto disks occurs

Table 6.5: Predictions from our models, assuming the standard cosmology and halo gas with $Z = 0.3Z_{\odot}$. Columns 3 and 4 list the total model luminosity and cooling radius in arcmin; column 5 gives the model luminosity after correction for the geometry of the source and background annuli; column 6 lists the ratio of the observed luminosity to the model luminosity after correction for the geometry.

Target	Halo model	Model $L_X/10^{41}$ ergs/s	R_{cool} (arcmin)	Geo. corrected Model $L_X/10^{41}$ ergs/s	Ratio of obs. 95% upper limit to model
NGC2841	Isothermal	10.4 ± 0.10	35.1 ± 13.7	5.5 ± 0.06	< 0.07
	NFW	7.6 ± 0.74	48.0 ± 18.1	4.5 ± 0.44	< 0.09
NGC4594	Isothermal	10.8 ± 0.18	18.7 ± 1.9	7.6 ± 0.13	< 0.16
	NFW	9.5 ± 0.63	23.0 ± 2.4	6.9 ± 0.46	< 0.17
NGC5529	Isothermal	3.4 ± 0.03	9.4 ± 0.4	1.3 ± 0.01	< 2.9
	NFW	2.4 ± 0.10	12.5 ± 0.5	0.9 ± 0.04	< 4.2

without radiating the binding energy of the gas at X-ray wavelengths (e.g. in a multiphase flow).

Our results encourage further work. The current upper limits are set by our ability to subtract the background count-rate in the images. This is exacerbated in the ROSAT data because the flux level at which individual point sources may be subtracted varies across the image on scales comparable to the spiral galaxy halos. Future work requires large-scale imaging with a detector that has a near-uniform detection threshold, such as the EPIC camera aboard the XMM-Newton satellite (Lumb et al. 1996). Such a survey would need to target a greater fraction of the nearby galaxies, setting firmer limits on the diffuse X-ray emission and eliminating variations in galaxy properties as a major source of uncertainty. If such work continues to fail to make detections, then we may be forced to conclude that spiral disks are no longer growing at the present epoch. This would agree with traditional models of galaxy formation in which galaxies accrete their mass at early times and evolve as more or less closed systems to the present-day. Clearly, there is much to learn from future studies of the X-ray halos of spiral galaxies.

Chapter 7

A Comparison of Semi-Analytic and Smoothed Particle Hydrodynamics Galaxy Formation

7.1 Introduction

The properties of galaxies in the Universe are determined by the behaviour of both the dark matter and the baryonic material from which they are made. The dynamics of the dark matter, which are determined by gravity alone (unless “self-interacting”, or collisional dark matter is considered, see, for example, Moore et al. 2000), are now reasonably well understood. N-body simulations (e.g. Davis et al. 1985) provide an accurate description of the evolution of structure into the highly non-linear regime where dark matter halos form (see, for example, Jenkins et al. 1998, Colín et al. 1999). Analytically, the Press-Schechter theory (Press & Schechter 1974) predicts to within $\sim 50\%$ the distribution of halo masses found in N-body simulations for a specified cosmology, whilst theoretically motivated fitting functions do even better (Sheth & Tormen 1999, Sheth, Mo & Tormen 1999, Jenkins et al. 2000). Extensions of this theory (Bond et al. 1991, Bower 1991, Lacey & Cole 1993) predict, with reasonable accuracy, the hierarchical build-up of halos through the mergers of smaller progenitors (see, for example, Lacey & Cole 1994, Somerville et al. 1998). The behaviour of the baryonic matter, on the other hand, is less well understood. The dynamics of the gas are not determined by gravity alone but also by hydrodynamical forces and radiative processes. Since gas must cool into dense lumps before it can turn into stars, these processes are crucial for galaxy formation (see, for

example, Binney 1977, Rees & Ostriker 1977, Silk 1977, White & Rees 1978).

In this Chapter, we compare two widely used techniques for modelling the behaviour of gas as it forms into galaxies: semi-analytic modelling and direct simulation using smoothed particle hydrodynamics (SPH). Semi-analytic models applied to cold dark matter (CDM) cosmologies (Cole 1991, White & Frenk 1991, Lacey & Silk 1991, Kauffmann, White & Guiderdoni 1993, Cole et al. 1994, Somerville & Primack 1999, Cole et al. 2000, Chapter 2) have met with considerable success in explaining many of the observed properties of the galaxy population, such as the luminosity function, the distributions of colour and morphological type, the counts as a function of magnitude and redshift (e.g. White & Frenk 1991, Lacey et al. 1993, Kauffmann, Guiderdoni & White 1994, Cole et al. 1994, Kauffmann 1995, Baugh, Cole & Frenk 1996a,b, Kauffmann 1996a,b, Kauffmann & Charlot 1998a,b, Chapter 2), the properties of Lyman-break galaxies (Baugh et al. 1998b, Governato et al. 1998), and the clustering of galaxies (Kauffmann, Nusser & Steinmetz 1997, Kauffmann et al. 1999a,b, Diaferio et al. 1999, Baugh et al. 1999, Chapters 3 & 4).

The SPH technique (Lucy 1977, Gingold & Monaghan 1977) has been used by many authors to model galaxy formation (e.g. Katz & Gunn 1991, Katz, Hernquist & Weinberg 1992, Navarro & White 1993, Evrard, Summers & Davis 1994, Steinmetz & Müller 1994,1995, Katz, Weinberg & Hernquist 1996, Frenk et al. 1996, Weil, Eke & Efstathiou 1998, Navarro & Steinmetz 1999, Pearce et al. 1999). These simulations have been successful in producing objects with approximately the mass of galaxies and, in cosmological simulations, with approximately the right abundance. However, to date no simulation has been able to produce a realistic, rapidly rotating spiral galaxy starting from cosmological initial conditions (e.g. Navarro & Steinmetz 1999).

Both techniques require a number of simplifying assumptions in order to model the evolution of cooling gas. For example, semi-analytic models assume that dark matter halos and their associated gas component are spherically symmetric, and that gas is efficiently shock-heated when halos collapse. SPH, on the other hand, assumes that gas is well represented by a set of discrete particles. The two methods have different strengths and limitations. Semi-analytic modelling can follow a large dynamic range of scales and is sufficiently flexible that the effects of varying assumptions and parameter values can be readily explored. SPH, on the other hand, does not impose any restrictions on geometry and solves directly the approximate evolution equations for gravitationally coupled dark matter and dissipative gas. Limited resolution, however, restricts the accessible dynamic

range and the expense of large simulations makes it impractical to carry out extensive parameter space explorations. In both approaches, a phenomenological model for star formation and feedback must be coupled to the evolution of dark matter and gas in order to calculate observable properties of galaxies. Generally, such models are more easily implemented in semi-analytic models than in SPH simulations in which the behaviour of the phenomenological model itself often depends on resolution.

The main aim of this Chapter is to determine the extent to which the two techniques of semi-analytic modelling and SPH simulation produce consistent results for the evolution of cooling gas in the cosmological setting relevant to galaxy formation. There are several ways in which such a comparison might be carried out. In this Chapter, we adopt the statistical approach of comparing the properties of *populations* rather than of individual objects. This comparison is motivated by two considerations. Firstly, we wish to understand how the different approximations inherent in the two techniques translate into differences in the *average* properties of the two models. Secondly, both techniques have been used (and continue to be used) to study statistical properties such as galaxy luminosity functions and spatial correlation functions. It is clearly important to test how reproducible these bulk properties are using these rather different modelling techniques. A complementary approach, well suited to future study, is to compare the properties of individual objects modelled using the two techniques. A secondary aim of this Chapter is to assess how the neglect of sub-resolution processes in SPH simulations, i.e. star formation and feedback, can affect the properties of objects above the resolution limit.

When differences between the results of the two modelling techniques do arise, it is difficult to know which of the two, if either, is giving the “correct” answer. In some cases, however, we can explore the reasons for a disagreement by altering parameters in the semi-analytic model which describe a specific physical process, for example, the galaxy merger timescale. In this way, we can identify specific areas of disagreement in which further theoretical work is required.

This Chapter is laid out as follows. In §7.2, we briefly describe the SPH and semi-analytic models, and discuss in greater detail their specific implementation in this work. In §7.3, we compare several properties of the galaxies calculated using the two techniques, and uncover the reasons for some of the differences by varying key parameters in the semi-analytic models. Finally, in §7.4, we present our conclusions.

7.2 The SPH Simulations and Semi-Analytic Models

We now present the models of galaxy formation employed in this Chapter. The semi-analytic model has been described in detail in Chapter 2. Since the SPH technique is described in detail elsewhere, we give only a brief overview here, referring the reader to the appropriate references for further details where applicable.

7.2.1 SPH Simulations

Techniques

The SPH technique is a Lagrangian method in which the gaseous component of the universe is described by a set of tracer gas elements represented by particles within the simulation volume. Estimates of local gas properties (and their spatial derivatives) for each particle are derived by smoothing over the properties of the N_{SPH} nearest neighbour particles (see Monaghan (1992) for a review). For simulations of galaxy formation, the gas must also be able to cool radiatively.

The simulation volume is initially populated with dark matter and gas particles, with a spatial distribution derived from a cosmological power spectrum. The equations of gravity and hydrodynamics are then solved over a succession of small timesteps in order to propagate the particle distribution forwards in time until the present day.

Simulation Specifics

Simulations were carried out for two of the cold dark matter cosmological models studied by Jenkins et al. (1998): Λ CDM (mean mass density parameter, $\Omega_0 = 0.3$; cosmological constant, $\Lambda_0 = 0.7$; Hubble constant, $h = 0.7$; and rms linear fluctuation amplitude in $8h^{-1}\text{Mpc}$ spheres, $\sigma_8 = 0.9$), and SCDM ($\Omega_0 = 1.0$, $\Lambda_0 = 0$, $h = 0.5$, $\sigma_8 = 0.6$). The baryon fraction in each cosmology was set, from nucleosynthesis constraints, to be $\Omega_b h^2 = 0.015$ (Copi, Schramm & Turner 1995). The simulations, which were carried out using the parallel AP3M-SPH code of Pearce & Couchman (1997), have 128^3 particles of each species in boxes of side $70h^{-1}\text{Mpc}$ and $50h^{-1}\text{Mpc}$ in the Λ CDM and SCDM models respectively. The gas mass per particle is therefore 1.4×10^9 and $1.0 \times 10^9 h^{-1} M_\odot$ in the Λ CDM and SCDM cosmologies respectively. Since we adopt $N_{\text{SPH}} = 32$, the smallest resolved objects have a gas mass of 4.5×10^{10} and $3.2 \times 10^{10} h^{-1} M_\odot$ in the two cosmologies. An unevolving gas metallicity of 0.3 times the Solar value was assumed, taking into account that gas cooling in objects above the resolution threshold will already have been

processed by several previous levels of the merger hierarchy. The SPH simulations use a cooling function which is a series of power-law fits to the results of Raymond, Cox & Smith (1976). If instead the tabulated cooling function of Sutherland & Dopita (1993) is used (this being the function used in the semi-analytic models) then around 10% more cold gas results by $z = 0$ (Kay et al. 1999).

We employed a comoving β -spline gravitational softening equivalent to a Plummer softening length of $35h^{-1}\text{kpc}$ for $z > 2.5$ in ΛCDM and $25h^{-1}\text{kpc}$ for $z > 1.5$ in SCDM. At lower redshifts, the softening remained fixed at $10h^{-1}\text{kpc}$ in physical coordinates, and the minimum SPH spatial resolution was also set to match this value. Approximately 10,000 timesteps were used in each simulation to evolve from $z = 50$ to $z = 0$. With our chosen parameters, the simulations are able to follow the cooling of gas into galactic dark matter halos. The resulting “galaxies” typically have 50-1000 particles.

Dark matter halos were identified at $z = 0, 0.1, 0.2, 0.3, 0.5, 1, 2$ and 3 . Simulation outputs at higher redshifts were available, but so few galaxies have formed beyond $z = 3$ in these small volumes that no statistically meaningful comparison could be made at higher redshifts. Halos were found using the friends-of-friends (FOF) group finding algorithm (Davis et al. 1985) with a linking length of $b = 0.2$ times the mean interparticle separation in SCDM. In ΛCDM , we assume halos form with the overdensity predicted by the model of Eke, Cole & Frenk (1996) and therefore use a linking length which grows from $b = 0.164$ at $z = 0.0$ to $b = 0.200$ at $z = 3$. Consistent definitions of halo overdensities are adopted in the semi-analytic models considered.

“Galaxies” were also found using this algorithm, but with a much smaller linking length of $b = 0.0168$. Galaxies are made of gas particles that have cooled below 12,000 K and represent overdensities of $\gtrsim 100,000$. A complete description of the simulations and the properties of the simulated galaxies is given by Pearce et al. (2000).

Assumptions and Limitations

The key assumption of the SPH technique is that the evolution of gas may be approximated by the evolution of a set of particles. Each particle may be thought of as a packet of gas that “carries” with it the thermodynamical properties of the system.

The smoothing inherent in the SPH technique introduces problems whenever gas properties vary discontinuously (or at least on scales much smaller than the smoothing scale). In the case of shocks, an artificial viscosity term is used to capture the shock and prevent it from being smoothed away by the SPH algorithm. Another example of this kind of

problem occurs in a multiphase gas (although see Ritchie & Thomas 2000), in which the sharp boundary between phases is smoothed over, causing the phases to diffuse into one another. This problem can lead to runaway cooling in the centres of dark matter halos, as happened, for example, in one of the simulations of Frenk et al. (1996). The simulations of Pearce et al. (1999) attempt to circumvent this problem by ignoring the contribution of cold ($T < 12,000\text{K}$) particles in the computation of the densities of hot ($T > 10^5\text{K}$) particles. (For a complete discussion of this approximation see Thacker et al. 1999) An important consequence of this approximation is that galaxies in the simulations match the shape of the observed galaxy luminosity function, a constraint used to fix several of the physical parameters in the semi-analytic models.

A further limitation arises from the fact that the sizes of the galaxies that form in the simulations are determined primarily by the gravitational force softening length rather than by any real physical process. This raises the possibility of enhanced tidal disruption, drag, and merging within dark matter halos. Whilst the softening length is kept fixed in physical coordinates at low redshift, it is fixed in comoving coordinates at high redshift, as described in §7.2.1. Thus for $z > 1.5$ the *physical* softening length in the ΛCDM simulation is larger than in the SCDM simulation and, as a result, non-physical effects due to softening may be expected to be more pronounced at early times in the ΛCDM simulation.

Although a variety of prescriptions have been tried in attempts to model supernovae feedback in SPH simulations, usually by converting cold gas into “star particles” which then inject thermal and kinetic energy into the surrounding gas (e.g. Navarro & White 1993, Steinmetz & Müller 1995, Katz, Weinberg & Hernquist 1996), this process remains poorly understood. In cosmological SPH simulations, gas can only begin to cool efficiently in objects well above the minimum resolved halo mass, around several times $10^{11}h^{-1}M_{\odot}$ in our case. Thus, resolution effects prevent all the gas from cooling in small halos at high redshift, a process that, in reality, is probably due to feedback from supernovae or other energetic sources. The resolution of our simulation was, in fact, chosen to ensure that the fraction of baryons that cools by the present in the SCDM model is comparable to the observed fraction of cold gas and stars in galaxies today. This was achieved by carrying out several test simulations with varying resolution until the desired cold gas fraction was obtained (Kay et al. 1999).

7.2.2 Semi-Analytic Models of Galaxy Formation

Techniques

We briefly review aspects of the semi-analytic model relevant to the present Chapter. A full description of the model was given in Chapter 2.

In semi-analytic models, some of the processes involved in galaxy formation (e.g. the growth of dark matter halos by mergers of smaller halos) are followed using analytic solutions and Monte-Carlo techniques. Other, more uncertain processes, such as feedback from supernovae are modelled by means of simple, physically motivated rules. Typically, each such rule contains one or two free parameters which are constrained using observations of galaxies in the local Universe.

In semi-analytic models, the dynamics of the gas are strongly coupled to the evolution of dark matter halos and to the processes of star formation and feedback. Recall that the starting point for our own modelling is the set of dark matter halos at $z = 0$ drawn from the Press-Schechter mass function. A merging history for each halo is then constructed using the extended Press-Schechter formalism. Beginning with the earliest progenitor halo, our model assumes that gas (initially assumed to have zero metallicity) is shock-heated to the virial temperature of the halo, after which it begins to cool according to a specified cooling function. Any gas that does cool forms a galaxy at the centre of the halo within which stars begin to form at a specified rate, producing both metals and supernovae. Supernovae reheat some of the gas in the galaxy, ejecting it back out into the surrounding halo. (This gas is not allowed to cool again until the halo has merged to form part of a larger halo.) These processes continue until the halo mass has increased by a factor of two or more, either by merging with a larger halo, or by numerous accretions of smaller halos. Any leftover hot gas becomes part of the new halo, and the largest galaxy of the newly formed halo becomes the central galaxy, onto which further gas can cool. Any other galaxies become satellites in the new halo, and may eventually merge with the central galaxy due to energy loss by dynamical friction. The full model includes other processes such as stellar population synthesis and morphological evolution which are not directly relevant to the present work.

As well as this “full” semi-analytic (FSA) model, for the purposes of this work we have constructed a “stripped-down” semi-analytic (SDSA) model which is designed to be directly comparable to the SPH simulations (in a statistical sense). In the SDSA model, we switch off star formation and the associated supernovae feedback and chemical

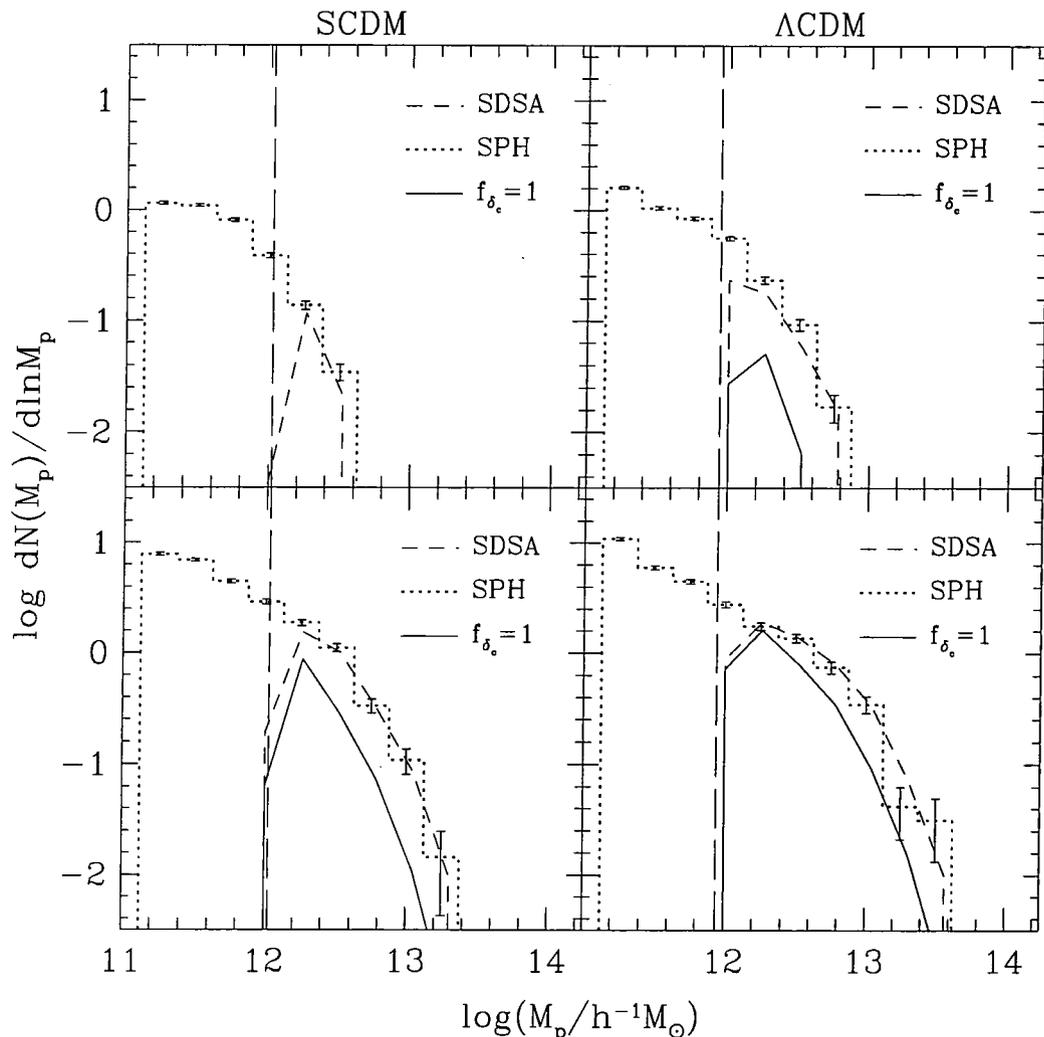


Figure 7.1: The mass function of dark matter halos at $z = 2$ which are progenitors of present-day halos of mass 10^{12} – $10^{13}h^{-1}M_{\odot}$ (upper panels) and 10^{13} – $10^{14}h^{-1}M_{\odot}$ (lower panels) in the SCDM (left-hand panels) and Λ CDM (right-hand panels) cosmologies. The dotted histograms are the mass functions in the SPH simulations. The thin solid lines show the mass functions obtained from the SDSA models using the standard definition of the extrapolated critical linear overdensity for collapse from the spherical collapse model (i.e. with the parameter $f_{\delta_c} = 1$), whilst the dashed lines show the SDSA model result when f_{δ_c} is chosen to give a good match to the progenitor mass function in the SPH simulations (c.f. eqn. 7.1). Vertical dashed lines show the resolution limit imposed on the SDSA model merger trees.

enrichment, since these processes are not included in the SPH calculation. Instead, we assume a fixed metallicity of 0.3 times the Solar value, just as in the simulations. We also mimic the SPH resolution by truncating halo merger trees at N'_{SPH} times the dark matter particle mass in the simulation, and by switching off gas cooling when the hot gas mass is less than N'_{SPH} times the gas particle mass. The parameter N'_{SPH} is set to $2 \times N_{\text{SPH}} = 64$, i.e. twice the number of particles in the SPH smoothing kernel. This value was chosen since it allows the SDSA model to match the position of the peak in the galaxy mass function in the SPH simulations (as will be shown in §7.3.2, see Fig. 7.7). The FSA model has no such truncation of merger trees and has effectively unlimited resolution (in practice, we resolve progenitor halos down to masses several hundred times smaller than in the SPH and SDSA models).

To further emulate conditions in the SPH simulations in the SDSA model, we replace the Press-Schechter formula for the mass function of dark matter halos (which is commonly used in semi-analytic models, including our FSA model) with the formula proposed by Sheth, Mo & Tormen (1999) which provides a better match to the results of large N-body simulations (Jenkins et al. 2000). Although this ensures that the abundance of halos at $z = 0$ in the SDSA model is similar to that in the SPH simulations, it does not guarantee that the distribution of progenitor halo masses will also be the same. In fact, as has been shown by Somerville et al. (1998), we find that at high redshifts the merger trees in our SDSA model underpredict the abundance of progenitor halos seen in the SPH simulations. We reconcile the progenitor distributions in the two models by adopting a simple, empirical correction. We scale the value of the extrapolated critical linear threshold for collapse from the spherical collapse model, δ_c , by a redshift-independent factor, f_{δ_c} , such that $\delta_c^{\text{eff}} = f_{\delta_c} \delta_c$. (Thus, in the standard form of the extended Press-Schechter theory, $f_{\delta_c} = 1$.) This approach is similar to that proposed by Tormen (1998). The factor f_{δ_c} is allowed to be a function of the $z = 0$ halo mass, M_{halo} . We find that the following simple form for f_{δ_c} provides a reasonable fit to progenitor mass functions in the SPH simulations over the redshift range 0 to 3:

$$f_{\delta_c} = \begin{cases} 1 + 0.09 [\log(M_{\text{halo}}/h^{-1}M_{\odot}) - 16.56], & \text{SCDM} \\ 1 + 0.14 [\log(M_{\text{halo}}/h^{-1}M_{\odot}) - 15.64], & \Lambda\text{CDM}. \end{cases} \quad (7.1)$$

This correction to the extended Press-Schechter theory is based purely on an empirical fit to our two simulations. It is designed to minimize differences in the statistical properties of the dark halos in the two models, so that we can focus on differences in their gas-dynamical properties. Given the limited statistics provided by our relatively small

simulation volumes, this correction should not be regarded as appropriate in a general sense and may not be accurate for larger volumes or for different power spectra to those considered here. In Fig. 7.1 we show the mass function of progenitor halos at $z = 2$ for parent halos in the mass ranges 10^{12} – $10^{13}h^{-1}M_{\odot}$ and 10^{13} – $10^{14}h^{-1}M_{\odot}$, defined so that $(dN(M_P)/d \ln M_P)d \ln M_P$ is the number of progenitors in the mass range $\ln M_P$ to $\ln M_P + d \ln M_P$ per parent halo. The dashed lines indicate the distribution from our SDSA model with the above correction, whilst the solid lines show the model without the correction. The correction succeeds in ensuring that the mass functions of progenitor halos are statistically similar in the SDSA and SPH models.

Specifics

The semi-analytic model of Chapter 2 was used to simulate galaxy formation in a large sample of dark matter halos spanning a wide range in mass. Specifically, for the SDSA model we simulated halos in 28 mass bins spaced uniformly in the logarithm of halo mass between 10^{11} and $10^{15}h^{-1}M_{\odot}$. For each mass bin, 100 halos were simulated. In the FSA model the mass range was extended down to $10^{10}h^{-1}M_{\odot}$, and between 4 and 20 halos were simulated in each mass bin (fewer halos were simulated for the most massive bins as these are computationally more expensive). To study the clustering of galaxies in the semi-analytic models, we used the techniques of Chapter 3 and populated halos in the SPH simulations with “semi-analytic” galaxies.

The cosmological parameters required as input into the semi-analytic model (Ω_0 , Λ_0 , h , σ_8 , Ω_b) were set to the same values used in the SPH simulations. The remaining parameters of the FSA model were set equal to the values given in Table 2.1 (for SCDM we used the parameters of the τ CDM model). Parameters were chosen so as to obtain a model which produces a reasonable match to the local B and K-band luminosity functions and other local data (as described in Chapter 2). These parameters can be split into two classes: those that affect the results of the SDSA model and those that do not. The latter, however, are still important for specifying the behaviour of the FSA model. Parameters that do not affect the SDSA model are those which govern star formation, feedback from supernovae and the production of metals by stars. We do not discuss them in any detail here but refer the reader to Chapter 2 for a full description.

The parameters which do affect the results of the SDSA model (and whose specific values tend to be inspired by the results of simulations) are the following: (i) f_{df} , the dynamical friction coefficient that determines the merger timescale for galaxies orbiting

in halos (see eqn. 2.35); (ii) the dark matter density profile; (iii) the gas density profile; (iv) the progenitor halo mass resolution; and (v) the critical mass for cooling ((iv) and (v) are both specified by N'_{SPH}). Unless otherwise stated, we set $f_{\text{df}} = 1$, $N'_{\text{SPH}} = 64$, and assume (a) that the dark matter density profile has the form proposed by Navarro, Frenk & White (1996,1997, hereafter NFW), namely,

$$\rho(r) \propto \frac{1}{r/r_s(1 + [r/r_s]^2)}, \quad (7.2)$$

where r_s is a scale-length, and (b) that the gas density has an isothermal profile at large radii, and a constant density core of size r_c , i.e.

$$\rho(r) \propto \frac{1}{r^2 + r_c^2}. \quad (7.3)$$

The core radius is initially set to some fraction of the NFW scale-length, r_s , of the dark matter halo. Our standard choice for this fraction is 0.33, motivated by the results of hydrodynamical simulations of cluster formation (Navarro, Frenk & White 1995, Eke, Navarro & Frenk 1998). As explained in Chapter 2, the gas that is able to cool in a halo is the densest gas, which has the lowest entropy. When halos merge to form a new halo this low entropy gas, which would normally settle into the inner parts of the halo, is missing. We take this into account by increasing the core radius of later generations of halos so that the gas density at the virial radius is the same as it would have been if no gas had cooled in progenitors. The model also allows us the option of keeping the core radius fixed, which we explore below. The inclusion of a core in the hot gas profile prevents the formation of extremely bright galaxies in the centres of groups and clusters, which would otherwise lead to a disagreement with the shape of the bright end of the observed galaxy luminosity function. The reasons for choosing this particular profile are therefore identical in spirit to those for preventing runaway cooling in the SPH simulations (see §7.2.1).

Assumptions and Limitations

Semi-analytic models make several assumptions in the treatment of gas in order to obtain simple, analytic solutions to complex hydrodynamical processes. We have already mentioned the important assumptions of spherical symmetry and of the shock-heating of the gas to the virial temperature of its associated halo. The hot gas is then further assumed to settle into a distribution with a universal form. Finally, the amount of gas that is able to cool by time t after the formation of the halo is identified with the gas contained within

the radius at which the cooling time equals t . Once it has cooled, this gas is assumed to flow to the centre of the halo, where it is available for star formation, provided that the free-fall time for the gas is also less than t . We shall refer to this as the “cooling radius” prescription.

7.3 Comparison of the Two Models

In this section we compare several properties of the galaxy populations that form in our models and consider how this comparison is affected by varying certain assumptions and parameter values.

7.3.1 Properties of Halo Gas

We begin by comparing the most basic quantities calculated by each technique, namely the fraction of gas in the hot and cold phases, both globally and as a function of dark matter halo mass. For these purposes, we define a ‘hot halo gas phase’ as gas hotter than 10^5K ; a ‘galaxy phase’ represented by cool, dense gas in the SPH simulation and SDSA, and also including stars in disks and spheroids in the FSA; and an ‘uncollapsed gas phase’ consisting of everything else — i.e. gas outside hot, virialised halos. Note that for the galaxy phase, we consider only galaxies with a mass greater than N'_{SPH} gas particles in the SPH and SDSA models, but include galaxies of all masses in the FSA model.

In the Press-Schechter (or Sheth-Mo-Tormen) theory, all the matter in the universe is deemed to be in halos of some mass, and semi-analytic models assume that gas in halos is shock heated to the halo virial temperature. We can therefore determine the fraction of gas in the uncollapsed gas phase in the FSA model simply by integrating over the analytical mass function (Sheth, Mo & Tormen 1999) from zero mass to the mass corresponding to a virial temperature of 10^5K . According to the spherical top-hat model of halo formation, the mass corresponding to 10^5K is:

$$M_{10^5\text{K}} = 3.5 \times 10^{10} (1+z)^{-3/2} \Omega_0^{-1/2} \left(\frac{200}{\Delta_c(z)} \right)^{1/2} h^{-1} M_\odot, \quad (7.4)$$

where $\Delta_c(z)$ is the overdensity of a newly formed, virialised dark matter halo at redshift z (e.g. Eke, Cole & Frenk 1996). Since some halos hotter than 10^5K are not resolved in the SDSA model, the integration in this case is carried out from zero mass to $M_{10^5\text{K}}$ or to N'_{SPH} dark matter particle masses, whichever is largest. This estimate does not correspond exactly to the situation in the SPH simulation in which the largest halos are

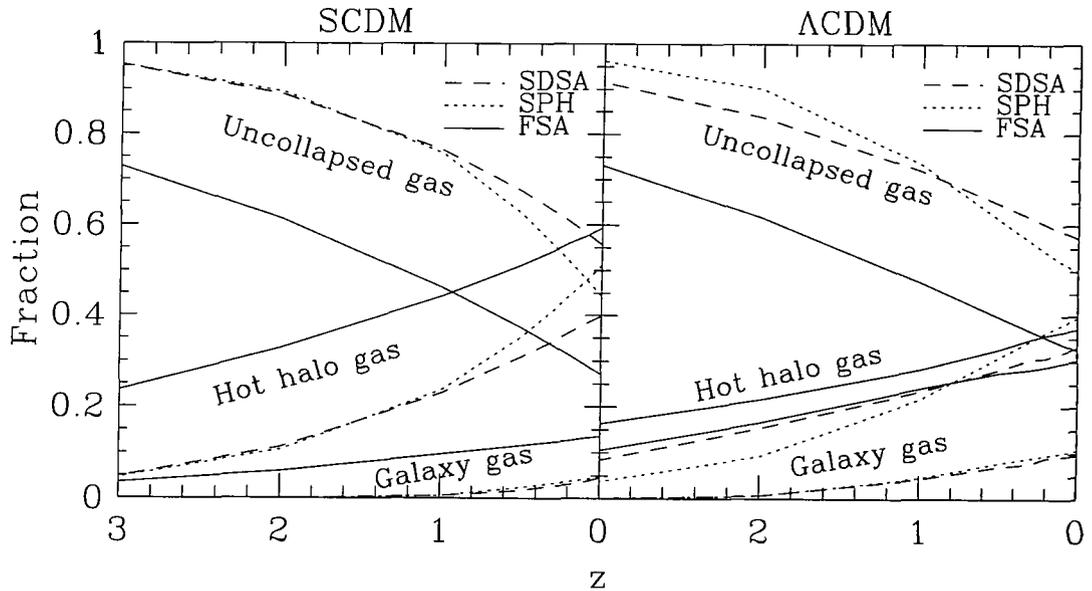


Figure 7.2: The global fraction of gas in each of three phases (reading lines from top to bottom): uncollapsed gas, hot halo gas and galaxy gas. Dotted lines correspond to the SPH simulation, dashed lines to the SDSA model, and solid lines to the FSA model.

surrounded by gas at temperatures above 10^5 K which extends beyond the virial radius. Because of this, the SDSA model calculation will effectively overestimate the amount of uncollapsed gas relative to the SPH simulation. On the other hand, gas in the SPH simulation tends to be slightly more extended than assumed in the semi-analytic model (i.e. the simulated clusters tend to have a baryonic content slightly smaller than the universal baryon fraction within a radius enclosing an overdensity of 200 – see e.g. Frenk et al. 1999). These two effects counteract each other to some degree.

The amount of gas in the hot and galaxy phases depends upon the rate at which gas cools. Therefore this comparison tests model assumptions relating to the process of gas cooling, such as spherical symmetry and the cooling radius prescription in the semi-analytic models or the effects of smoothing in SPH. This test will therefore be sensitive to the choice of gas density profile in the semi-analytic models and to N_{SPH} in the simulations. Since this comparison is concerned only with the total amount of gas in different phases, it is insensitive to the way in which the gas is apportioned into galaxies within a single halo, at least in the SPH and SDSA models. In the FSA model, some dependence on galaxy merger rates may exist, since merging can affect the star formation rate in a galaxy and thus alter the amount of gas reheated by feedback, as well as the rate of chemical evolution, which in turn alters the cooling rates in subsequent generations of

halos.

Global Gas Fractions

Figure 7.2 shows the fraction of gas in each of the three phases: hot halo, galaxy and uncollapsed, as a function of redshift. In both cosmologies, the uncollapsed gas fraction in the SDSA model is quite close to, although somewhat larger (by $\lesssim 0.1$), than in the SPH simulation at low redshifts. At $z = 0$, the fractional difference is $\lesssim 30\%$. Given the caveats mentioned above, this level of agreement is pleasing. In the FSA model, the fraction of gas in the uncollapsed phase is significantly lower than in the SPH simulation and the SDSA models. The only differences between the FSA and SDSA models in this calculation is the numerical resolution and the use of the Press-Schechter mass function in the FSA model and the Sheth-Mo-Tormen in the SDSA model. It turns out that these differences contribute about equally to the discrepancy in the fraction of uncollapsed gas in the two cases at $z = 0$. (The Press-Schechter mass function contains more low temperature halos than the Sheth-Mo-Tormen mass function.) At higher redshift, resolution effects are the dominant factor. The gas belonging to sub-resolution halos is classed as uncollapsed gas in the SDSA and SPH cases, but is accounted for as hot halo or galaxy gas in the FSA case.

In the galaxy phase, which is the most interesting from the perspective of galaxy formation, the SDSA and SPH models again agree extremely well at all redshifts in both cosmologies. The fractional difference, defined as $|M_{\text{SDSA}} - M_{\text{SPH}}|/M_{\text{SPH}}$, where M_{SDSA} and M_{SPH} are the masses of galaxy gas in the SDSA and SPH models respectively, is close to 50% at $z = 2$ in the SCDM cosmology, but is much smaller in Λ CDM. Half of the $z = 0$ mass of galaxy gas has cooled by $z \approx 0.5$ and $z \approx 1.0$ in SCDM and Λ CDM respectively. Below these redshifts, the fractional difference between the SDSA and SPH models is everywhere less than 25%. The agreement between the hot halo gas in the SDSA and SPH models is also very good. For Λ CDM, at $z = 0$, the two models differ by only 20%. This is a non-trivial result. One of the most uncertain assumptions of the semi-analytic technique is that gas is shock-heated to the virial temperature of the parent halo. It is therefore reassuring that the amount of hot halo gas turns out to be similar to that seen in the SPH simulations.

The FSA model predicts significantly more galaxy gas than the SPH and SDSA models. Since the merger trees in this model have effectively no mass resolution limit, gas cools very efficiently at high redshifts in small halos. However, feedback reheats a signifi-

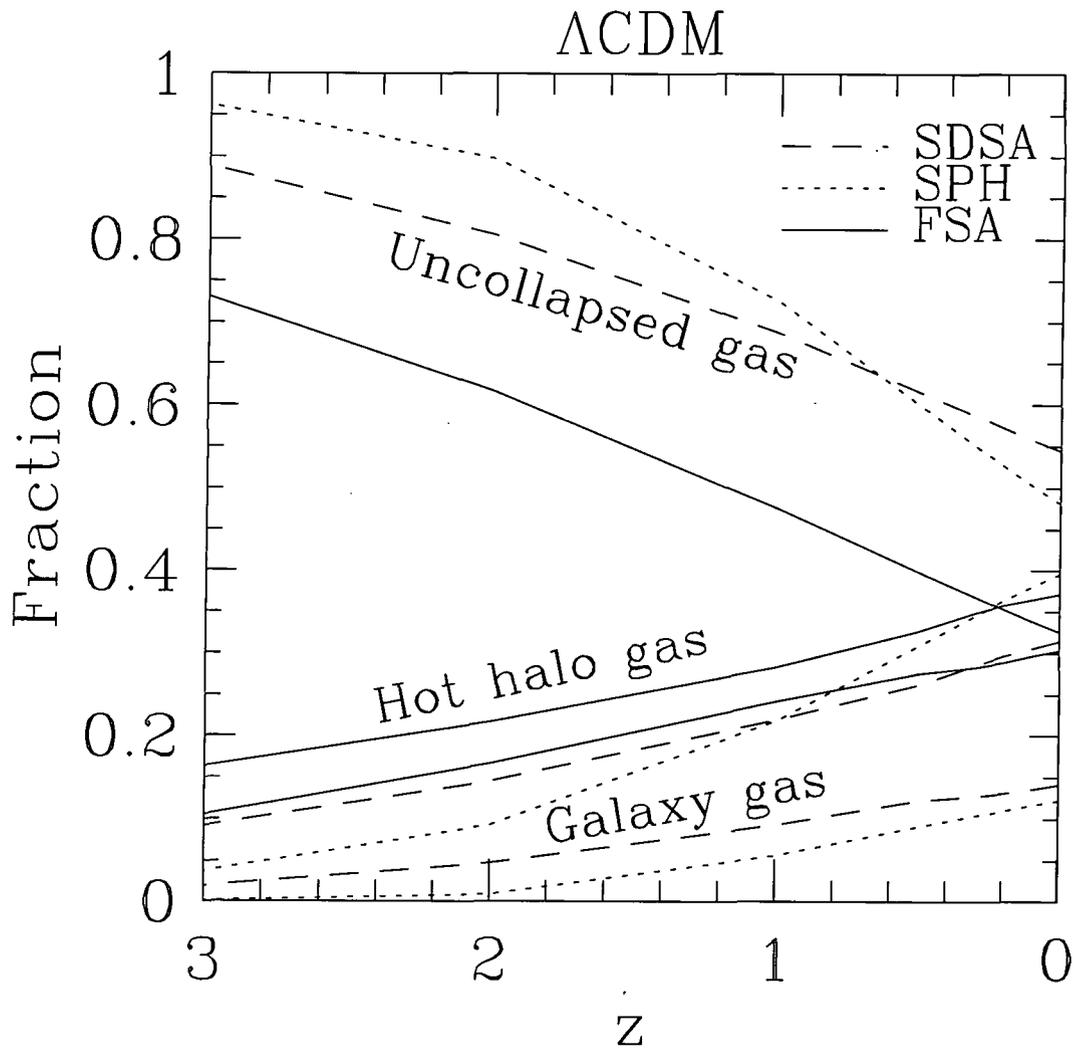


Figure 7.3: The global fraction of gas in each of three phases (reading lines from top to bottom): uncollapsed gas, hot halo gas and galaxy gas. Results are shown for the Λ CDM cosmology with $N'_{\text{SPH}} = 32$. Dotted lines correspond to the SPH simulation, dashed lines to the SDSA model, and solid lines to the FSA model.

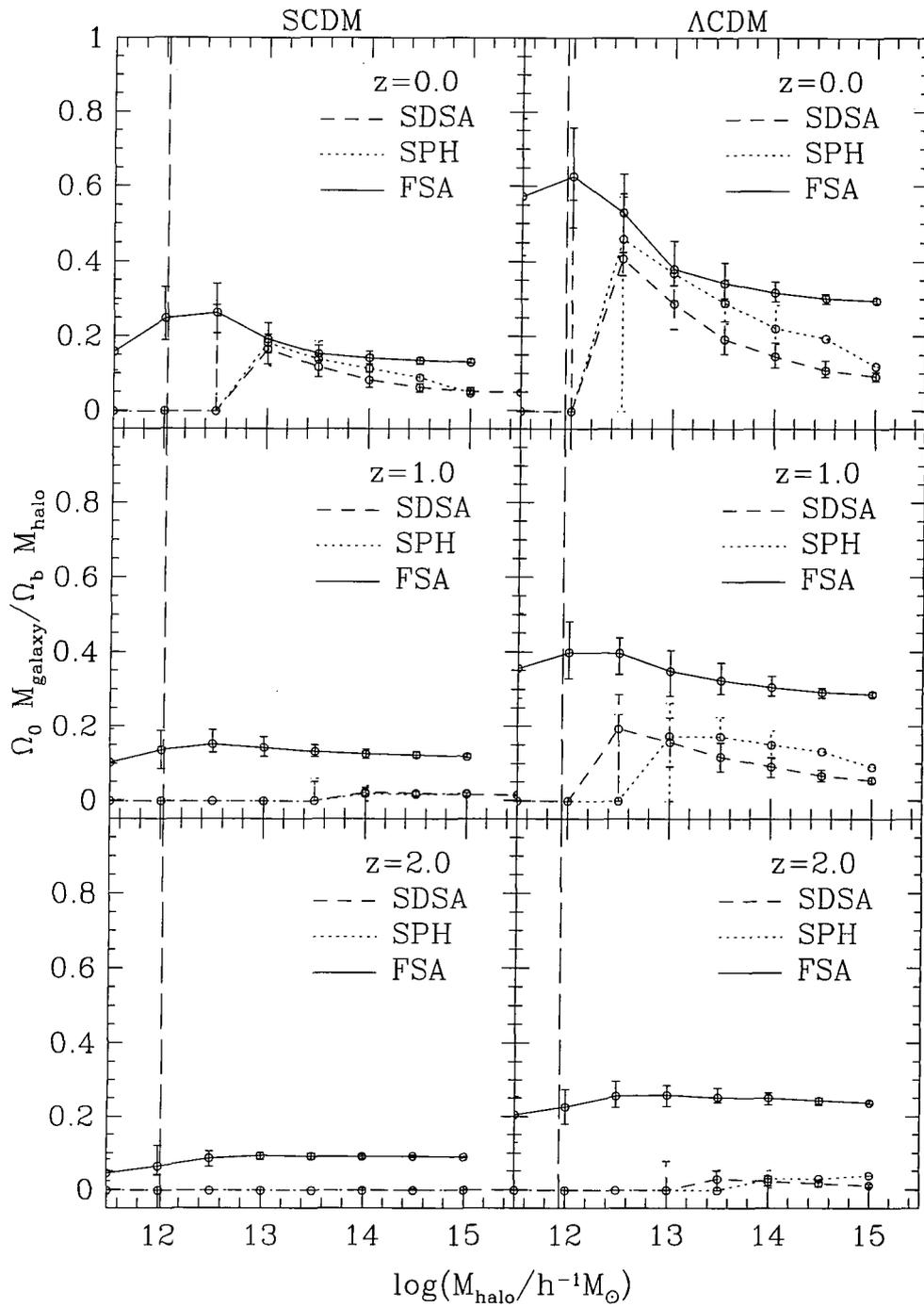


Figure 7.4: (continued on next page)

Figure 7.4: (*continued*) The mass of gas relative to the total (i.e. dark matter plus gas) halo mass at $z = 0$ in progenitor galaxies, as a function of present-day halo mass. (At $z = 0$ we include all galaxies in the halo, whilst at higher redshifts we include all galaxies in progenitors of this halo.) The ordinate shows this fraction multiplied by the universal baryon fraction. Only progenitor galaxies more massive than N'_{SPH} gas particles are considered. Results are shown at $z = 0, 1$ and 2 for SCDM (left-hand panels) and Λ CDM (right-hand panels). The solid lines show the median of the distribution in the FSA model, with errorbars indicating the 10 and 90 percentiles. The dotted lines show the corresponding fraction for galaxies in the SPH simulation, and the dashed lines for galaxies in the SDSA model. The long-dashed, vertical lines indicate the mass of those halos which, on average, contain a total gas mass (including the hot and cold components) equal to 64 times the SPH gas particle mass (assuming that the gas mass is Ω_b/Ω_0 times the total halo mass).

cant fraction of this gas, reducing the galaxy phase fraction. In SCDM, stronger feedback is required to match the observed galaxy luminosity function than in Λ CDM, and this is reflected in a smaller galaxy gas fraction.

Since it is N'_{SPH} that determines which halos are resolved in the SDSA model, this parameter affects the gas fractions in all three phases. For example, reducing N'_{SPH} from 64 to 32 worsens the agreement between the galaxy phase fractions in the SDSA and SPH models in the Λ CDM cosmology as shown in Fig. 7.3, although the two models are still reasonably close, differing only by approximately 15% at $z = 0$. The differences produced by this entirely plausible change in N'_{SPH} are indicative of the degree of uncertainty inherent in these comparisons.

Galaxy Gas Fractions in Halos

Figure 7.4 shows the mass of “galaxy-phase” gas, expressed as a fraction of the total (i.e. dark matter plus gas) halo mass and scaled by the universal baryon fraction. At $z > 0$ the quantity shown is the fraction of gas which exists in progenitors of the $z = 0$ halos in each mass bin. This is a bulk quantity for each halo and is independent of the way in which gas is divided among the galaxies that reside in each halo. The SDSA model predicts somewhat less galaxy gas than the SPH simulation, particularly in the Λ CDM cosmology. The difference is typically $\sim 50\%$, except at low masses where the agreement is better. (We do not plot error bars for the largest mass bins in the SPH simulations

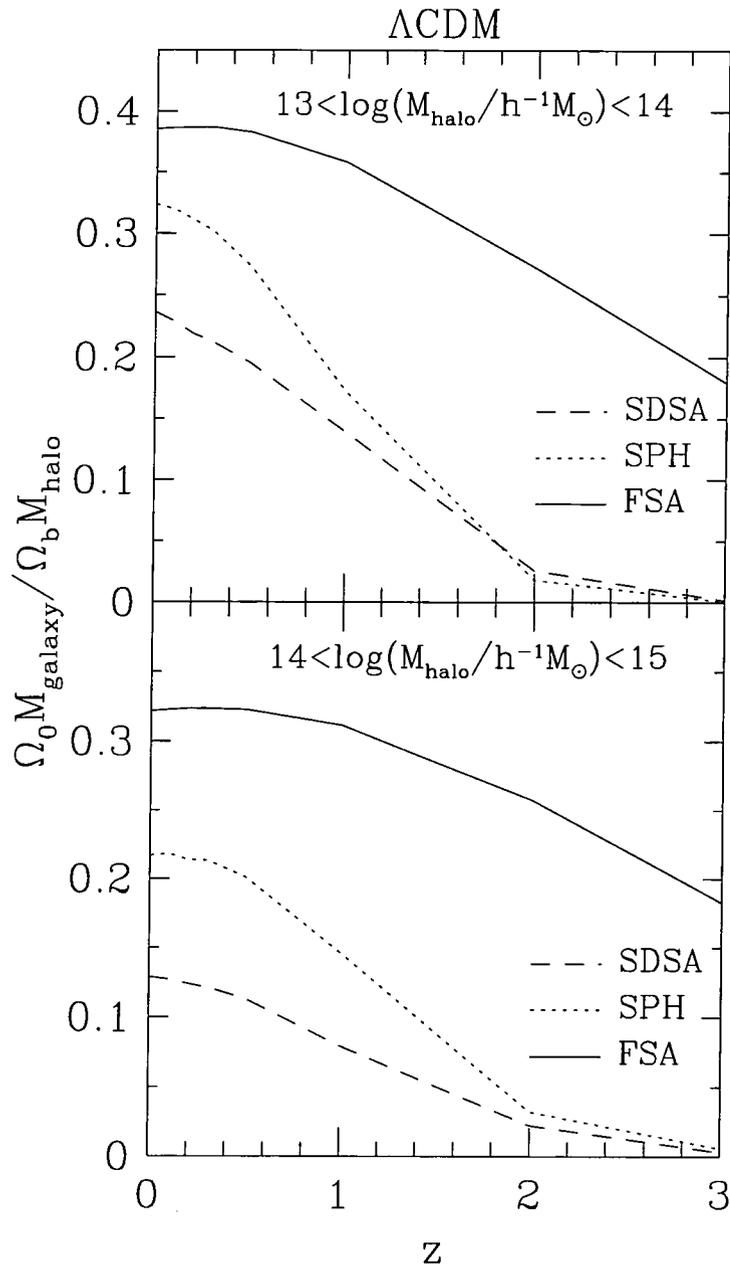


Figure 7.5: The mean mass of gas in the galaxy phase relative to the total (i.e. dark matter plus gas) mass of the $z = 0$ halo, scaled by the universal baryon fraction, as a function of redshift in progenitors of present-day halos of mass 10^{13} – $10^{14}h^{-1}M_{\odot}$ (upper panel) and 10^{14} – $10^{15}h^{-1}M_{\odot}$ (lower panel). Results are shown for the Λ CDM cosmology. Dashed lines show the SDSA models, dotted lines the SPH models and solid lines the FSA models.

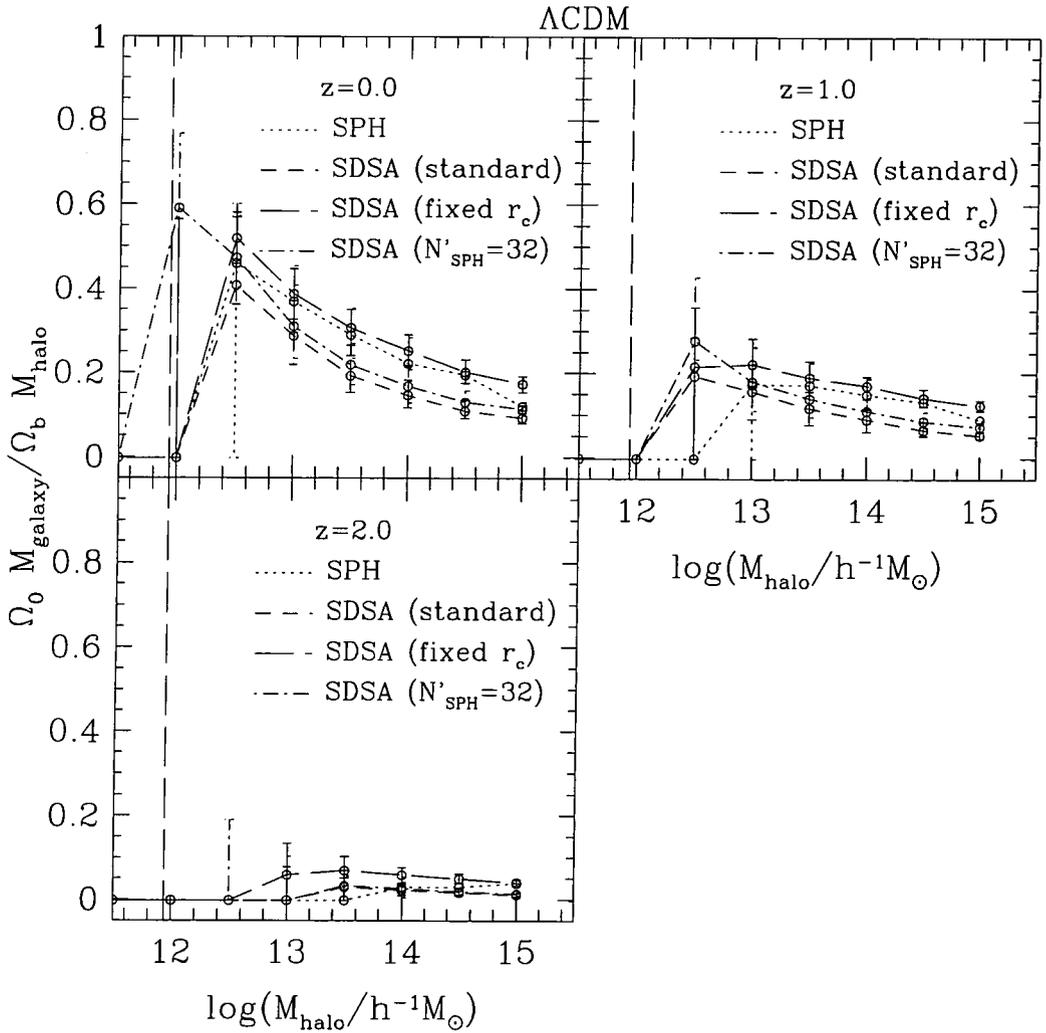


Figure 7.6: The mass of gas relative to the total (i.e. dark matter plus gas) halo mass at $z = 0$ in progenitor galaxies, as a function of present-day halo mass in the Λ CDM cosmology. (At $z = 0$ we include all galaxies in the halo, whilst at higher redshifts we include all galaxies in progenitors of this halo.) The ordinate shows this fraction multiplied by the universal baryon fraction. The dotted line shows the median, and the errorbars the 10 and 90 percentiles of the distribution in the SPH simulation. The remaining lines show predictions calculated from the SDSA model for: standard parameters (short-dashed line); a hot gas core radius which is fixed at 0.33 times the NFW scale length (long-dashed line); and a critical mass for cooling equal to 32 times the SPH gas particle mass (dot-dashed line). The long dashed, vertical line indicates the mass at which halos contain a total mass in gas (i.e. hot halo and galaxy gas) equal to 64 times the SPH gas particle mass (assuming that the gas mass is Ω_b/Ω_0 times the total halo mass). Results are shown at $z = 0, 1$ and 2 as indicated in the panels.

because they contain only one or two halos.) We note that this is significantly more than the 10% change in cooled gas mass found by Kay et al. (1999) when the Sutherland & Dopita (1993) cooling function was used in SPH simulations (which, in any case, increased the amount of gas that cooled). At higher redshifts, the two model curves become closer, indicating different mean rates of gas cooling. At the smallest masses plotted, the SPH and SDSA curves turn over near the critical cooling mass (marked on the plot by the vertical long dashed line) because of resolution effects, while the FSA model turns over because of the effects of feedback which begin to become efficient at reheating cold galaxy gas in halos below $\sim 10^{12}h^{-1}M_{\odot}$.

As noted in §7.3.1, the FSA model cools much more gas into the galaxy phase than the SPH and SDSA models. Much of this gas cools at high redshift in low mass progenitor halos (which are unresolved in the SDSA and SPH models) as may be seen from the lower panels of Fig. 7.4.

Differences in the cooling rates in the Λ CDM cosmology can be seen more clearly in Fig. 7.5, where we plot the mean fraction of cold gas in progenitor galaxies for two bins of present-day halo mass: 10^{13} – $10^{14}h^{-1}M_{\odot}$ (upper panel) and 10^{14} – $10^{15}h^{-1}M_{\odot}$ (lower panel). At high redshifts, the SPH cooling rate is very similar to the cooling rate in the SDSA model, but it becomes faster at lower redshifts. Similar trends are seen in the SCDM cosmology, although the differences are smaller, leading to very similar present-day gas fractions in the two models. In Λ CDM, the SPH model ends up with 30-70% more cool gas (i.e. the gas mass in the SPH model is between 1.3 and 1.7 times that in the SDSA model). In both cosmologies, the net cooling rate in the FSA model is slower than in the SPH model (particularly in the lower range of halo mass), but comparable to that in the SDSA model. Whilst the FSA model resolves more progenitors, which speeds cooling, feedback reduces the effective cooling rate. However, since the FSA model contains progenitor halos at much higher redshifts than the other models (due to its greater resolution), more gas has already cooled in the FSA model by $z = 2$ than in either the SPH or SDSA models.

Since the amount of gas that can cool depends upon the density profile assumed for the hot halo and on the critical cooling mass, it is interesting to see how sensitive the agreement between the SDSA model and the SPH simulation is to variations in these parameters. We show the result of this test in Fig. 7.6 for the Λ CDM cosmology. (The trends are similar for SCDM.) Halving the critical mass required for gas cooling (i.e. reducing N'_{SPH} from 64 to 32) in the SDSA model has a small effect, marginally improving

the agreement with the SPH simulation. (Note that for $N'_{\text{SPH}} = 32$, we plot results only for objects more massive than 64 gas particles in order to compare directly to the other curves.) The critical cooling mass is not, of course, a precise number, and this comparison suggests that we may have been too conservative in setting it equal to $2 \times N_{\text{SPH}}$. In any case, it appears that the agreement between the galaxy gas mass as a function of halo mass in the SDSA and SPH models is better than one might have expected.

Keeping the gas core radius fixed at $0.33r_s$ in all halos, rather than letting it grow as in our standard model, allows significantly more gas to cool in the SDSA model. This brings this model into excellent agreement with the SPH model at $z = 0$, as may be seen in Fig. 7.6, but at the expense of a slightly larger cooled gas mass at higher redshifts. Adopting a smaller core radius (e.g. $r_c = 0.15r_s$) makes very little difference. Finally, we note that assuming a gas profile which traces that of the dark matter leads to $\sim 50\%$ more cooling in the highest mass halos (note that this model is not shown in Fig. 7.6).

7.3.2 Properties of Galaxies

We now consider properties of the models that are sensitive to the way in which gas is apportioned amongst individual galaxies, rather than just the total amount of gas in a halo. In the SPH simulation, the number of galaxies that form in a given halo and their masses are determined by resolution effects, the gas cooling rate and the galaxy merger rate. In the semi-analytic models, the choice of density profile for the hot corona determines the cooling rate of the gas. We also expect the masses and numbers of galaxies to be affected by the choice of f_{df} , which controls the rate at which galaxies merge within a halo and, in the case of SDSA model, the choice of N'_{SPH} , which determines the minimum halo mass in which galaxies can form.

Galaxy Mass Function

Figure 7.7 shows the differential galaxy mass functions in the models at $z = 0, 1$ and 2 . (Recall that, by definition, M_{galaxy} consists only of cold gas in the SPH and SDSA cases, but can also include stars in a disk and spheroid in the FSA case.) The agreement between the SPH and semi-analytic models is somewhat different in the two cosmologies. In the SCDM case, there is quite good agreement for large galaxy masses, but the SPH simulation produced about twice as many galaxies with mass of a few times $10^{11}M_{\odot}$ than the SDSA model. In the Λ CDM case, the SPH simulation produced about 3–4 times as many galaxies as the SDSA model over most of the mass range above the critical cooling

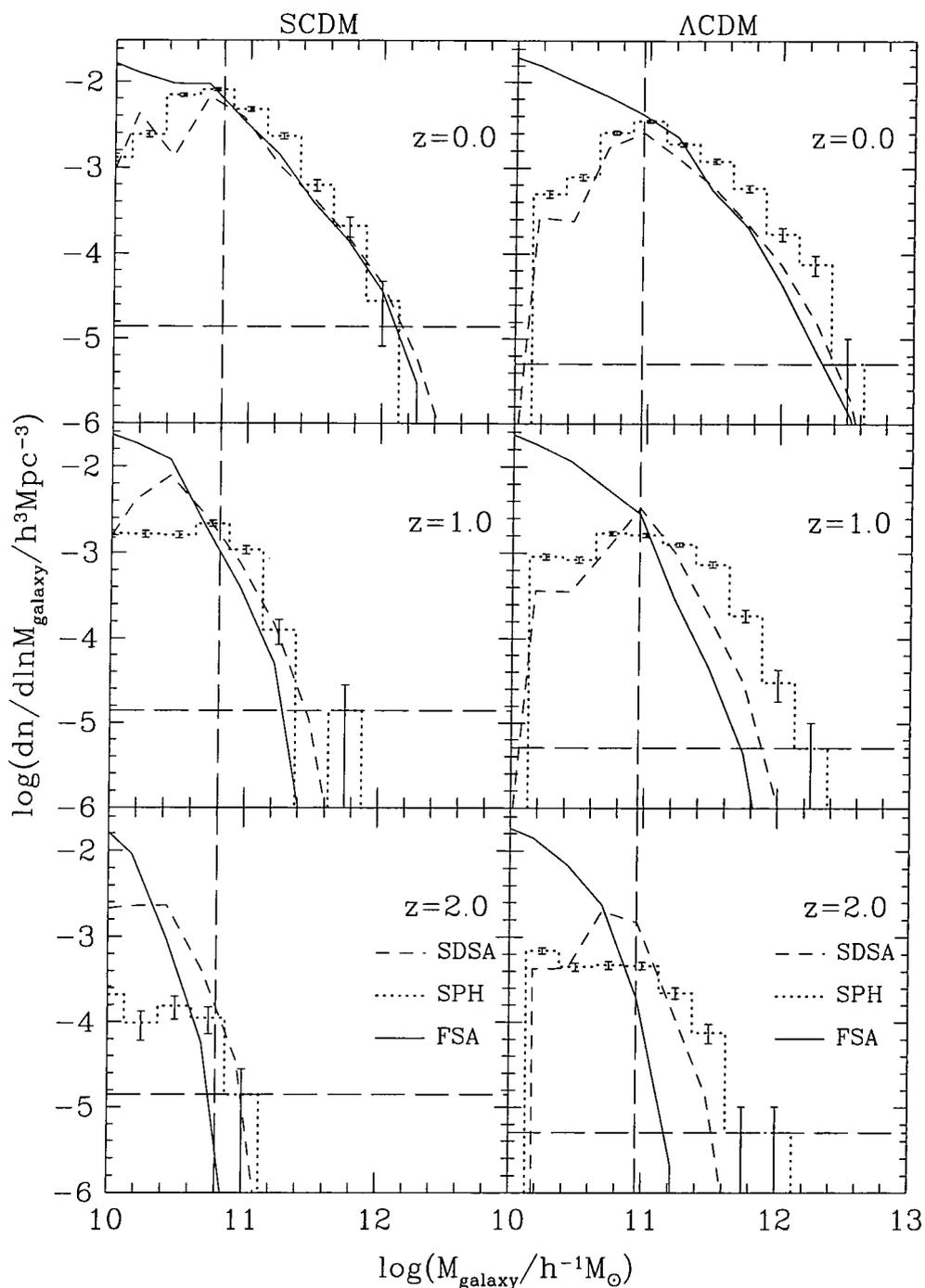


Figure 7.7: The differential galaxy mass function in the SCDM (left-hand panels) and Λ CDM (right-hand panels) cosmologies at various epochs. The solid line corresponds to the FSA model, the dotted histogram to the SPH simulation and the dashed line to the SDSA model. The vertical dashed line indicates the critical mass for cooling in the SPH simulations (which is 64 times the gas particle mass), whilst the horizontal dashed line indicates the abundance corresponding to one object in the entire SPH simulation volume. Results are shown at $z = 0, 1$ and 2 as indicated in each panel.

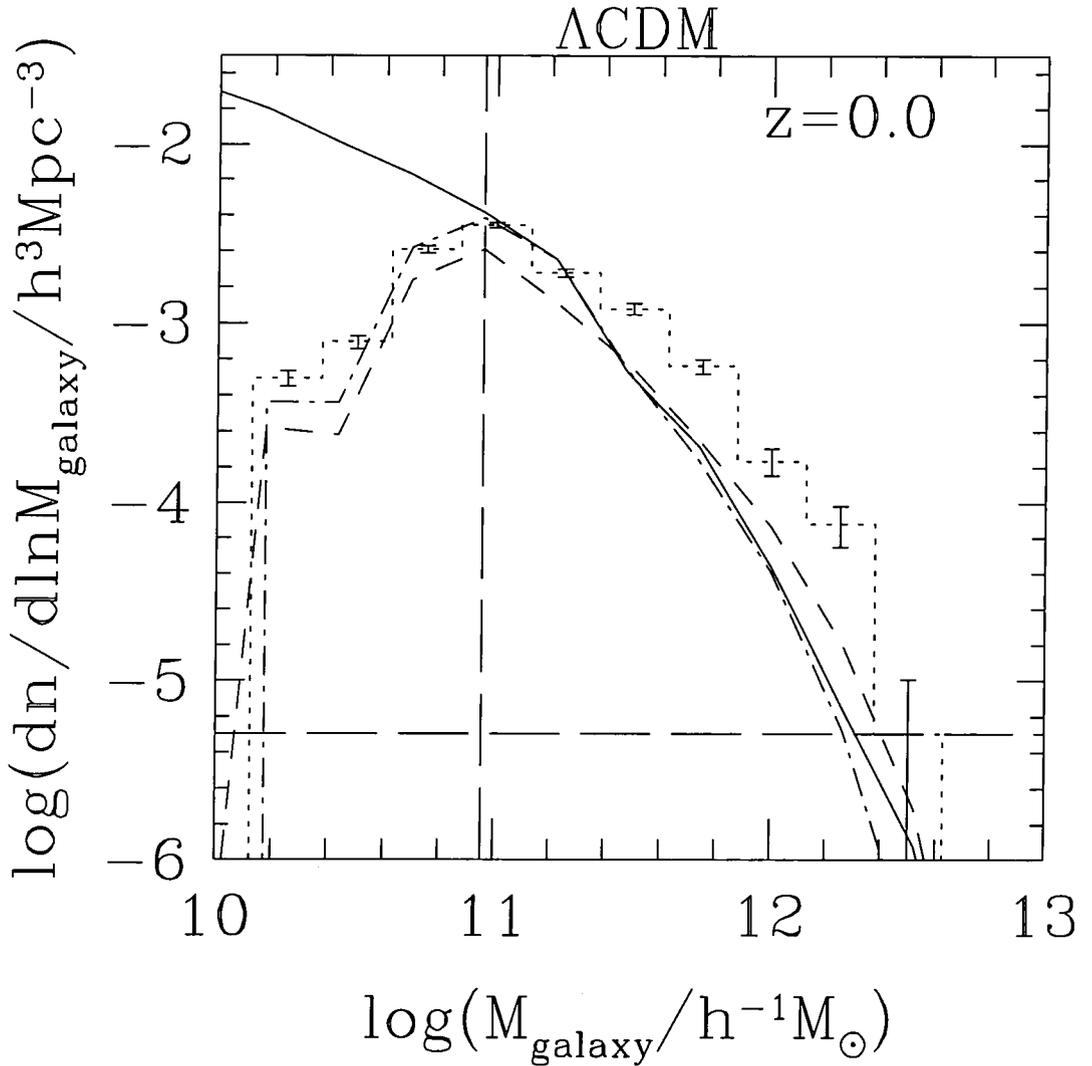


Figure 7.8: The differential galaxy mass function at $z = 0$ in the Λ CDM cosmology. The solid line corresponds to the FSA model, the dotted histogram to the SPH simulation and the dashed line to the SDSA model. The dot-dashed line shows the results for the SDSA model using the standard Press-Schechter dark matter halo mass function and the standard extended Press-Schechter formalism (i.e. with $f_{\delta_c} = 1$) for generating merger trees. The vertical dashed line indicates the critical mass for cooling in the SPH simulations (which is 64 times the gas particle mass), whilst the horizontal dashed line indicates the abundance corresponding to one object in the entire SPH simulation volume.

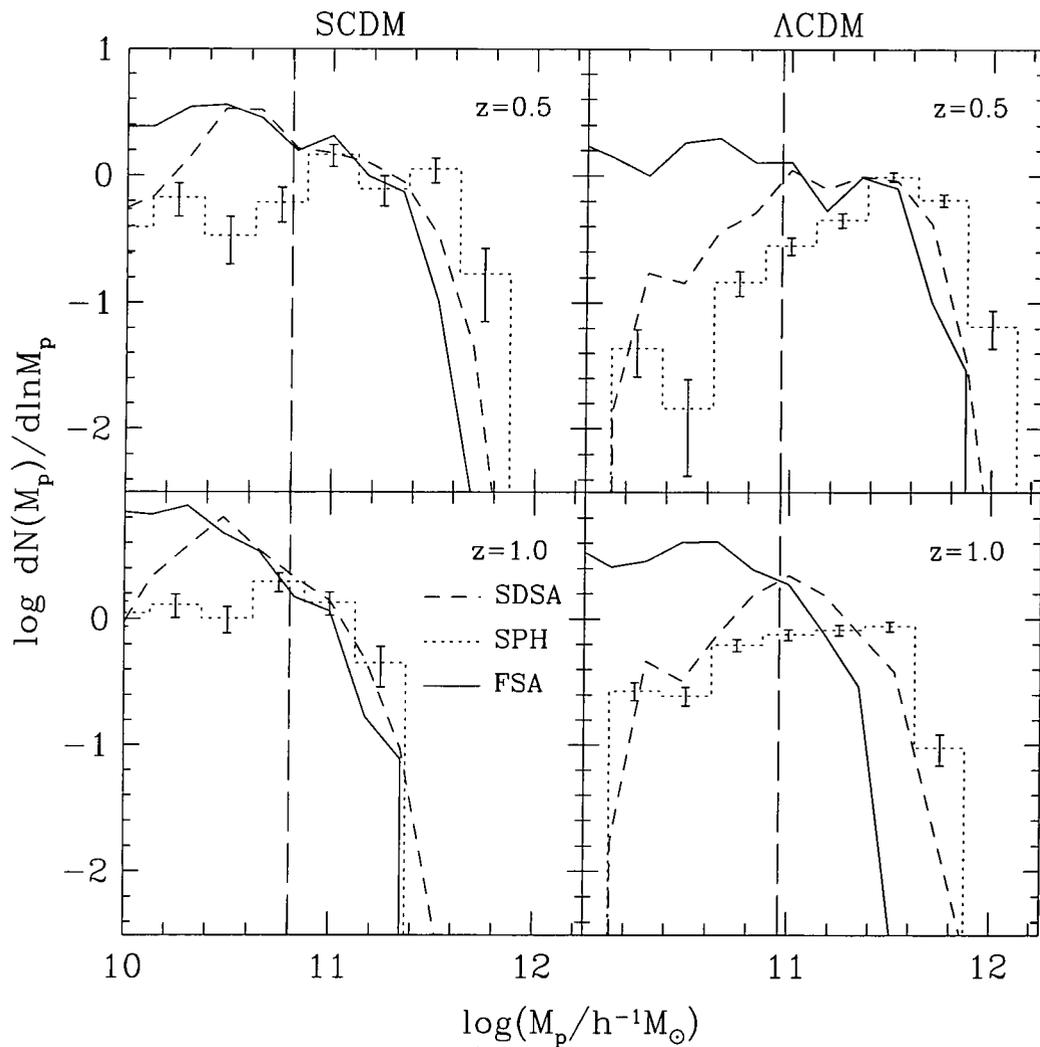


Figure 7.9: The mass function of galaxies which are progenitors of present-day galaxies with (cold gas plus stellar) masses $10^{11.5}-10^{12}h^{-1}M_\odot$. Results are shown at $z = 0.5$ and 1 for SCDM (left-hand panels) and Λ CDM (right-hand panels). The solid line corresponds to the FSA model, the dotted histogram to the SPH simulation and the dashed line to the SDSA model. The vertical dashed line indicates the mass corresponding to 64 SPH gas particles.

mass. If viewed as a difference in mass at fixed abundance, the discrepancy can be as large as a factor of almost 2. For a fixed mass-to-light ratio, this would lead to the bright end of the luminosity function being approximately 0.75 magnitudes brighter in the SPH simulation than in the SDSA model. The differences between SDSA and SPH models are similar at higher redshifts, whilst the FSA mass function evolves more rapidly.

First, we note the interesting behaviour at the low mass end of the mass function. Below the critical cooling mass, 9.0×10^{10} and $6.4 \times 10^{10} h^{-1} M_{\odot}$ in Λ CDM and SCDM respectively, the abundance of galaxies in the SDSA and SPH models drops very sharply, due to the effects of resolution. These prevent the formation of small halos and also inhibit the cooling of gas in marginally resolved halos. In the FSA model, on the other hand, the number of galaxies continues to rise towards small masses. Although feedback becomes gradually more efficient towards lower masses, this still allows many low mass galaxies to form. In both the SDSA and SPH models some galaxies do form with mass below the cooling limit. In the former case this can occur because cooling is prevented only in halos with a total hot gas mass lower than the limit. It is still possible for cooling in a more massive halo to create a galaxy below the mass threshold simply because there has not been enough time for the total hot gas mass of the halo to cool. The SPH simulations also contain galaxies below the mass limit because the cutoff is not sharp, but marks the point at which cooling becomes efficient. These objects represent young galaxies in the process of formation and, except at early times, they make up a small fraction of the total number of galaxies. In addition, in the SPH model there is a small contribution at low masses from galaxies that are being tidally disrupted in clusters.

In Fig. 7.8 we show the effect on the SDSA galaxy mass function (in the Λ CDM cosmology) of using the standard Press-Schechter halo mass function and $f_{\delta_c} = 1$, rather than the Sheth-Mo-Tormen halo mass function and f_{δ_c} as given by eqn. (7.1). Using the latter, more accurate, mass function produces somewhat more massive galaxies. The choice of f_{δ_c} alters the mass function of progenitors of the more massive galaxies, but has little effect elsewhere.

Some insight into the origin of the differences between the galaxy mass functions in the SPH and SDSA models may be obtained by considering the mass function of galaxy progenitors. For present-day galaxies with masses in the range $10^{11.5} - 10^{12} h^{-1} M_{\odot}$, Fig. 7.9 shows progenitor mass functions at $z = 0.5$ and $z = 1$. These are normalized to the present day abundance of galaxies (i.e. $(dN(M_P)/d \ln M_P) d \ln M_P$ is the number of progenitors in the mass range $\ln M_P$ to $\ln M_P + d \ln M_P$ per parent galaxy) so that

the differences in the $z = 0$ mass functions seen in Fig. 7.7 do not propagate through to this comparison. It is immediately obvious from this figure that the SDSA galaxies have fewer high-mass progenitors than the galaxies in the SPH simulation, particularly in the Λ CDM cosmology.

The discrepancy between the SPH and semi-analytic galaxy mass functions seen in Figs. 7.7 and 7.9 could be due either to differences in the galaxy merger rates or in the gas cooling rates in the two models. We first consider the possibility that the SPH and SDSA galaxy merger rates may be different, leading to different numbers of galaxies forming in each halo. As we discussed in §7.2.1, merging in the SPH simulations may be artificially enhanced by non-physical effects introduced, for example, by the use of softened forces. The strength of these effects will depend in a non-trivial way on the formation epoch of each galaxy. Whilst we cannot attempt to mimic the details of such effects in the semi-analytic model, we can explore the consequences of a global change in the merger timescale. In the semi-analytic model, when a galaxy falls into a larger halo, it is assumed to sink to the centre in a time proportional to the dynamical friction timescale for an object orbiting in an isothermal halo. As shown in Appendix A.4, this may be written as:

$$\tau_{\text{mrg}} = f_{\text{df}} \theta_{\text{orb}} \tau_{\text{dyn}} \frac{0.3722}{\ln(\Lambda_{\text{Coulomb}})} \frac{M_{\text{H}}}{M_{\text{sat}}}, \quad (7.5)$$

where f_{df} is a dimensionless parameter; M_{H} is the mass of the halo in which the satellite orbits; M_{sat} is the mass of the satellite galaxy including the mass of the dark matter halo in which it formed (not including the mass of the satellite's dark matter halo leads to an overestimate of the dynamical friction timescale as shown by Navarro, Frenk & White 1995); τ_{dyn} is the dynamical time of the large halo and the Coulomb logarithm is $\ln(\Lambda_{\text{Coulomb}}) = \ln(M_{\text{H}}/M_{\text{sat}})$. The variable θ_{orb} contains the dependence of the merger timescale on the orbital parameters of the galaxy and is chosen from the distribution found in N-body simulations by Tormen (1997), as described in §2.4.1.

Cole et al. (2000) point out that the formula in eqn. (7.5) has been derived on the basis of a number of assumptions, for example that the satellite galaxy may be treated as a point mass. Thus, whilst they recommend a default value of $f_{\text{df}} = 1$, they also allow themselves the freedom to choose a different value if required to produce a realistic model. Recently, Colpi, Mayer & Governato (1999) have examined the validity of eqn. (7.5) in detail using both analytic and numerical techniques. On the basis of their investigations, they suggest small modifications to the formula, in particular a slightly different dependence of θ_{orb} on the orbital eccentricity. They also find that the effects of tidal stripping produce $f_{\text{df}} \approx 2.7$

(for a specific model of the satellite's dark matter halo). Both of these changes act to increase τ_{mrg} , and so reduce the merger rate. Since the SDSA model already contains too few high mass galaxies compared to the SPH simulations, a slower merger rate would merely increase the discrepancy.

At the resolution of our SPH simulations, an infalling satellite galaxy will lose nearly all of its original dark matter halo shortly after entering the larger halo. On the other hand, in the SPH simulations, merging with the central object is driven not only by dynamical friction, but also by drag due to viscous effects as the satellite moves through the hot halo of the cluster (Frenk et al. 1996, Tittley, Couchman & Pearce 1999). These processes may drive the effective f_{df} to a value less than unity.

The smaller the value of f_{df} , the faster a galaxy will sink to the centre of its host halo. In Fig. 7.10 we show the effect of reducing the merging rate from the default value, $f_{\text{df}} = 1$, to $f_{\text{df}} = 0.4$ on the SDSA galaxy mass function in the Λ CDM model. This change improves the match between the SDSA and SPH models at $z = 0$, but only slightly, and mostly at high masses. Below $10^{12}h^{-1}M_{\odot}$, the two models still differ significantly. At higher redshift the improvement is slightly better, but the SDSA model still contains fewer high-mass objects than the SPH simulation. Thus, a change in the global merger timescale does not help to reconcile the two models. This is perhaps not too surprising since this simple alteration cannot mimic any redshift dependence of the merger timescales in the SPH simulation. Decreasing f_{df} further produces more high mass galaxies, at the expense of depleting the number of lower mass galaxies, thereby worsening the agreement between the SDSA and SPH galaxy mass functions at the low mass end.

A second possible explanation for the discrepancy between the galaxy mass functions is that the relative gas cooling rates in the SPH and SDSA models are different. We investigate the effects of changing the rate at which gas cools in the SDSA model, by keeping the core radius of the hot gas density profile equal to a fixed fraction of the NFW scale radius in the halo. This has the effect of allowing gas to cool more rapidly in any given halo. The galaxy mass function of the SDSA model for this case is shown by the dot-dashed line in Fig. 7.10. The agreement with the SPH simulation is now much improved.

In the next section we will explore how changing merger and cooling rates affects the number of galaxies found in halos of a given mass and thus identify the underlying cause of the differences between the mass functions in the SDSA and SPH models.

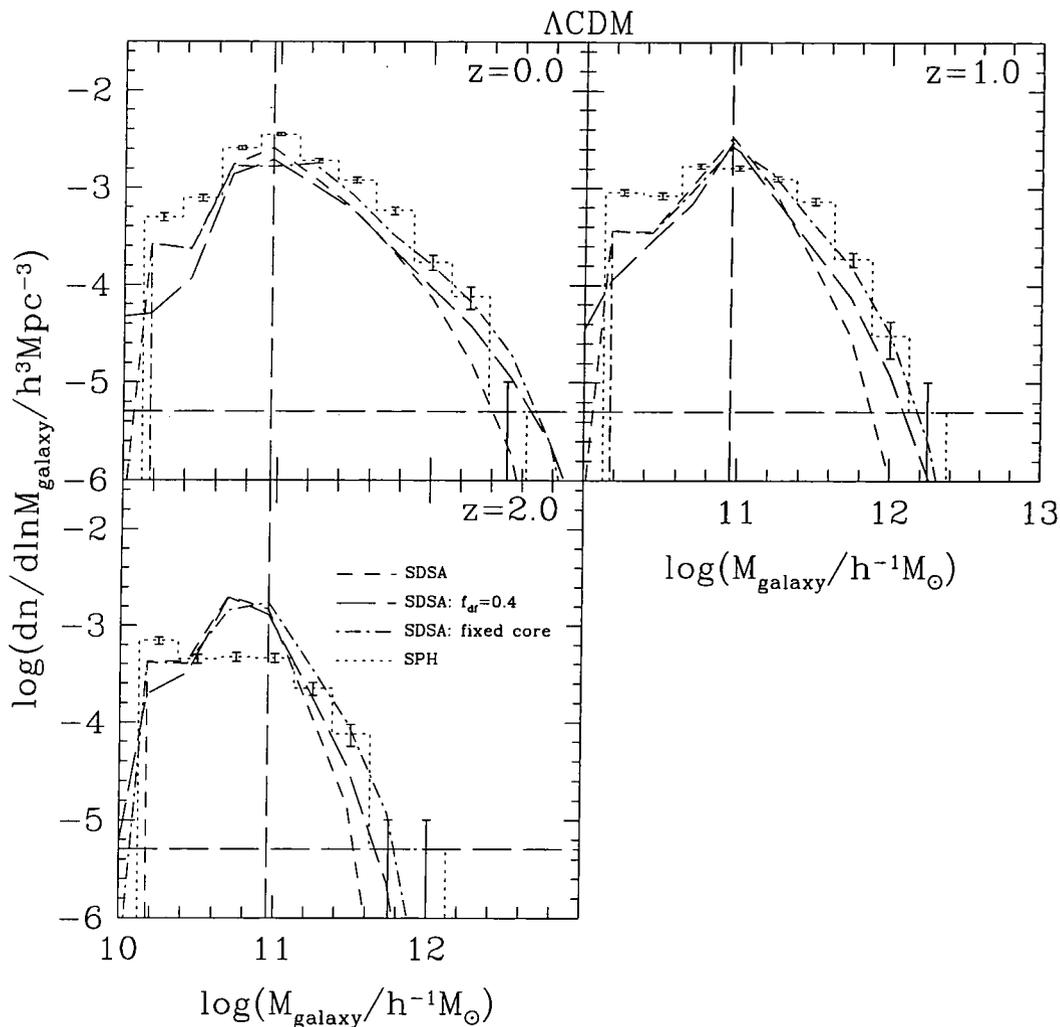


Figure 7.10: The differential galaxy mass function at selected redshifts in the Λ CDM cosmology predicted by the SDSA model when (i) using a faster merging rate (long-dashed line) and (ii) keeping the core radius of the hot gas density profile fixed (dot-dashed line). For reference, the short-dashed line shows the standard SDSA model. Dotted histograms show the SPH simulation results. The vertical dashed lines indicate the critical mass for cooling in the SPH simulations (which is 64 times the gas particle mass), whilst the horizontal dashed line indicates the abundance corresponding to one object in the entire SPH simulation volume. Results are shown at $z = 0, 1$ and 2 , as indicated in each panel.

Number of Galaxies per Halo

The redshift evolution of the mean number of progenitor galaxies per halo in the Λ CDM cosmology is shown in Fig. 7.11, for two ranges of present-day halo masses: 10^{13} – $10^{14}h^{-1}M_{\odot}$ (upper panel) and 10^{14} – $10^{15}h^{-1}M_{\odot}$ (lower panel). (Qualitatively similar results are obtained for the SCDM cosmology.) In the SPH simulation there are 165 halos in the lower mass range and 12 in the higher mass range. Since the galaxy abundance is a steep function of galaxy mass (c.f. Fig. 7.7) the number of galaxies per halo above some particular mass cut is always dominated by galaxies close to the cutoff. As a result, the number of galaxies per halo more massive than 64 SPH particles may be affected by differences in the way in which the resolution limit works in SPH and SDSA models. Therefore, we plot results for galaxies more massive than 64 SPH gas particles (thin lines), but also for those more massive than 128 SPH gas particles (heavy lines) which are much less affected by resolution effects.

The shape of the abundance curves in Fig. 7.11 is determined by the relative rates of galaxy formation and merging and has the same basic form in the SDSA and SPH models. The number of galaxies more massive than 64 SPH gas particles predicted by the SDSA model in halos of mass 10^{13} – $10^{14}h^{-1}M_{\odot}$ at the present day agrees well with that found in the SPH simulation. The level of agreement, however, is less impressive at higher redshift and for higher mass halos. For the 128 particle selection the SDSA model contains *fewer* galaxies per halo than the SPH model in both ranges of halo mass and at all redshifts.

In the preceding subsection, we saw that increasing the merger rate in the Λ CDM SDSA model leads to slightly better agreement between its galaxy mass function and that of the SPH simulation. A faster merger rate, however, depletes the number of galaxies by combining them into larger galaxies. In Fig. 7.12, we show the effect of reducing f_{df} from 1 to 0.4 on the evolution of the progenitor population in the SDSA model. For galaxies more massive than 64 SPH gas particles, the rate at which N_{gal} increases at high redshift is now slower than before and results in a lower value of N_{gal} at the present day. The net effect is to bring the SDSA curve closer to the SPH results for the 10^{14} – $10^{15}h^{-1}M_{\odot}$ halo sample, but to increase the discrepancy for the 10^{13} – $10^{14}h^{-1}M_{\odot}$ halos. For galaxies more massive than 128 SPH gas particles, the already low galaxy numbers are depleted even further, exacerbating the discrepancy with the SPH results. We conclude that the discrepancies in the mass functions in the Λ CDM SDSA and SPH models are not due to

differences in the galaxy merger rates in the two models.

The effects of altering the cooling rate in the SDSA model are also illustrated in Fig. 7.12 (dot-dashed lines). As before, we have varied the cooling rate simply by keeping the core radius of the hot gas density profile fixed. For galaxies more massive than 64 SPH gas particles, the SDSA model now overpredicts the mean number of galaxies per halo at all redshifts, particularly in the most massive halos. The increase is driven by galaxies near the 64 particle cutoff whose mass increases due to the additional cooling of gas. (A similar behaviour occurs if the cooling rate is enhanced by reducing the threshold for cooling from $N'_{\text{SPH}} = 64$ to 32, rather than by changing the density profile of the gas.)

A cleaner test can be made by looking at galaxies more massive than 128 SPH gas particles. In this regime, the SDSA model with enhanced cooling is in excellent agreement with the SPH simulation. As we saw earlier, such a modification of the SDSA model also produces a galaxy mass function (Fig. 7.10) and a distribution of cold gas in halos of different mass (Fig. 7.6) that are very similar to those in the SPH simulation. We conclude therefore that the main reason for the differences we have found between the SDSA model and the SPH simulation is that gas cools more efficiently in massive halos in the SPH simulation. This difference, and the corresponding differences in the mass function, are larger in Λ CDM than in SCDM, presumably because of the larger time interval during which gas can cool in Λ CDM (as structures tend to form earlier in this cosmology). Increasing the cooling rate in the SDSA model slightly spoils the excellent agreement with the SPH simulation on the global phase fractions (Fig. 7.2). The effect, however, is small (a maximum difference of 50% as opposed to the original 25%) and of the same order as the effect of assuming a different resolution limit for the SDSA model (Fig. 7.3).

7.3.3 Spatial Distribution of Galaxies

As a final comparison, we consider the clustering of galaxies, as measured by the two-point correlation function. This is plotted in Fig. 7.13 for galaxies more massive than N'_{SPH} gas particles. Note that this selection criterion (which picks out only rather massive galaxies) is very different from that considered in Chapter 3 and, as a result, the correlation functions plotted in Fig. 7.13 are quite different from those in Chapter 3. Note also that the relatively small volume of the simulations affects the determination of the correlation function for pair separations greater than a few Mpc. To compute the correlation function in the SDSA and FSA models, we make the further assumption that the galaxies trace

the mass within each dark matter halo.

The agreement of the correlation functions of all three models in both cosmologies is very good, except on scales below $1h^{-1}\text{Mpc}$ in the SCDM cosmology, where the semi-analytic models have a lower amplitude than the SPH simulation. The good agreement is not completely unexpected as the mean number of galaxies per halo is similar in all models, but it does demonstrate that the assumption that galaxies trace the mass within individual halos is a reasonable approximation, at least for studies of the galaxy correlation function. In particular, all the models predict very similar evolution in the correlation function, with galaxies becoming strongly biased at high redshifts.

In Fig. 7.14 we show the effects of reducing the merger timescale in the ΛCDM cosmology on the galaxy two-point correlation function. The enhanced merger rate significantly lowers the SDSA correlation function below that of the SPH model. Enhancing the cooling rate of gas (by using a fixed gas core radius) in the SDSA model has a negligible effect on the correlation function. This is further evidence that the main difference between the SDSA and SPH models is not a gross difference in galaxy merger rates, but rather a difference in the efficiency with which gas cools.

7.4 Discussion and Conclusions

In this Chapter we have explored the consistency of two very different techniques commonly used to study galaxy formation, semi-analytic modelling and SPH simulations. For this purpose, we constructed a simplified or ‘stripped-down’ semi-analytic model (SDSA), designed to mimic the conditions of the SPH simulations as closely as possible. An artificial resolution limit, similar to that in our SPH simulations, was imposed on the SDSA model and the standard star formation and feedback prescriptions normally used in semi-analytic modelling were stripped out. Furthermore, the dark matter halo mass function proposed by Sheth, Mo & Tormen (1999) was used in the SDSA model instead of the standard Press-Schechter formula since the former provides a better match to the mass function in N-body simulations. Similarly, we also modified the extended Press-Schechter theory in order to better match the distribution of progenitor halo masses obtained in the SPH simulations. In this way, we minimize differences in the evolution of dark matter halos and concentrate on the aspect of interest in this work: the evolution of the gas. In this Chapter, we have focussed on a *statistical* analysis of the differences between the SDSA and SPH models. As a secondary aim, we have also compared the outcome of the

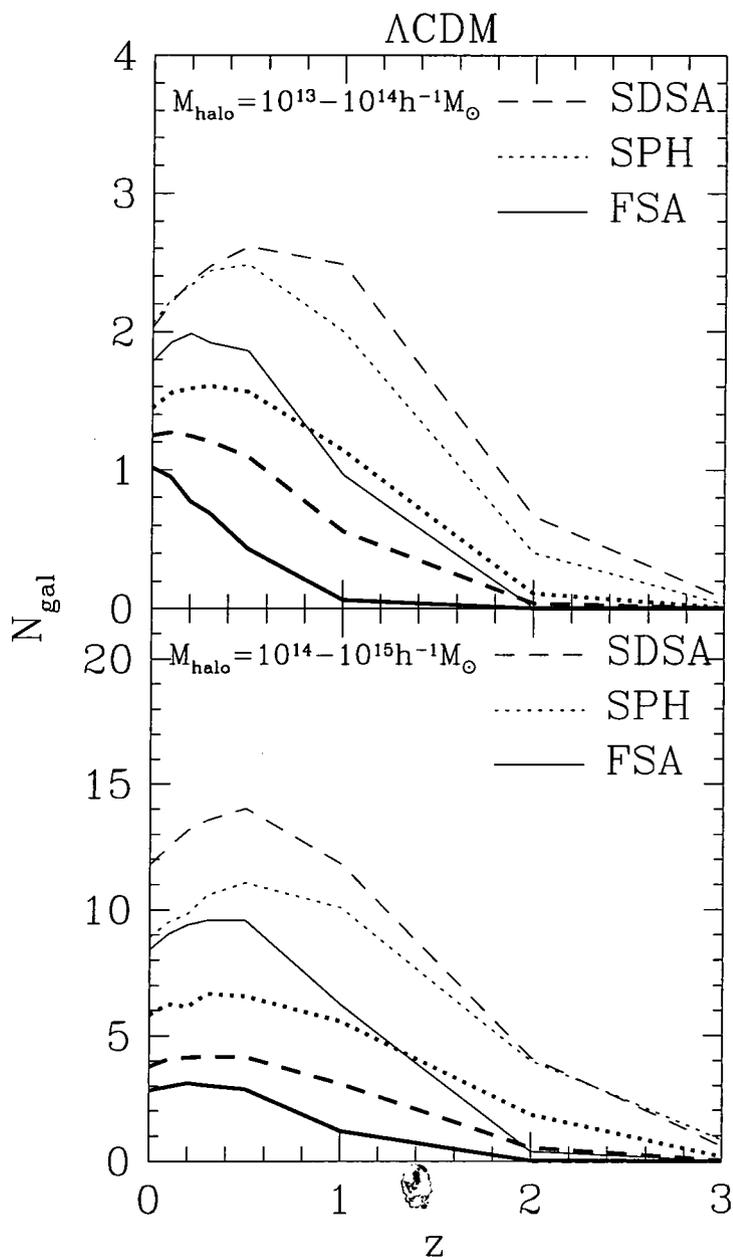


Figure 7.11: The mean number of progenitor galaxies with mass greater than 64 SPH particles (thin lines) and 128 particles (heavy lines) in the Λ CDM cosmology, as a function of redshift, for halos of present-day mass $10^{13} - 10^{14} h^{-1} M_{\odot}$ (upper panel) and $10^{14} - 10^{15} h^{-1} M_{\odot}$ (lower panel). Dashed, dotted and solid lines show results from the SDSA, SPH and FSA models respectively.

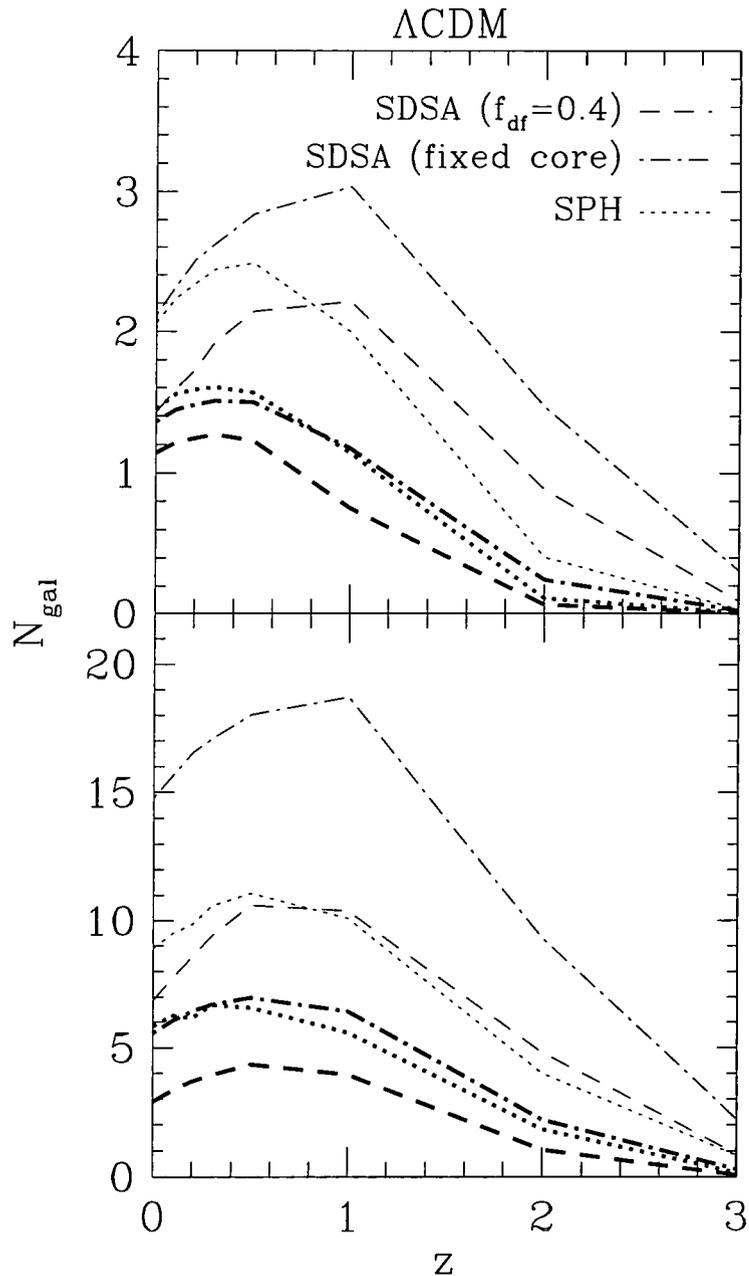


Figure 7.12: The mean number of progenitor galaxies with mass greater than 64 SPH particles (thin lines) and 128 particles (heavy lines), as a function of redshift in halos of present-day mass $10^{13}-10^{14} h^{-1} M_{\odot}$ (upper panel) and $10^{14}-10^{15} h^{-1} M_{\odot}$ (lower panel) in the Λ CDM cosmology. The predictions of the SDSA model are shown in the cases where (i) a faster merger rate is used (dashed line) and (ii) the cooling rate is enhanced by keeping the core radius fixed in the hot gas density profile (dot-dashed line). The dotted line shows the results from the SPH simulation.

SPH simulations with a full semi-analytic model, in which the star formation and feedback prescriptions are restored and no artificial resolution limit is imposed. The motivation for this latter comparison is a desire to assess how the realism of a typical SPH simulation is likely to be compromised by the neglect, or very rough treatment, of star formation and feedback.

For our main comparisons, we considered five properties of the gas distribution in two versions of the CDM cosmology (SCDM and Λ CDM), over the redshift range 0 to 2. In order of decreasing generality, these are: (i) the global fractions of gas in three distinct phases – hot halo, galactic (i.e. cold and dense gas) and uncollapsed; (ii) the amount of cold galactic gas in halos of different mass; (iii) the abundance of objects as a function of their cold gas mass; (iv) the number of “galaxies” as a function of halo mass; and (v) the correlation function of galaxies. The main conclusion of this Chapter is that the agreement between the SPH simulation and the stripped-down version of the semi-analytic model is much better than a pessimist might have expected. The tests we have carried out generally show reasonable agreement between the two techniques, although we have found significant differences in the cold gas mass functions in the two models. Since in both approaches the gas physics are necessarily treated in an approximate fashion, it seems to us inappropriate to regard one as ‘correct’ and the other as ‘wrong,’ in the instances where discrepancies arise.

Specifically, our main results may be summarized as follows:

(i) Over the simulation volume as a whole, the global amounts of gas in the three main phases that develop (hot, cold galactic and uncollapsed) are very similar in the SDSA model and the SPH simulation. From $z = 0.5$ (the redshift by which about half the final amount of galaxy gas has cooled) to the present, the maximum difference is never larger than 25%. Perhaps surprisingly, the SDSA model produces slightly *less* hot gas than the SPH simulation at all epochs (although this difference may not be significant given the uncertainties inherent in this comparison). The distribution of cold gas in halos of different mass also agrees quite well (to within $\sim 50\%$) with the largest differences occurring in the most massive halos. These results apply to the SPH and SDSA models in their standard forms, i.e. with parameters chosen independently of this comparison in order to match certain observational constraints. Thus, we conclude that in spite of its various approximations (e.g. spherical symmetry), the simple model of gas cooling normally employed in semi-analytic models gives similar results, on average, to the SPH simulations.

(ii) Even though the total amount of cold gas is similar in the SDSA model and the SPH simulations, the actual cold gas mass functions (i.e. the abundance of galaxies as a function of cold gas mass) are different, particularly in the Λ CDM cosmology. In this case, the SPH simulation produced three to four times more high-mass galaxies than the SDSA model, or equivalently, at a fixed abundance, galaxies in the SPH simulation are, on average, twice as massive as their SDSA counterparts. We were able to discount a difference in galaxy merging rates as the dominant source of this discrepancy. Instead, we identified as the culprit a difference in the efficiency of cooling in massive galaxies: more gas cools into these galaxies in the SPH simulation than in the standard SDSA model. Thus, artificially increasing the cooling rate of gas in the SDSA model (by assuming that the core radius of the gas density profile remains fixed rather than growing with time as in our standard model) leads to excellent agreement with the SPH galaxy mass functions and also with the mass functions of galaxy progenitors.

(iii) The spatial distribution of galaxies, as characterised by the two-point correlation function, is remarkably similar in the semi-analytic and SPH models. This is true not only of the present-day distribution, but also of the clustering at high redshift. This conclusion also holds even when the cooling rate is enhanced in the SDSA model as discussed in (ii).

The primary limitation of current cosmological SPH simulations is the relatively poor resolution attainable even with the largest computers. By comparing our SPH or SDSA models to the full semi-analytic model (FSA), we gain some idea of how important these resolution effects are in practice. Furthermore, since the FSA model includes prescriptions for star formation and feedback that are not modelled at all in the SPH simulations we have considered, we can also assess how important these processes are in determining the properties of hot and cold gas in the moderately large galaxies that form in our SPH simulations. The total amount of gas that can cool in an SPH simulation is determined by the resolution limit. On the other hand, because of its intrinsically high resolution, the semi-analytic model traces the evolution of gas even in small halos which the SPH simulation assigns to the uncollapsed phase. As a result, not only is the fraction of halo gas (hot and cold) larger in the FSA model than in the SPH simulation, but its cold gas mass function extends to smaller masses that can be resolved in the SPH simulation. The FSA cold gas mass function has a much sharper cut-off at the massive end than either the SDSA or SPH models.

In summary, our comparisons demonstrate a higher level of consistency than was perhaps expected between the results of SPH simulations and the more idealized semi-

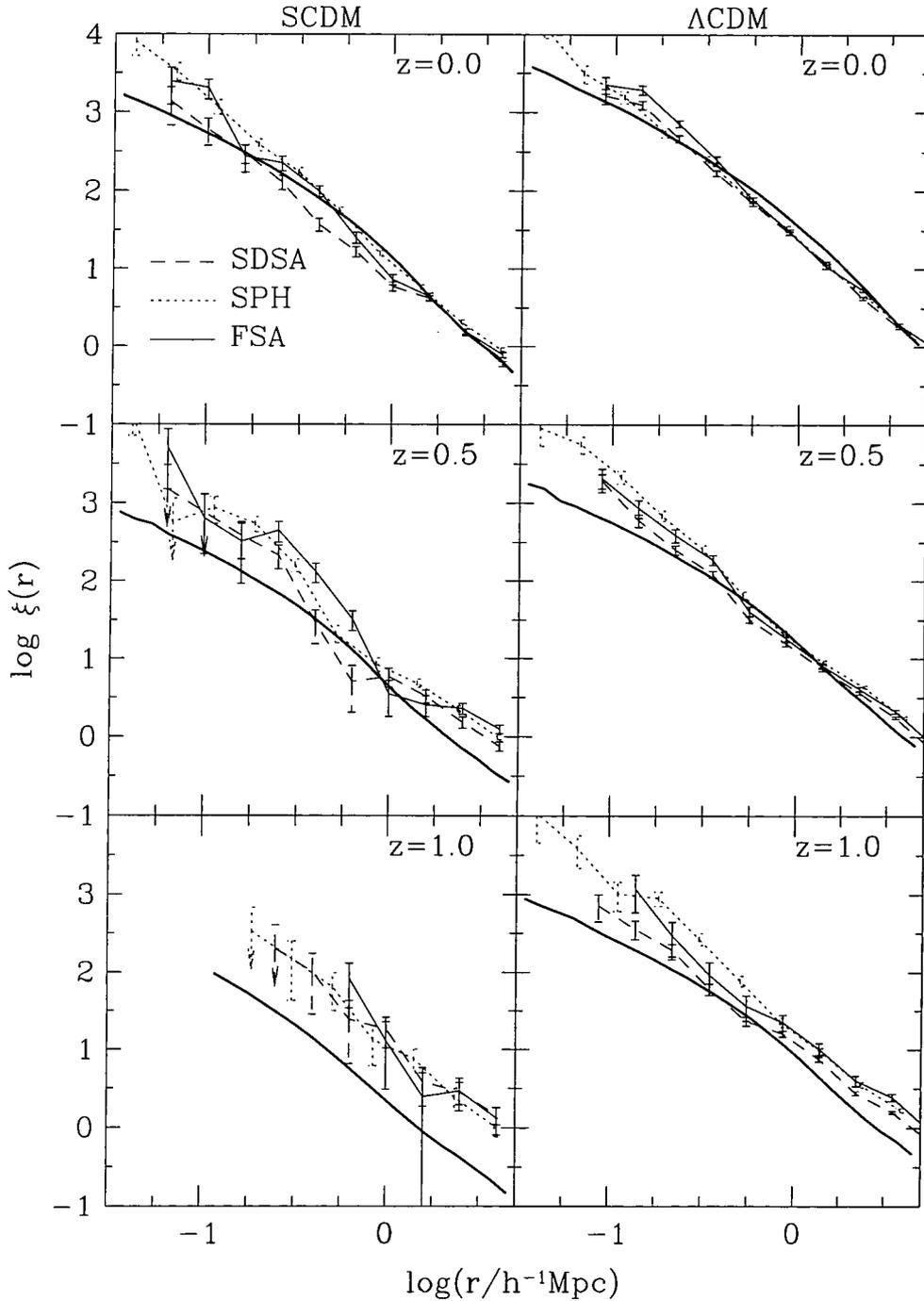


Figure 7.13: Two-point galaxy correlation functions from the SPH, FSA and SDSA models are shown as the dotted, solid and dashed lines respectively. The thick solid line shows the correlation function of dark matter in the simulations. Galaxies from all the models are chosen to have a mass in cold gas plus stars greater than 64 times the SPH gas particle mass. SCDM and Λ CDM cosmologies are shown in left and right-hand panels respectively. Results are shown at $z = 0, 0.5$ and 1 as indicated in each panel. Downwards-pointing arrows are used where the lower errorbar drops below zero.

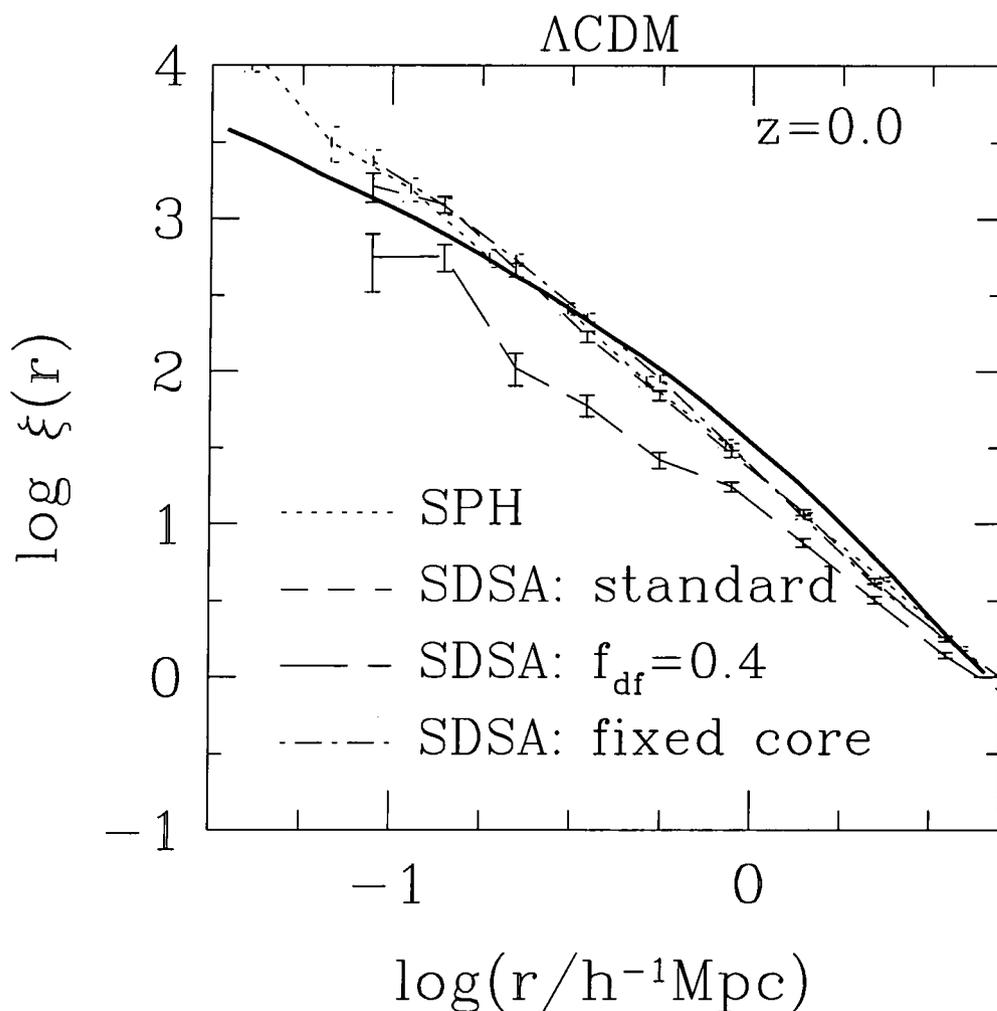


Figure 7.14: The Λ CDM galaxy two-point correlation function in the SPH and SDSA models at $z = 0$. The SPH model is shown by the dotted line, whilst the standard SDSA model is indicated by the short-dashed line. The long-dashed line shows the SDSA model with an enhanced merger rate ($f_{\text{df}} = 0.4$), whilst the dot-dashed line shows the SDSA model with a fixed gas core radius. The heavy solid line shows the correlation function of dark matter.

analytic models. A particularly uncertain component of the semi-analytic treatment is the assumption that gas cools from a quasi-equilibrium state established when the gas is shock-heated to the virial temperature of a halo during collapse. Our comparisons do not test this assumption directly, only its net effect on the amount of gas that cools. Globally, this turns out to be very similar to the SPH result. However, the semi-analytic model tends to produce somewhat less cool gas in massive halos than the SPH simulation, particularly in the Λ CDM cosmology. We stress that due to the limited resolution of our SPH simulations, our conclusions are restricted to massive galaxies, with baryonic mass of $\gtrsim 10^{11} M_{\odot}$. It will be important to check whether our results still hold in higher resolution simulations. In this Chapter we have focussed on statistical properties of the galaxy population. Agreement at this level does not necessarily imply agreement on the properties of galaxies on a case-by-case basis. This should prove an interesting avenue for future work to investigate. Our present results, however, provide useful support for the reliability of current techniques for modelling galaxy formation in a cosmological context.

In this Thesis we have studied galaxy formation and clustering using a detailed and physically realistic model. Our aims were two-fold. Firstly, we have examined in detail two aspects of galaxy formation: the clustering properties of galaxies and the process of reionization. Secondly, we have made initial investigations aimed at determining just how realistic this physical model is.

8.1 Applications of the Model

We have used our semi-analytic model in this Thesis to study the problems of galaxy clustering and the reionization of the intergalactic medium by stars.

8.1.1 Galaxy Clustering

The application of our semi-analytic model to the problem of galaxy clustering in Chapters 3 and 4 has shown that the physics of galaxy formation are capable of creating a scale dependent bias between the distributions of galaxies and mass for pair separations less than about $1h^{-1}\text{Mpc}$ (i.e. on scales comparable to the sizes of dark matter halos). On small scales our model actually predicts an *antibias* (i.e. the galaxies are less strongly clustered than the mass). This bias is successful at explaining the observed clustering pattern of galaxies in a particular cosmology. The same model helps to explain the low observed velocity dispersion of galaxies, predicting a pairwise velocity dispersion $\sim 40\%$ lower than that of the dark matter on Mpc scales and in reasonable agreement with the results of the Las Campanas Redshift Survey. The model can also reproduce clustering trends that depend on the internal properties of galaxies, with elliptical galaxies being more strongly clustered than spiral galaxies and with the very brightest galaxies (those two magnitudes brighter than L_*) having a correlation function roughly three times greater than that of L_* galaxies.

8.1.2 Reionization

In Chapter 5 we studied cosmological reionization and demonstrated that whilst the model of galaxy formation we consider may well be entirely consistent with the Gunn-Peterson constraint there are two large uncertainties which prevent a definitive statement being made, namely the degree of clumping in the IGM and the fraction of ionizing photons that can escape from the galaxies that produce them. In a Λ CDM cosmology and for a physical calculation of the escape fraction, we have shown that reionization occurs by $z \approx 6.1$ if the IGM has uniform density, and by $z \approx 4.5$ for an IGM clumped similarly to the dark matter (in a τ CDM cosmology reionization does not occur until $z \approx 4.1$ and $z \approx 3.8$ for the same two assumptions about gas clumping). Despite these unknowns we have shown that the impact of reionization on the cosmic microwave background can be robustly determined, and should be measurable with, for example, the Atacama Large Millimeter Array (ALMA) in the not too distant future, providing strong constraints on galaxy formation at high redshifts.

8.2 Tests of the Model Ingredients

Semi-analytic modelling possesses a number of unique strengths that make it a particularly powerful technique. Many of these strengths have been highlighted by the work of Chapters 3, 4 and 5. Firstly, as demonstrated by our study of galaxy clustering in this Thesis, the models can be used to interpret observational results in a physical way. The second strength of the technique is the ability to model many aspects of galaxy formation consistently within a single framework. This feature makes such models extremely flexible, allowing them to be applied to a wide range of studies, and also imposes important constraints on the modelling process. Finally, and perhaps most importantly, semi-analytic modelling techniques have demonstrated the importance of the various *physical* processes (i.e. gas cooling, feedback from supernovae, galaxy merging etc.) which determine the properties of galaxies.

8.2.1 The Growth of Galactic Disks

Whilst these strengths and the previous successes of such models give hope that they will have continued success in the future we have attempted in this Thesis to explore the limitations of these models. Only by finding where the models fail can we hope to improve both the model and our understanding of galaxy formation. Specifically, in Chapter 6 we

have confronted one of the fundamental assumptions of such models (the growth of galactic disks by accretion of gas from a surrounding hot halo) with the available observational data. The results show that the actual X-ray emission is roughly an order of magnitude less than theoretically expected. This problem suggests that either the assumption is incorrect or galaxies somehow manage to hide their gas accretion from us. In either case, further study will be crucial to determine just how realistic this fundamental assumption is.

8.2.2 Comparison to Simulations of Gas Hydrodynamics

We have also compared the results of the semi-analytic technique with an alternative approach to galaxy formation, namely direct hydrodynamical simulation, in Chapter 7. Here the results are broadly pleasing (although this is of course a matter of opinion), with the two models producing galaxies with similar properties and similar evolutionary trends. However, the differences between the models are still often larger than the error bars on observational data which they are trying to explain (e.g. the abundance of galaxies of a particular mass can differ by as much as a factor of four between the two models). Such a situation is far from ideal. The study presented in this Thesis has uncovered possible reasons for the discrepancies between the two models, namely that the simple estimates of gas cooling rates in semi-analytic models may be systematically too low.

8.3 Future Prospects

As is the case with any scientific study, the work in this Thesis has answered some important questions, but raised many more. The testing of the assumptions and approximations of the semi-analytic technique should now be repeated in far greater detail, and such testing should be made an ongoing process to ensure that such models remain accurate and their results well understood. Another important advance would be to improve the modelling of star formation and feedback, perhaps adopting a scheme motivated by the physical structure of the interstellar medium (see, for example, Efstathiou 2000). This would allow constraints to be placed on ISM physics from global galaxy formation considerations.

It has also become clear that such models face some strong tests in the near future, particularly from observational data reaching ever higher redshifts. In particular the Next Generation Space Telescope (NGST) promises to probe the very earliest stages of galaxy

formation. Building realistic and robust models to predict the properties of galaxies at $z \sim 10$ and greater will be a key requirement to understanding NGST's results. This is not an easy goal, as at these high redshifts many physical processes previously ignored in such models may well be of great importance.

Within the next five to ten years it is also hoped that experiments such as MAP and Planck will determine the cosmological parameters to high accuracy so these no longer need to be considered as free parameters of the model. A step in this direction has already been made by the Boomerang (de Bernardis et al. 2000) and Maxima experiments (Hanany et al. 2000). The new generation of galaxy surveys, such as 2dF and SDSS, will provide the definitive measurement of key statistics (such as the galaxy luminosity function) at low redshifts. These statistics are presently used to constrain the models, but are quite uncertain observationally, even at just one magnitude below L_* . Once the luminosity function is determined accurately and selection effects understood, all models will have to reproduce these results before they can be considered to be realistic.

Such tests are crucial to allow us to decide which aspects of galaxy formation we understand and what remains unknown. The present situation was nicely summarized by P. J. E. Peebles twenty years ago (Peebles 1980):

The phenomenological basis of physical science can be indirect, and has to be particularly so in the physics of the very early universe. The dread limiting case is a picture of what happened in the remote past which seamlessly agrees with physical theory that is well motivated by the evidence from the laboratory and extragalactic astronomy at low redshifts, but is tested by no empirical evidence except what went into its construction. Would this picture be a part of physical cosmology? Perhaps it is fortunate that there is no need to decide, for the pictures now under construction are far from seamless. Cosmology is in an exciting state because we have a rich and growing list of problems and a growing observational base that may allow us to find a few solutions. The historical record here and in other physical sciences suggests that as the puzzles and conundrums we know about are laid to rest, they will be replaced by still more interesting ones.

What seems clear is that these models, in their present form, are unlikely to describe completely the highly accurate observational data expected in the near future. This is of course the ideal situation for any scientific theory to be in, as it is only then that we can

hope to learn something new.

Appendix A

A Semi-Analytic Model of Galaxy Formation: Supporting Theory

A.1 Fundamental Cosmology

The backbone of the semi-analytic model is the extended Press-Schechter theory, which describes both the abundance and merger histories of dark matter halos. This theory combines two models of perturbation growth in an expanding universe: linear perturbation theory and the spherical collapse model. We will describe these models in Appendix A.1.2 and Appendix A.1.3 respectively. First, however, we describe some of the fundamental cosmology which will be used in the derivations of these models.

A.1.1 The Evolution of a Homogeneous and Isotropic Universe

The gravitational evolution of the universe is described by Einstein's General Theory of Relativity. Under the assumptions that the Universe is both homogeneous and isotropic (the so-called "Cosmological Principle") it can be shown that the space-time metric must have the Friedmann-Robertson-Walker form (e.g. Robertson 1935):

$$ds^2 = c^2 dt^2 - a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2 d\Omega^2 \right], \quad (\text{A.1})$$

where ds is the space-time separation of two events with coordinate separations dt (cosmic time, with c being the speed of light), dr (comoving radial distance) and $d\Omega = \sin\theta d\theta d\phi$ (θ and ϕ being the usual spherical polar coordinates). The expansion factor $a(t)$ describes the scale of the Universe as a function of time and is conventionally set equal to 1 at the present day such that r measures proper distance at the present time. The constant K determines the curvature of the universe,

$$K \begin{cases} > 0; & \text{closed universe} \\ = 0; & \text{flat universe} \\ < 0; & \text{open universe.} \end{cases} \quad (\text{A.2})$$

The evolution of the scale parameter $a(t)$ can be derived from Einstein's field equations, and is specified by the Friedmann cosmological equations (Peebles 1980)

$$\dot{a}^2 + Kc^2 = \frac{8\pi}{3}G\rho(t)a^2 + \frac{\lambda c^2}{3}a^2 \quad (\text{A.3})$$

$$\ddot{a} = -\frac{4\pi}{3}G\left(\rho(t) + 3\frac{p(t)}{c^2}\right)a + \frac{\lambda c^2}{3}a, \quad (\text{A.4})$$

where G is the gravitational constant, λ is the cosmological constant, and $\rho(t)$ and $p(t)$ are the density and pressure respectively of matter and radiation in the Universe. Here we will always be concerned with a dust model in which the density and pressure due to radiation are negligible and in which the pressure due to matter is zero. Then eqns. (A.3) and (A.4) simplify to

$$H^2(t) = H_0^2\left(\Omega_0 a^{-3} + \Lambda_0 + \Omega_K a^{-2}\right), \quad (\text{A.5})$$

and

$$\ddot{a} = -\frac{4\pi}{3}G\rho(t)a, \quad (\text{A.6})$$

where $\Omega_0 = 8\pi G\rho(t_0)/3H_0^2$ (the density parameter), $\Lambda_0 = \lambda c^2/3H_0^2$ (the scaled cosmological constant, hereafter simply cosmological constant), $\Omega_K = -Kc^2/H_0^2$ and $H(t) = \dot{a}(t)/a(t)$ (the Hubble function). Here a subscript 0 denotes the value of a quantity at the present day (time $t = t_0$) and $\dot{}$ denotes differentiation with respect to t . Note that $\Omega_0 + \Lambda_0 + \Omega_K = 1$ from the above when $t = t_0$.

Rewriting eqn. (A.5) in a more convenient form,

$$\dot{a} = H_0\left(\Omega_0 a^{-1} + \Lambda_0 a^2 + \Omega_K\right)^{1/2}. \quad (\text{A.7})$$

Then, for example, when $\Lambda_0 = \Omega_K = 0$ (and therefore $\Omega_0 = 1$) we have

$$\dot{a} = H_0 a^{-1/2}, \quad (\text{A.8})$$

and so

$$\int_0^a a'^{1/2} da' = H_0 \int_0^t dt', \quad (\text{A.9})$$

and hence

$$a(t) = \left(\frac{3}{2}H_0 t\right)^{2/3}. \quad (\text{A.10})$$

For other cosmologies $a(t)$ admits the following parametric solutions (e.g. Peebles 1980, Coles & Lucchin 1995):

$$\begin{aligned} \Omega_0 < 1, \Lambda_0 = 0 & \begin{cases} a(\eta) = \frac{\Omega_0}{2(1-\Omega_0)} (\cosh \eta - 1) \\ t(\eta) = \frac{1}{2H_0} \frac{\Omega_0}{(1-\Omega_0)^{3/2}} (\sinh \eta - \eta), \end{cases} \\ \Omega_0 > 1, \Lambda_0 = 0 & \begin{cases} a(\theta) = \frac{\Omega_0}{2(\Omega_0-1)} (1 - \cos \theta) \\ t(\theta) = \frac{1}{2H_0} \frac{\Omega_0}{(\Omega_0-1)^{3/2}} (\theta - \sin \theta), \end{cases} \\ \Omega_0 + \Lambda_0 = 1, & a(t) = \left(\frac{\Omega_0}{1-\Omega_0} \right)^{1/3} \sinh^{2/3} \left[\frac{3H_0 t}{2} (1 - \Omega_0)^{1/2} \right]. \end{aligned} \quad (\text{A.11})$$

A.1.2 Linear Growth of Perturbations

Here we derive the equations governing the growth of perturbations in the Universe in the linear regime. In Newtonian theory the evolution of the cosmological fluid is governed by the continuity (mass conservation) equation, Euler's equation and Poisson's equation, which are, respectively,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (\text{A.12})$$

$$\rho \frac{D\mathbf{u}}{Dt} = -\nabla p - \rho \nabla \Phi, \quad (\text{A.13})$$

$$\nabla^2 \Phi = 4\pi G \rho, \quad (\text{A.14})$$

where \mathbf{u} is the velocity of the fluid, Φ is the gravitational potential, ∇ denotes differentiation with respect to position \mathbf{r} and D/Dt denotes the Lagrangian derivative $\partial/\partial t + \mathbf{u} \cdot \nabla$. These equations apply in physical coordinates, but for convenience we choose to express them in terms of comoving position, $\mathbf{x} = \mathbf{r}/a$, conformal time $d\tau = dt/a$ and peculiar velocity $\mathbf{v} = a d\mathbf{x}/dt = \mathbf{u} - \mathbf{x} da/dt$ (here $\mathbf{x} da/dt = (\mathbf{x}/a) da/d\tau$ is the velocity due to the expansion of the Universe) and a density contrast $\delta = (\rho - \bar{\rho})/\bar{\rho}$ (where $\bar{\rho}$ is the mean density of the Universe). Eqn. (A.12) then becomes

$$\frac{\partial}{\partial \tau} (\bar{\rho}[1 + \delta]) + \nabla \cdot (\bar{\rho}[1 + \delta][\mathbf{v} + \mathbf{x} da/dt]) = 0, \quad (\text{A.15})$$

where now $\dot{}$ and ∇ are differentiation with respect to τ and \mathbf{x} respectively. With no perturbations (i.e. $\delta = 0$, $v = 0$ and $\Phi = 0$) this reduces to

$$\dot{\bar{\rho}} + 3\bar{\rho} da/dt = 0. \quad (\text{A.16})$$

Subtracting eqn. (A.16) from eqn. (A.15) then gives

$$\dot{\delta} + \delta \left(\frac{\dot{\bar{\rho}}}{\bar{\rho}} + 3 \frac{da}{dt} \right) + \nabla \cdot [(1 + \delta)\mathbf{v}] = 0. \quad (\text{A.17})$$

However, $\dot{\bar{\rho}}/\bar{\rho} + 3da/dt = 0$ from eqn. (A.16), so this becomes

$$\dot{\delta} + \nabla[(1 + \delta)\mathbf{v}] = 0. \quad (\text{A.18})$$

When moving to comoving coordinates we use the fact that the partial derivative of $\mathbf{u}[t, \mathbf{r}]$ at fixed \mathbf{r} is given by

$$\left(\frac{\partial \mathbf{u}}{\partial t}\right)_{\mathbf{r}} = a^{-1} \left(\frac{\partial \mathbf{u}}{\partial \tau}\right)_{\mathbf{x}} - a^{-1} \frac{\partial \mathbf{r}}{\partial \tau} \cdot \nabla_{\mathbf{r}} \mathbf{u}, \quad (\text{A.19})$$

and so the operator $D/Dt = \partial/\partial t + \mathbf{u} \cdot \nabla$ in physical coordinates becomes

$$a^{-1} \left[\frac{\partial}{\partial \tau} - \mathbf{v} \cdot \nabla \right] \quad (\text{A.20})$$

in comoving coordinates. Hence eqn. (A.13) becomes

$$a^{-1} \bar{\rho}(1 + \delta) \left[\frac{\partial}{\partial \tau} + \mathbf{v} \cdot \nabla \right] (\mathbf{v} + \mathbf{x}\dot{a}/a) = -a^{-1} \nabla p - a^{-1} \bar{\rho}(1 + \delta) \nabla \Phi. \quad (\text{A.21})$$

The unperturbed solution is then

$$\frac{\partial}{\partial \tau} \left(\mathbf{x} \frac{\dot{a}}{a} \right) = 0. \quad (\text{A.22})$$

Subtracting eqn. (A.22) from (A.21) leaves

$$(1 + \delta)(\dot{\mathbf{v}} + \mathbf{v} \cdot \nabla \mathbf{v} + \mathbf{v}\dot{a}/a) = -\nabla p/\bar{\rho} - (1 + \delta)\nabla \Phi, \quad (\text{A.23})$$

where we have used the identity $\mathbf{v} \cdot \nabla \mathbf{x} = \mathbf{v}$. Since $\delta \ll 1$ in the linear regime this reduces to

$$\dot{\mathbf{v}} + \mathbf{v} \cdot \nabla \mathbf{v} + \mathbf{v}\dot{a}/a = -\nabla p/\bar{\rho} - \nabla \Phi. \quad (\text{A.24})$$

Finally, eqn. (A.14) becomes

$$\nabla^2 \Phi = 4\pi G \bar{\rho} a^2 \delta. \quad (\text{A.25})$$

We now ignore any terms above first order in eqns. (A.18) and (A.24) to obtain

$$\begin{aligned} \dot{\delta} + \nabla \cdot \mathbf{v} &= 0 \\ \Rightarrow \ddot{\delta} + \nabla \cdot \dot{\mathbf{v}} &= 0 \end{aligned} \quad (\text{A.26})$$

and

$$\begin{aligned} \dot{\mathbf{v}} + \frac{\dot{a}}{a} \mathbf{v} &= -\frac{\nabla p}{\bar{\rho}} - \nabla \Phi \\ \Rightarrow \nabla \cdot \dot{\mathbf{v}} + \frac{\dot{a}}{a} \nabla \cdot \mathbf{v} &= -\frac{\nabla^2 p}{\bar{\rho}} - \nabla^2 \Phi. \end{aligned} \quad (\text{A.27})$$

Eliminating $\nabla \cdot \dot{\mathbf{v}}$ between eqns. (A.26) and (A.27), $\nabla \cdot \mathbf{v}$ with eqn. (A.12) and eliminating Φ using Poisson's equation (A.14) we find

$$\ddot{\delta} + \frac{\dot{a}}{a} \dot{\delta} = \frac{\nabla^2 P}{\bar{\rho}} + \frac{3}{2} \Omega(\tau) \left(\frac{\dot{a}}{a}\right)^2 \delta. \quad (\text{A.28})$$

For dust models the pressure is zero, so the above equation involves no spatial derivatives and hence can be solved by a separation of variables yielding solutions of the form $\delta(\mathbf{x}, \tau) = A(\mathbf{x})f(\tau)$. The general solution to this second-order, linear differential equation is then of the form

$$\delta(\mathbf{x}, \tau) = A(\mathbf{x})f_1(\tau) + B(\mathbf{x})f_2(\tau). \quad (\text{A.29})$$

For example, in an Einstein-de Sitter cosmology we have $a \propto \tau^2 \propto t^{2/3}$ and $\Omega(\tau) = 1$, hence

$$\ddot{\delta} + 2\dot{\delta}/\tau - 6\delta/\tau^2 = 0. \quad (\text{A.30})$$

Solutions are of the form $f(\tau) \propto \tau^\alpha$ where $\alpha = 2$ or -3 for the growing or decaying mode solution respectively. We will be interested here in the growing mode solutions. In the Einstein-de Sitter universe therefore perturbations grow as $\delta \propto a$ in the linear regime.

For other cosmologies the equivalent relations are (Peebles 1980, Peebles 1984):

$$\begin{aligned} \Omega_0 < 1, \Lambda_0 = 0; & \quad \delta \propto 3 \sinh \eta (\sinh \eta - \eta) / (\cosh \eta - 1)^2 - 2 \\ \Omega_0 > 1, \Lambda_0 = 0; & \quad \left\{ \begin{aligned} \delta \propto \frac{6}{1 - \cos \theta} \left[1 - \left(\frac{1 + \cos \theta}{1 - \cos \theta} \right)^{1/2} \tan^{-1} \left(\frac{1 - \cos \theta}{1 + \cos \theta} \right)^{1/2} \right] - 1, 0 < \theta < \pi \\ \delta \propto \frac{6}{1 - \cos \theta} \left[1 - \left(\frac{1 + \cos \theta}{1 - \cos \theta} \right)^{1/2} \left\{ \tan^{-1} \left(\frac{1 - \cos \theta}{1 + \cos \theta} \right)^{1/2} - \pi \right\} \right] - 1, \pi < \theta < 2\pi \end{aligned} \right. \\ \Omega_0 + \Lambda_0 = 1; & \quad \delta \propto (x / (\Omega_0^{-1} - 1)^{1/3}) (x^3 + 1)^{1/2} \int_0^1 (1 + x^3 y^{6/5})^{-3/2} dy, \\ & \quad a(x) = (\Omega_0^{-1} - 1)^{-1/3} x. \end{aligned} \quad (\text{A.31})$$

The growth factor, $D(a)$, is then defined as the ratio of the growing mode overdensity at expansion factor a to its amplitude at the present day ($a = 1$).

A.1.3 Spherical Top-Hat Collapse

Having determined the growth of perturbations in the linear regime we will now describe the simplest possible model for the non-linear growth of a density perturbation. We consider a spherical region in an otherwise homogeneous universe of mean density $\bar{\rho}$ which has radius $r(t)$ and overdensity $\delta(t)$. In Newtonian physics the smoothly distributed matter outside of this region would have no influence on the evolution of the perturbation. A similar result, known as Birkhoff's theorem (Birkhoff 1923, p. 253), holds in General

Relativity. Therefore we can apply the standard Newtonian equations of motion to a test particle at the edge of the spherical region:

$$\frac{d^2r}{dt^2} = -\frac{GM}{r^2} = -\frac{4\pi G}{3}\bar{\rho}(t)(1+\delta)r, \quad (\text{A.32})$$

where the mass of the perturbation is $M = 4\pi\bar{\rho}(1+\delta)r^3/3$. Note that the cosmological scale factor, $a(t)$, evolves as (eqn. A.6)

$$\frac{d^2a}{dt^2} = -\frac{4\pi G}{3}\bar{\rho}(t)a, \quad (\text{A.33})$$

which has the same form as the previous equation. Hence the perturbation evolves in the same way as a universe of different density, but the same initial time and expansion rate. Multiplying eqn. (A.32) by dr/dt and integrating we find the energy equation

$$\int \frac{dr}{dt} \frac{d^2r}{dt^2} dt = -GM \int r^{-2} \frac{dr}{dt} dt$$

$$\frac{1}{2} \left(\frac{dr}{dt} \right)^2 - \frac{GM}{r} = E, \quad (\text{A.34})$$

where E is the constant of integration. Note that the first term on the left is the kinetic energy per unit mass whilst the second term is the gravitational potential energy per unit mass. Hence the constant E is just the specific energy of the perturbation. Since we want to consider a perturbation which will collapse we choose $E < 0$ (i.e. a self-bound perturbation). Then

$$\frac{dr}{dt} = \left(\frac{2GM}{r} - 2|E| \right)^{1/2}, \quad (\text{A.35})$$

or, equivalently,

$$\frac{r^{1/2}dr}{(GM - |E|r)^{1/2}} = \sqrt{2}dt. \quad (\text{A.36})$$

To solve this differential equation let $r = GM(1 - \cos\theta)/2|E| = GM \cos^2(\theta/2)/|E|$ and hence $dr = GM \cos(\theta/2) \sin(\theta/2)d\theta$. Eqn. (A.36) then becomes

$$\frac{GM \sin(\theta/2) \cos(\theta/2)d\theta}{|E|^{3/2}(1 - \cos^2(\theta/2))^{1/2}} = \sqrt{2}dt, \quad (\text{A.37})$$

which simplifies to give

$$\frac{GM}{2|E|^{3/2}}(1 - \cos\theta)d\theta = \sqrt{2}dt, \quad (\text{A.38})$$

and hence

$$\frac{GM}{(2|E|)^{3/2}}(\theta - \sin\theta) = t - t_0. \quad (\text{A.39})$$

Note that this solution is essentially identical to the expansion factor solution for a closed universe given in eqn. (A.11).

We can therefore write

$$r = r_m(1 - \cos \theta)/2, \tag{A.40}$$

where $r_m = GM/|E|$ is the maximum radius of the perturbation (occurring at $\theta = \pi$). When $t = 0, r = 0$ (i.e. $\theta = 0$) so evidently $t_0 = 0$ and we can write

$$t_m(\theta - \sin \theta)/\pi = t, \tag{A.41}$$

where $t_m = GM/(2|E|)^{3/2}$ is the time at which the perturbation reaches its maximum radius.

For later purposes we wish to match the early evolution of this perturbation to the equivalent linear theory solution. A relation between r and t when θ is small can be derived by expanding eqns. (A.40) and (A.41) as a series in θ :

$$\frac{r}{r_m} = \frac{\theta^2}{4} - \frac{\theta^4}{48} + \dots; \quad \frac{t}{t_m} = \frac{1}{\pi} \left(\frac{\theta^3}{6} - \frac{\theta^5}{120} + \dots \right). \tag{A.42}$$

Solving the second of these for θ to first order gives

$$\theta \approx \left(\frac{6\pi t}{t_m} \right)^{1/3}. \tag{A.43}$$

Using this result in the second order expression $\theta^3 = 6\pi t/t_m + \theta^5/20$ then gives

$$\begin{aligned} \theta^3 &= \frac{6\pi t}{t_m} \left(1 + \frac{1}{20} \left[\frac{6\pi t}{t_m} \right]^{2/3} \right) \\ \theta^2 &= \left(\frac{6\pi t}{t_m} \right)^{2/3} \left(1 + \frac{1}{30} \left[\frac{6\pi t}{t_m} \right]^{2/3} \right), \end{aligned} \tag{A.44}$$

where the second bracket has been expanded to first order in $(6\pi t/t_m)^{2/3}$. Substituting this last relation into the expansion for r gives

$$\frac{r}{r_m} = \frac{1}{4} \left(\frac{6\pi t}{t_m} \right)^{2/3} \left(1 + \frac{1}{30} \left[\frac{6\pi t}{t_m} \right]^{2/3} - \frac{1}{12} \left[\frac{6\pi t}{t_m} \right]^{2/3} \left[1 + \frac{1}{30} \left\{ \frac{6\pi t}{t_m} \right\}^{2/3} \right]^2 \right). \tag{A.45}$$

Keeping only terms of order $(6\pi t/t_m)^{2/3}$ in the brackets gives

$$\frac{r}{r_m} = \frac{1}{4} \left(\frac{6\pi t}{t_m} \right)^{2/3} \left(1 - \frac{1}{20} \left[\frac{6\pi t}{t_m} \right]^{2/3} \right). \tag{A.46}$$

At some early time $t_i \ll t_m$ (such that $\delta_i \ll 1$) let the perturbation radius be r_i . Since the overdensity of the perturbation at this time was small we have

$$M = 4\pi \bar{\rho}_i r_i^3/3. \tag{A.47}$$

At some later time t the mass is unchanged and hence the density of the sphere is

$$\rho(t) = \frac{3M}{4\pi r(t)^3} = \bar{\rho}_i \left(\frac{r_i^3}{r(t)} \right)^3. \tag{A.48}$$

Dividing by the mean density at time t gives

$$1 + \delta(t) = \frac{\bar{\rho}_i}{\bar{\rho}(t)} \frac{r_i^3}{r^3(t)}. \quad (\text{A.49})$$

In an Einstein-de Sitter universe the mean density varies as $\bar{\rho} \propto a^{-3} \propto t^{-2}$ from eqn. (A.10), so using eqn. (A.46) to eliminate r_i and $r(t)$ (noting that $r_i/r_m \approx (6\pi t_i/t_m)^{2/3}/4$ since $t_i \ll t_m$) gives

$$\begin{aligned} 1 + \delta(t) &= \left(1 - \frac{1}{20} \left[\frac{6\pi t}{t_m}\right]^{2/3}\right)^{-3}, \\ \delta(t) &\approx \frac{3}{20} \left(\frac{6\pi t}{t_m}\right)^{2/3}. \end{aligned} \quad (\text{A.50})$$

Since from eqns. (A.40) and (A.41) the sphere collapse to a singularity at time $t = 2t_m$ the extrapolated linear overdensity at collapse is

$$\delta_c = \delta(2t_m) = \frac{3}{20} (12\pi)^{2/3} \approx 1.686. \quad (\text{A.51})$$

We will also be interested in the mean density of the collapsed region after it has reached virial equilibrium. At its radius of maximum expansion, r_m , the perturbation has zero velocity, so from eqn. (A.34) its energy is purely gravitational and equal to

$$E = -\frac{GM}{r_m}. \quad (\text{A.52})$$

After virialisation $2T + W = 0$, where T is the internal energy of the perturbation and W its gravitational energy. Hence, $E = T + W = W/2$ and so

$$\begin{aligned} -\frac{GM}{2r_H} &= -\frac{GM}{r_m}, \\ r_H &= r_m/2, \end{aligned} \quad (\text{A.53})$$

assuming that energy is conserved in the collapse. After virialisation the collapsed region has a radius equal to half its radius at maximum expansion. The ratio of the mean density of the collapsed region to the critical density, $\rho_c(2t_m) = \bar{\rho}(2t_m)\Omega(2t_m)^{-1}$, is

$$\Delta_{\text{vir}} = \frac{\Omega(2t_m)}{\bar{\rho}(2t_m)} \frac{8 \times 3M}{4\pi r_m^3} = 8\Omega(2t_m) \frac{\bar{\rho}_i}{\bar{\rho}(2t_m)} \left(\frac{r_i}{r_m}\right)^3, \quad (\text{A.54})$$

where we have used eqn. (A.47). Using eqn. (A.46) to leading order and the facts that for an Einstein-de Sitter cosmology $\bar{\rho}(t) \propto t^{-2}$ and $\Omega(\tau) = 1$ this becomes

$$\begin{aligned} \Delta_{\text{vir}} &= 8 \left(\frac{t_i}{2t_m}\right)^{-2} \left[\frac{1}{4} \left(\frac{6\pi t_i}{t_m}\right)^{2/3}\right]^3 \\ \Delta_{\text{vir}} &= 18\pi^2 \approx 178. \end{aligned} \quad (\text{A.55})$$

In other cosmologies both δ_c and Δ_{vir} depend upon redshift. For an $\Omega_0 < 1, \Lambda_0 = 0$ universe the relevant expressions have been derived by Lacey & Cole (1993):

$$\delta_c(\eta) = \frac{3}{2} \left(\frac{3 \sinh \eta (\sinh \eta - \eta)}{(\cosh \eta - 1)^2} - 2 \right) \left(1 + \left[\frac{2\pi}{\sinh \eta - \eta} \right]^{2/3} \right), \quad (\text{A.56})$$

$$\Delta_{\text{vir}}(\eta) = 8\pi^2 \left[\frac{(\cosh \eta - 1)^2}{\sinh \eta (\sinh \eta - \eta)} \right]^2. \quad (\text{A.57})$$

In $\Omega_0 + \Lambda_0 = 1$ cosmologies solutions have been found by Eke, Cole & Frenk (1996), but these have no simple analytic form and must be calculated numerically.

A.2 Halo Angular Momentum

The angular momentum within the virial radius of a halo in which V_{rot} is constant with radius is given by

$$\begin{aligned} J(r_{\text{H}}) &= \int_0^{r_{\text{H}}} \int_0^\pi \int_0^{2\pi} V_{\text{rot}} r' \sin \theta \rho(r') r'^2 \sin \theta d\phi d\theta dr' \\ &= \pi^2 V_{\text{rot}} \int_0^{r_{\text{H}}} r'^3 \rho(r') dr'. \end{aligned} \quad (\text{A.58})$$

The total energy, E_{H} , of the halo is the sum of kinetic and potential energies, T_{H} and W_{H} respectively. The gravitational self-binding energy of the material within r_{H} is given by (Binney & Tremaine 1987)

$$W(r_{\text{H}}) = \frac{4\pi}{2} \int_0^\infty \phi(r') \rho(r') r'^2 dr', \quad (\text{A.59})$$

where $\rho(r)$ and $\phi(r)$ are the density and gravitational potential respectively due to the matter within r_{H} only (i.e. we ignore any exterior matter as we are interested only in the self-binding energy of the sphere).

Using Poisson's equation, $\nabla^2 \phi(r) = 4\pi G \rho(r)$, and $\phi \nabla^2 \phi = \nabla \cdot \phi \nabla \phi - |\nabla \phi|^2$ this becomes

$$W(r_{\text{H}}) = -\frac{1}{2G} \left[\int_0^\infty |\nabla \phi(r')|^2 r'^2 dr' - \int_0^\infty (\nabla \cdot \phi \nabla \phi) r'^2 dr' \right]. \quad (\text{A.60})$$

By use of the divergence theorem and the fact that the potential and its derivative fall to zero at infinity the second term in the above equation vanishes, leaving

$$W(r_{\text{H}}) = -\frac{1}{2G} \int_0^\infty |\nabla \phi(r')|^2 r'^2 dr'. \quad (\text{A.61})$$

Finally using the fact that $|\nabla \phi(r)| = GM(r)/r^2$ (and noting that $M(r)$ is constant for $r > r_{\text{H}}$) we find

$$W(r_{\text{H}}) = -\frac{G}{2} \left[\int_0^{r_{\text{H}}} \frac{M(r')^2}{r'^2} dr' + \frac{M^2(r_{\text{H}})}{r_{\text{H}}} \right]. \quad (\text{A.62})$$

The kinetic energy of material inside the virial radius is

$$T(r_H) = 2\pi \int_0^{r_H} v^2(r') \rho(r') r'^2 dr', \quad (\text{A.63})$$

where $v(r)$ is the three dimensional velocity dispersion of dark matter at radius r . In terms of the mean one-dimensional velocity dispersion, $\sigma(r) = v(r)/\sqrt{3}$, we have

$$T(r_H) = 6\pi \int_0^{r_H} \sigma^2(r') \rho(r') r'^2 dr'. \quad (\text{A.64})$$

Under the assumption that the velocity dispersion is isotropic the Jeans equation, $d(\rho\sigma^2)/dr = -\rho GM(r)/r^2$, can be used, in which case integrating the above equation by parts gives

$$T(r_H) = 2\pi \left[r_H^3 \rho(r_H) \sigma^2(r_H) + \int_0^{r_H} GM(r') \rho(r') r' dr' \right] \quad (\text{A.65})$$

providing that $r^3\rho(r)\sigma^2(r)$ vanishes as r tends to 0. For the standard choice of NFW halo profiles the Jean's equation is integrated to $r = \infty$ to obtain $\sigma(r)$, assuming that the NFW profile and hydrostatic equilibrium apply at all r . This is an approximation as beyond the virial radius material will still be infalling. However, $\sigma(r)$ derived in this way has been found to be in good agreement with N-body simulations (Cole & Lacey 1996). It should also be noted that these definitions of W and T do not satisfy $2T(r_H) + W(r_H) = 0$ (the value being slightly greater than unity). If r_H were extended to infinity then the virial theorem would be satisfied exactly by the NFW profile. A similar behaviour was found in N-body simulations by Cole & Lacey (1996). The above definitions are consistent with their definition of λ_H .

Combining eqns (A.58), (A.62) and (A.65) with eqn. (2.22) we find

$$V_{\text{rot}} = \lambda_H \frac{GM_H^{5/2}}{\pi^2} \left(\int_0^{r_H} r'^3 \rho(r') dr' \right)^{-1} \left(2\pi \left[r_H^3 \rho(r_H) \sigma^2(r_H) + \int_0^{r_H} GM(r') \rho(r') r' dr' \right] - \frac{G}{2} \left[\int_0^{r_H} \frac{M(r')^2}{r'^2} dr' + \frac{M^2(r_H)}{r_H} \right] \right)^{-1}, \quad (\text{A.66})$$

or, more simply,

$$V_{\text{rot}} = A(a_{\text{NFW}}) \lambda_H V_H, \quad (\text{A.67})$$

where

$$A(a_{\text{NFW}}) = \frac{G^{1/2} M_H^2 r_H^{1/2}}{\pi^2} \times \left(\int_0^{r_H} r'^3 \rho(r') dr' \right)^{-1} \left(2\pi \left[r_H^3 \rho(r_H) \sigma^2(r_H) + \int_0^{r_H} GM(r') \rho(r') r' dr' \right] - \frac{G}{2} \left[\int_0^{r_H} \frac{M(r')^2}{r'^2} dr' + \frac{M^2(r_H)}{r_H} \right] \right)^{-1}, \quad (\text{A.68})$$

and $V_H = \sqrt{GM_H/r_H}$. The value of $A(a_{\text{NFW}})$ depends only rather weakly on a_{NFW} as described by Cole et al. (2000).

A similar relation holds for the case of a singular isothermal sphere halo potential, although here a slightly different approach must be adopted. If the kinetic energy is calculated by using the Jeans equation to derive $\sigma(r)$, then $2T_H(r_H)/|W_H(r_H)|$ is found to be significantly greater than unity. The large discrepancy arises because the density profile is extrapolated beyond the virial radius. This works reasonably well for an NFW profile, whose mass diverges only logarithmically with r (and for which the gravitational potential is a declining function of radius for large r), but fails for the isothermal sphere model where the mass does not rapidly converge ($M(r) \propto r$). For such profiles we adopt the approach of Mo, Mao & White (1998) and evaluate the binding energy using eqn. (A.62), but then assume the virial relation $2T_H(r_H)/|W_H(r_{\text{vir}})| = 1$ in order to find the kinetic and total energies. Using the relations $M(r) = V_H^2 r/G$ and $\rho(r) = (M_H/4\pi r_H^3)(r_H/r)^2$, which are valid for the isothermal halo, the binding energy can be found from eqn. (A.62) to be

$$W(r_H) = -\frac{GM_H^2}{2r_H} \left[1 + \int_0^{r_H} \frac{dr}{r_H} \right] = -M_H V_H^2. \quad (\text{A.69})$$

Thus the total energy is

$$E_H(r_H) = -M_H V_H^2/2. \quad (\text{A.70})$$

The angular momentum of the halo can be evaluated from eqn. (A.58), and is

$$J_H(r_H) = \frac{\pi}{4} V_{\text{rot}} r_H \int_0^{r_H} \frac{r dr}{r_H^2} = \frac{\pi}{8} V_{\text{rot}} r_H M_H. \quad (\text{A.71})$$

Inserting eqns. (A.70) and (A.71) into eqn. (2.22) results in the expression

$$V_{\text{rot}} = A \lambda_H V_H, \quad (\text{A.72})$$

where $A = 8\sqrt{2}/\pi$. For an isothermal sphere with a constant density core the value of A falls slowly as the core radius is increased. The value of A in this case is given by Cole et al. (2000).

A.3 Gas Equations

Under the assumption that both \dot{M}_{cool} (the rate of accretion of gas by a galaxy from its surrounding hot halo) and Z_{hot} (the metallicity of that accreted gas) are approximately constant over a short time step the eqns. (2.30) to (2.34) can be solved straightforwardly as follows. In time t since the beginning of a time step the mass of gas accreted is

$$\Delta M_{\text{acc}} = \dot{M}_{\text{cool}} t. \quad (\text{A.73})$$

Using eqn. (2.26) in eqn. (2.30) gives

$$\begin{aligned}\dot{M}_* &= (1-R)M_{\text{cold}}/\tau_* \\ \ddot{M}_* &= (1-R)\dot{M}_{\text{cold}}/\tau_* = (1-R)[\dot{M}_{\text{cool}} - (1-R+\beta)\psi]/\tau_*.\end{aligned}\quad (\text{A.74})$$

Eliminating ψ from this last expression using eqn. (2.30) gives the second-order differential equation

$$\ddot{M}_* + \frac{\dot{M}_*}{\tau_{\text{eff}}} = \frac{1-R}{1-R+\beta} \frac{\dot{M}_{\text{cool}}}{\tau_{\text{eff}}}, \quad (\text{A.75})$$

where $\tau_{\text{eff}} = \tau_*/(1-R+\beta)$. This has a solution of the form $M_* = M_*^0 + \Delta M_*$, where M_*^0 is the initial mass of stars and the increase in the mass of long-lived stars, ΔM_* is

$$\Delta M_* = M_{\text{cold}}^0 \frac{1-R}{1-R+\beta} \left[1 - \exp\left(-\frac{t}{\tau_{\text{eff}}}\right) \right] - \dot{M}_{\text{cool}} \tau_{\text{eff}} \frac{1-R}{1-R+\beta} \left[1 - \frac{t}{\tau_{\text{eff}}} - \exp\left(-\frac{t}{\tau_{\text{eff}}}\right) \right], \quad (\text{A.76})$$

where we have required $\dot{M}_* = (1-R)M_{\text{cold}}^0/\tau_*$ at $t = 0$, with M_{cold}^0 being the mass of cold gas at the start of the timestep, in accordance with eqn. (2.30). The changes in cold and hot gas masses can be found in terms of these quantities as follows:

$$\begin{aligned}\dot{M}_{\text{cold}} &= \dot{M}_{\text{cool}} - \frac{1-R+\beta}{1-R} \dot{M}_* \\ \Rightarrow \Delta M_{\text{cold}} &= \Delta M_{\text{acc}} - \frac{1-R+\beta}{1-R} \Delta M_*,\end{aligned}\quad (\text{A.77})$$

and

$$\begin{aligned}\dot{M}_{\text{hot}} &= -\dot{M}_{\text{cool}} + \frac{\beta}{1-R} \dot{M}_* \\ \Rightarrow \Delta M_{\text{hot}} &= -\Delta M_{\text{acc}} + \frac{\beta}{1-R} \Delta M_*.\end{aligned}\quad (\text{A.78})$$

The corresponding change in the mass of metals can be found similarly. Substituting eqn. (2.27) into eqn. (2.34) we find

$$\dot{M}_{\text{cold}}^Z = \dot{M}_{\text{cool}} Z_{\text{hot}} + \frac{p(1-e)}{1-R} \dot{M}_* - \frac{1-R+\beta}{1-R} \dot{M}_*^Z \quad (\text{A.79})$$

$$\Rightarrow \Delta M_{\text{cold}}^Z = \Delta M_{\text{acc}}^Z + \frac{(1-e)p}{1-R} \Delta M_* - \frac{1-R+\beta}{1-R} \Delta M_*^Z, \quad (\text{A.80})$$

where $\Delta M_{\text{acc}}^Z = \dot{M}_{\text{cool}} Z_{\text{hot}} t$. For the metals in the hot gas using eqn. (2.27) in eqn. (2.34) gives

$$\dot{M}_{\text{hot}}^Z = -\dot{M}_{\text{cool}} Z_{\text{hot}} + \frac{pe}{1-R} \dot{M}_* + \frac{\beta}{1-R} \dot{M}_*^Z \quad (\text{A.81})$$

$$\Rightarrow \Delta M_{\text{hot}}^Z = -\Delta M_{\text{acc}}^Z + \frac{ep}{1-R} \Delta M_* + \frac{\beta}{1-R} \Delta M_*^Z. \quad (\text{A.82})$$

The change in the mass of metals in stars can be found by substituting eqn. (2.27) into eqn. (2.33) and differentiating with respect to time to obtain

$$\ddot{M}_*^Z = (1 - R)\dot{M}_{\text{cool}}^Z/\tau_*. \quad (\text{A.83})$$

Eliminating \dot{M}_{cool}^Z from this equation using eqn. (A.80) yields

$$\begin{aligned} \ddot{M}_*^Z + \frac{\dot{M}_*^Z}{\tau_{\text{eff}}} &= \frac{1 - R}{1 - R + \beta} \left(\frac{\dot{M}_{\text{cool}} Z_{\text{hot}}}{\tau_{\text{eff}}} + \frac{p(1 - e)}{1 - R + \beta} \frac{\dot{M}_{\text{cool}}}{\tau_{\text{eff}}} \right) \\ &+ \frac{p(1 - e)}{1 - R + \beta} \frac{1 - R}{1 - R + \beta} \left(\frac{M_{\text{cool}}^0}{\tau_{\text{eff}}^2} - \frac{\dot{M}_{\text{cool}}}{\tau_{\text{eff}}} \right) \exp\left(-\frac{t}{\tau_{\text{eff}}}\right), \end{aligned} \quad (\text{A.84})$$

which, after requiring that $\dot{M}_*^{Z0} = (1 - R)Z_{\text{cold}}^0 M_{\text{cool}}^0/\tau_*$ from eqn. (2.33), has the solution

$$\begin{aligned} \Delta M_*^Z &= \frac{1 - R}{1 - R + \beta} \left(M_{\text{cool}}^{Z0} \left[1 - \exp\left(-\frac{t}{\tau_{\text{eff}}}\right) \right] - \dot{M}_{\text{cool}} \tau_{\text{eff}} Z_{\text{hot}} \left[1 - \frac{t}{\tau_{\text{eff}}} - \exp\left(-\frac{t}{\tau_{\text{eff}}}\right) \right] \right) \\ &+ \frac{(1 - e)p}{1 - R + \beta} \left\{ M_{\text{cool}}^0 \left[1 - \left(1 + \frac{t}{\tau_{\text{eff}}} \right) \exp\left(-\frac{t}{\tau_{\text{eff}}}\right) \right] \right. \\ &\left. - \dot{M}_{\text{cool}} \tau_{\text{eff}} \left[2 - \frac{t}{\tau_{\text{eff}}} - \left(2 + \frac{t}{\tau_{\text{eff}}} \right) \exp\left(-\frac{t}{\tau_{\text{eff}}}\right) \right] \right\}. \end{aligned} \quad (\text{A.85})$$

For the case of no cooling gas supply, $\dot{M}_{\text{cool}} = 0$, the above equations show that when $t \gg \tau_{\text{eff}}$ the mean metallicity of stars that have formed is

$$Z_* = \frac{\Delta M_*^Z}{\Delta M_*} = Z_{\text{cold}}^0 + \frac{(1 - e)p}{1 - R + \beta} = Z_{\text{cold}}^0 + p_{\text{eff}}, \quad (\text{A.86})$$

where

$$p_{\text{eff}} = \frac{(1 - e)p}{(1 - R + \beta)} \quad (\text{A.87})$$

is the effective yield produced by the model of star formation and feedback. Note that p_{eff} may depend on the potential well in which the galaxy forms through its dependence on β and e .

The net changes in masses and metal masses in each phase can then be evaluated over that time step.

A.4 Dynamical Friction Timescale

To derive the timescale on which a satellite's orbit decays we will assume a dark matter halo of mass M_{H} with an isothermal density profile truncated at radius r_{H} . The circular velocity of the halo is $V_{\text{C}}^2 = GM_{\text{H}}/r_{\text{H}}$ and is constant with radius. The density at radius r is $\rho_{\text{H}}(r) = V_{\text{C}}^2/4\pi Gr^2$. The gravitational potential at $r' < r_{\text{H}}$ is given by

$$\begin{aligned} \phi(r') &= \int_{\infty}^{r_{\text{H}}} \frac{GM_{\text{H}}}{r^2} dr + \int_{r_{\text{H}}}^{r'} \frac{V_{\text{C}}^2}{r^2} dr \\ \phi(r') &= -V_{\text{C}}^2 \left(1 - \ln \frac{r'}{r_{\text{H}}} \right). \end{aligned} \quad (\text{A.88})$$

It will be convenient to express the equations of motion under the influence of dynamical friction in terms of quantities relevant to a circular orbit. In such a circular orbit of radius r_C the tangential velocity is given by $v_\theta^2 = GM(r_C)/r_C = V_C^2$. The specific energy of this orbit is then,

$$E = \frac{1}{2}V_C^2 + \phi(r_C) = -\frac{1}{2}V_C^2 \left(1 - 2 \ln \frac{r_C}{r_H}\right), \quad (\text{A.89})$$

where we have used the expression for the specific potential energy from eqn. (A.88). Consequently

$$r_C(E) = r_H \exp\left(\frac{E}{V_C^2} + \frac{1}{2}\right). \quad (\text{A.90})$$

Using eqn. (A.90) the angular momentum of the circular orbit can be written as follows:

$$J_C(E) = V_C r_C(E) = V_C r_H \exp\left(\frac{E}{V_C^2} + \frac{1}{2}\right). \quad (\text{A.91})$$

In terms of these quantities the tangential, total and radial velocities of an orbit (not necessarily a circular one) with energy E and angular momentum J are

$$v_\theta = \epsilon J_C(E)/r, \quad (\text{A.92})$$

$$v^2 = 2(E - \phi(r)) = 2 \left[-\frac{1}{2}V_C^2 \left(1 - 2 \ln \frac{r_C}{r_H}\right) + V_C^2 \left(1 - \ln \frac{r'}{r_H}\right) \right] = V_C^2 \left(1 - 2 \ln \frac{r}{r_C(E)}\right), \quad (\text{A.93})$$

and,

$$v_r^2 = v^2 - v_\theta^2 = V_C^2 \left(1 - \frac{e^2 r_C^2}{r^2} - 2 \ln \frac{r}{r_C(E)}\right), \quad (\text{A.94})$$

where $\epsilon = J/J_C(E)$.

Under the action of dynamical friction (Chandrasekhar 1943b) the satellite galaxy will experience a force (Binney & Tremaine 1987, section 7.1)

$$\mathbf{F}_{\text{df}} = -4\pi G^2 \ln \Lambda_C \rho_H M_S^2 B(v/\sqrt{2}\sigma) \frac{\mathbf{v}}{v^3} \quad (\text{A.95})$$

where $\ln \Lambda_C \approx \ln(rv^2/GM_S)$ is the Coulomb logarithm (treating the satellite as a point mass) and

$$B(x) = \text{erf}(x) - \frac{2x}{\sqrt{\pi}} e^{-x^2}. \quad (\text{A.96})$$

The resulting rates of change of the specific energy and angular momentum of the satellite are therefore

$$\dot{E} = \mathbf{v} \cdot \mathbf{F}_{\text{df}}/M_S = v|F_{\text{df}}|/M_S, \quad (\text{A.97})$$

and

$$\dot{J} = |\mathbf{r} \wedge \mathbf{F}_{\text{df}}|/M_S = rv_\theta|F_{\text{df}}|/vM_S, \quad (\text{A.98})$$

where M_S is the mass of the satellite and we have used the fact that $|\mathbf{r} \wedge \mathbf{v}| = rv_\theta/v$. Combining eqns. (A.90), (A.91), (A.92), (A.93), (A.94), (A.95), (A.97) and (A.98) we find

$$\dot{j} = -(rv_\theta/v)|F_{\text{df}}|/M_S = -\epsilon J_C(E) \frac{GM_S \ln \Lambda_C B([1 - 2 \ln r/r_C(E)]^{1/2})}{V_C r^2 [1 - 2 \ln r/r_C(E)]^{3/2}}, \quad (\text{A.99})$$

and

$$\dot{E} = -v|F_{\text{df}}|/M_S = -\frac{GM_S V_C \ln \Lambda_C B([1 - 2 \ln r/r_C(E)]^{1/2})}{r^2 [1 - 2 \ln r/r_C(E)]^{1/2}}. \quad (\text{A.100})$$

We now determine the average rate of change of angular momentum and energy throughout one orbit. To do this we average eqns. (A.99) and (A.100) over the orbital period, using the relation $dt = dr/v_r$, where t is time. Therefore the average of some quantity Q is given by

$$\langle Q \rangle = \left(\int_{r_{\min}}^{r_{\max}} \frac{dr}{v_r} \right)^{-1} \int_{r_{\min}}^{r_{\max}} Q \frac{dr}{v_r}, \quad (\text{A.101})$$

where r_{\min} and r_{\max} are the minimum and maximum radius of the orbit found by solving eqn. (A.94) for $v_r = 0$. The orbit averaged rates of change of energy and angular momentum are then

$$\begin{aligned} \langle \dot{j} \rangle &= -\epsilon \frac{J_C(E)}{r_C^2(E)} \ln \Lambda_C \frac{GM_S}{V_C} \left(\int_{x_{\min}}^{x_{\max}} \frac{dx}{(1 - \epsilon^2 x^2 - 2 \ln x)} \right)^{-1} \\ &\times \left(\int_{x_{\min}}^{x_{\max}} \frac{B([1 - 2 \ln x]^{1/2})}{x^2 (1 - 2 \ln x)^{3/2} (1 - \epsilon^2 x^2 - 2 \ln x)} dx \right), \end{aligned} \quad (\text{A.102})$$

and

$$\begin{aligned} \langle \dot{E} \rangle &= -\frac{GM_S V_C}{r_C^2(E)} \ln \Lambda_C \left(\int_{x_{\min}}^{x_{\max}} \frac{dx}{(1 - \epsilon^2 x^2 - 2 \ln x)} \right)^{-1} \\ &\times \left(\int_{x_{\min}}^{x_{\max}} \frac{B([1 - 2 \ln x]^{1/2})}{x^2 (1 - 2 \ln x)^{1/2} (1 - \epsilon^2 x^2 - 2 \ln x)} dx \right), \end{aligned} \quad (\text{A.103})$$

where $x = r/r_C(E)$. We will now express these in terms of $\dot{\epsilon}$ and \dot{r}_C using $\langle \dot{j} \rangle = \epsilon V_C r_C (\dot{r}_C/r_C + \dot{\epsilon}/\epsilon)$ and $\dot{E} = -V_C^2 \dot{r}_C/r_C$, where we have assumed that ϵ and $r_C(E)$ are approximately constant over one orbit. In terms of these quantities eqns. (A.102) and (A.103) become

$$\begin{aligned} \frac{\dot{r}_C}{r_C} &= -\frac{GM_S}{V_C r_C^2(E)} \ln \Lambda_C \left(\int_{x_{\min}}^{x_{\max}} \frac{dx}{(1 - \epsilon^2 x^2 - 2 \ln x)} \right)^{-1} \\ &\times \left(\int_{x_{\min}}^{x_{\max}} \frac{B([1 - 2 \ln x]^{1/2})}{x^2 (1 - 2 \ln x)^{1/2} (1 - \epsilon^2 x^2 - 2 \ln x)} dx \right), \end{aligned} \quad (\text{A.104})$$

and

$$\frac{\dot{\epsilon}}{\epsilon} + \frac{\dot{r}_C}{r_C} = -\ln \Lambda_C \frac{GM_S}{V_C r_C^2} \left(\int_{x_{\min}}^{x_{\max}} \frac{dx}{(1 - \epsilon^2 x^2 - 2 \ln x)} \right)^{-1}$$

$$\begin{aligned}
& \times \left(\int_{x_{\min}}^{x_{\max}} \frac{B([1 - 2 \ln x]^{1/2})}{x^2(1 - 2 \ln x)^{3/2}(1 - \epsilon^2 x^2 - 2 \ln x)} dx \right) \\
\frac{\dot{\epsilon}}{\epsilon} = & -\ln \Lambda_C \frac{2GM_S}{V_C r_C^2} \left(\int_{x_{\min}}^{x_{\max}} \frac{dx}{(1 - \epsilon^2 x^2 - 2 \ln x)} \right)^{-1} \\
& \times \left(\int_{x_{\min}}^{x_{\max}} \frac{B([1 - 2 \ln x]^{1/2}) \ln x}{x^2(1 - 2 \ln x)^{3/2}(1 - \epsilon^2 x^2 - 2 \ln x)} dx \right). \quad (\text{A.105})
\end{aligned}$$

Note that if initially $\epsilon = 1$ then eqn. (A.105) implies that $\dot{\epsilon} = 0$ (since $x = 1$ and so $\ln x = 0$ in this case). Hence a circular orbit will remain circular under the influence of dynamical friction. In the case of $\epsilon = 1$ the evolution of $r_C = r$ is solvable analytically since eqn. (A.104) reduces to

$$\begin{aligned}
\int_0^{r_{C,i}} r_C dr_C &= -\frac{GM_S}{V_C} \ln \Lambda_C B(1) \int_{T_{df}}^0 dt \\
\frac{1}{2} r_{C,i}^2 &= \frac{GM_S}{V_C} \ln \Lambda_C B(1) T_{df}, \quad (\text{A.106})
\end{aligned}$$

where $r_{C,i}$ is the initial radius of the orbit and T_{df} is the time taken for the orbital radius to decay to zero. The dynamical friction timescale, T_{df} , is therefore

$$T_{df} = \frac{V_C r_{C,i}^2}{2GM_S \ln \Lambda_C B(1)}, \quad (\text{A.107})$$

or, using $GM_H/r_H = V_C^2$,

$$T_{df} = \frac{1}{2B(1) \ln \Lambda_C} \left(\frac{r_{C,i}}{r_H} \right)^2 \left(\frac{R_H}{V_C} \right) \left(\frac{M_H}{M_S} \right). \quad (\text{A.108})$$

In the more general case of non-circular orbits the evolution equations (A.104) and (A.105) must be solved numerically, yielding a result for T_{df} of the form (Lacey & Cole 1993)

$$T_{df} = \frac{0.3722 f(\epsilon_i)}{\ln \Lambda_C} \left(\frac{r_{C,i}}{r_H} \right)^2 \left(\frac{\pi r_H}{V_C} \right) \left(\frac{M_H}{M_S} \right), \quad (\text{A.109})$$

where ϵ_i is the initial circularity, and we have used $1/2\pi B(1) \approx 0.3722$. Lacey & Cole (1993) find that for $10^{-2} \leq \epsilon_i \leq 1$ the function $f(\epsilon_i)$ is fitted to an accuracy of better than 3% by the form $f(\epsilon_i) \approx \epsilon_i^{0.78}$.

Appendix B

Non-Uniform Reionization by Galaxies: Supporting Theory

B.1 Calculation of the Escaping Fraction

B.1.1 Escaping Fraction in the DS94 Model: Stars in Mid-Plane

In the model of Dove & Shull (1994, hereafter DS94), ionizing photons escape from galactic disks through “HII chimneys”, which are holes in the neutral gas layer ionized by OB associations. The OB associations are assumed to lie in the disk mid-plane, and to have a distribution of ionizing luminosities $dN/dS \propto S^{-2}$ for $S_1 < S < S_2$, $(dN/dS)dS$ being the number of associations with luminosities in the range S to $S + dS$ (Kennicutt, Edgar & Hodge 1989). The gas is assumed to have a Gaussian vertical distribution with scaleheight h_z . The fraction of Ly α photons escaping through chimneys on both sides of the disk at radius r is (Dove & Shull 1994, eqn. 24)

$$f_{\text{esc,gas}} = \begin{cases} 0 & \text{if } S_m \geq S_2 \\ \left[\ln \left(\frac{S_2}{S_m} \right) + \frac{9}{2} \left(\frac{S_m}{S_2} \right)^{1/3} - \frac{S_m}{2S_2} - 4 \right] / \ln \left(\frac{S_2}{S_1} \right) & \text{if } S_1 \leq S_m < S_2 \\ 1 + \left[\frac{9}{2} \left\{ \left(\frac{S_m}{S_2} \right)^{1/3} - \left(\frac{S_m}{S_1} \right)^{1/3} \right\} - \frac{1}{2} \left\{ \frac{S_m}{S_2} - \frac{S_m}{S_1} \right\} \right] / \ln \left(\frac{S_2}{S_1} \right) & \text{if } S_m < S_1, \end{cases} \quad (\text{B.1})$$

where S_m is defined as

$$S_m(r) = \pi^{3/2} n_0^2 \exp(-2r/r_{\text{disk}}) h_z^3 \alpha_{\text{H}}^{(2)}. \quad (\text{B.2})$$

Here $\alpha_{\text{H}}^{(2)}$ is the recombination coefficient for hydrogen for recombinations to all energy levels except the first, and we have assumed an exponential disk with radial scalelength r_{disk} , so that the hydrogen gas density is

$$n(r, z) = n_0 \exp(-r/r_{\text{disk}} - z^2/2h_z^2), \quad (\text{B.3})$$

r and z being the usual cylindrical polar coordinates.

Then, since $S_{\text{m}} \propto \exp(-2r/r_{\text{disk}})$ assuming h_z is constant with radius, we find from eqn. (B.1),

$$f_{\text{esc, gas}}(r) = \begin{cases} 0 & \text{if } S_{\text{m}}^0 e^{-2r/r_{\text{disk}}} \geq S_2 \\ \left[\ln \left(\frac{S_2}{S_{\text{m}}^0} \right) + \frac{2r}{r_{\text{disk}}} + \frac{9}{2} \left(\frac{S_{\text{m}}^0}{S_2} \right)^{1/3} e^{-2r/3r_{\text{disk}}} \right. \\ \quad \left. - \frac{S_{\text{m}}^0}{2S_2} e^{-2r/r_{\text{disk}}} - 4 \right] / \ln \left(\frac{S_2}{S_1} \right) & \text{if } S_1 \leq S_{\text{m}}^0 e^{-2r/r_{\text{disk}}} < S_2 \\ 1 + \left[\frac{9}{2} \left\{ \left(\frac{S_{\text{m}}^0}{S_2} \right)^{1/3} - \left(\frac{S_{\text{m}}^0}{S_1} \right)^{1/3} \right\} e^{-2r/3r_{\text{disk}}} \right. \\ \quad \left. - \frac{1}{2} \left\{ \frac{S_{\text{m}}^0}{S_2} - \frac{S_{\text{m}}^0}{S_1} \right\} e^{-2r/r_{\text{disk}}} \right] / \ln \left(\frac{S_2}{S_1} \right) & \text{if } S_{\text{m}}^0 e^{-2r/r_{\text{disk}}} < S_1, \end{cases} \quad (\text{B.4})$$

where S_{m}^0 is the value of S_{m} calculated for the density at the centre of the galaxy disk. Our model of galaxy formation calculates the radial scale length of each galaxy's disk and also the mass of cold gas present in that disk, and we assume that h_z/r_{disk} is constant. We can therefore determine n_0 and the ratios S_{m}^0/S_2 and S_{m}^0/S_1 . The escape fraction in eqn. (B.4) can be averaged over the galaxy disk if we assume the star formation rate (and hence OB association abundance) is proportional to the local gas surface density (Kennicutt 1989, Kennicutt 1997). This gives the fraction of all ionizing photons produced

by the galaxy which can escape into the IGM. The result is,

$$f_{\text{esc,gas}} = \begin{cases} \left[\left\{ \frac{352}{225} - \frac{4}{15} \ln \left(\frac{S_2}{S_m^0} \right) \right\} \left(\frac{S_2}{S_m^0} \right)^{1/2} \right. & \text{if } S_m^0 \geq S_2 \\ \left. - \left\{ \frac{352}{225} - \frac{4}{15} \ln \left(\frac{S_1}{S_m^0} \right) \right\} \left(\frac{S_1}{S_m^0} \right)^{1/2} \right] / \ln \left(\frac{S_2}{S_1} \right) \\ \left[\left\{ \frac{4}{15} \ln \left(\frac{S_1}{S_m^0} \right) - \frac{352}{225} \right\} \left(\frac{S_1}{S_m^0} \right)^{1/2} \right. & \text{if } S_1 \leq S_m^0 < S_2 \\ \left. - \frac{1}{18} \frac{S_m^0}{S_2} + \frac{81}{50} \left(\frac{S_m^0}{S_2} \right)^{1/3} + \ln \left(\frac{S_2}{S_m^0} \right) \right] / \ln \left(\frac{S_2}{S_1} \right) \\ 1 + \left[\frac{81}{50} \left\{ \left(\frac{S_m^0}{S_2} \right)^{1/3} - \left(\frac{S_m^0}{S_1} \right)^{1/3} \right\} - \frac{1}{18} \left\{ \frac{S_m^0}{S_2} - \frac{S_m^0}{S_1} \right\} \right] / \ln \left(\frac{S_2}{S_1} \right) & \text{if } S_m^0 < S_1. \end{cases} \quad (\text{B.5})$$

In Figure B.1 we show this average escape fraction as a function of the ratio S_m^0/S_2 for $S_2/S_1 = 1000$.

B.1.2 Escaping Fraction in DS94 Model: Stars Tracing Gas

In the DS94 model OB associations are assumed to lie in the midplane of the galaxy disk. If instead OB associations are spread throughout the gas layer, having the same vertical distribution as the cold gas, then the resulting escape fraction will be higher than that in the DS94 model. We assume the same density profile as before, given by eqn. (B.3). Consider an OB association emitting S ionizing photons per second, at position (r, z) in the disk. We make the assumption (as did DS94) that the radial variations in density can be ignored for calculating the escape fraction at radius r (which will be a valid assumption provided the size of the HII region formed is much less than r_{disk}). In order for any photons emitted into a cone of solid angle $d\Omega$ which makes an angle θ with the z -axis to escape the galaxy, the emission rate of photons into this cone must exceed the total recombination rate in the cone. This occurs for an ionizing luminosity $S_{\text{req}}(\theta)$, where,

$$S_{\text{req}}(\theta) \frac{d\Omega}{4\pi} = n_0^2 \alpha_{\text{H}}^{(2)} \exp\left(-2\frac{r}{r_{\text{disk}}}\right) \int_0^\infty l^2 \exp\left(-\frac{(z+l\cos\theta)^2}{h_z^2}\right) dl d\Omega, \quad (\text{B.6})$$

which can be written as $S_{\text{req}}^\pm(\theta) = \pm S_{\text{req}}^{0,\pm} / \cos^3 \theta$ (S_{req}^+ is the solution for $\cos \theta > 0$ and S_{req}^- is the solution for $\cos \theta < 0$), where

$$S_{\text{req}}^{0,\pm} = S_m^0 e^{-2r/r_{\text{disk}}} \left\{ \left[1 \mp \operatorname{erf} \left(\frac{z}{h_z} \right) \right] \left(1 + \frac{2z^2}{h_z^2} \right) - \frac{2}{\sqrt{\pi}} \frac{z}{h_z} \exp \left(-\frac{z^2}{h_z^2} \right) \right\}. \quad (\text{B.7})$$

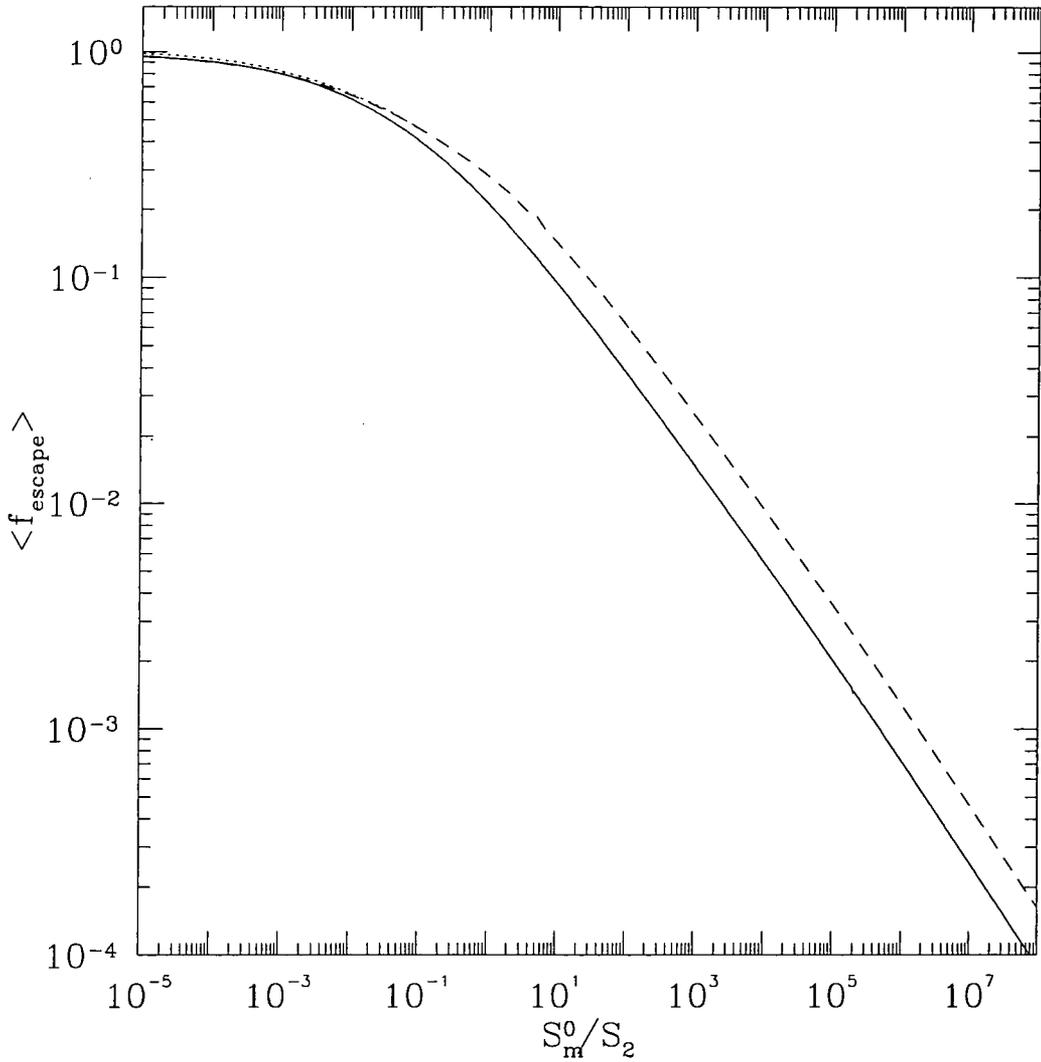


Figure B.1: The average escape fraction for a galactic disk in the DS94 model with OB associations in the disk mid-plane (solid line) and distributed as the cold gas (dashed line). Both models assume $S_2/S_1 = 1000$. The dotted line (visible just above the solid line at small values of S_m^0/S_2) shows the effects of accounting for radial variations in the gas density in the model where OB associations are distributed as the gas. This line only differs noticeably from the dashed line at the lowest values of S_m^0/S_2 .

This defines two critical angles, $\cos \theta_c^\pm(S) = \pm(S_{\text{req}}^{0,\pm}/S)^{1/3}$, such that photons can escape the galaxy only if $\theta < \theta_c^+(S)$ or $\theta > \theta_c^-(S)$.

$$\cos \theta_c^\pm(S) = \left(\frac{S_m^0 e^{-2r/r_{\text{disk}}}}{S} \left\{ \left[1 \pm \operatorname{erf} \left(\frac{z}{h_z} \right) \right] \left(1 + \frac{2z^2}{h_z^2} \right) - \frac{2}{\sqrt{\pi}} \frac{z}{h_z} \exp \left(-\frac{z^2}{h_z^2} \right) \right\} \right)^{1/3}. \quad (\text{B.8})$$

The total escaping fraction from this OB association is then given by,

$$f_{\text{esc,gas}}(S) = \frac{1}{2S} \left[\int_{\cos \theta_c^+}^1 [S - S_{\text{req}}^+(\theta)] d(\cos \theta) + \int_{-1}^{\cos \theta_c^-} [S - S_{\text{req}}^-(\theta)] d(\cos \theta) \right] \quad (\text{B.9})$$

$$f_{\text{esc,gas}}(S) = 1 - \frac{3}{4} \left(\frac{S_{\text{req}}^{0,+}}{S} \right)^{1/3} - \frac{3}{4} \left(\frac{S_{\text{req}}^{0,-}}{S} \right)^{1/3} + \frac{S_{\text{req}}^{0,+}}{4S} + \frac{S_{\text{req}}^{0,-}}{4S}. \quad (\text{B.10})$$

Averaging this escape fraction over the assumed OB association luminosity function then gives a mean escape fraction of

$$f_{\text{esc,gas}} = \begin{cases} 0 & \text{if } S_2 > S_m^0 e^{-\frac{2r}{r_{\text{disk}}}} \\ \left[\frac{1}{2} \ln \left(\frac{S_2^2}{S_{\text{req}}^{0,+} S_{\text{req}}^{0,-}} \right) + \frac{9}{4} \left\{ \left(\frac{S_{\text{req}}^{0,+}}{S_2} \right)^{1/3} + \left(\frac{S_{\text{req}}^{0,-}}{S_2} \right)^{1/3} - 2 \right\} \right] & \text{if } S_1 < S_m^0 e^{-\frac{2r}{r_{\text{disk}}}} \leq S_2 \\ -\frac{1}{4} \left\{ \frac{S_{\text{req}}^{0,+}}{S_2} + \frac{S_{\text{req}}^{0,-}}{S_2} - 2 \right\} / \ln \left(\frac{S_2}{S_1} \right) & \\ \left[\frac{9}{4} \left\{ \left(\frac{S_{\text{req}}^{0,+}}{S_2} \right)^{\frac{1}{3}} - \left(\frac{S_{\text{req}}^{0,+}}{S_1} \right)^{\frac{1}{3}} + \left(\frac{S_{\text{req}}^{0,-}}{S_2} \right)^{\frac{1}{3}} - \left(\frac{S_{\text{req}}^{0,-}}{S_1} \right)^{\frac{1}{3}} \right\} \right] & \text{if } S_m^0 e^{-\frac{2r}{r_{\text{disk}}}} \leq S_1. \\ -\frac{1}{4} \left\{ \frac{S_{\text{req}}^{0,+}}{S_2} - \frac{S_{\text{req}}^{0,+}}{S_1} + \frac{S_{\text{req}}^{0,-}}{S_2} - \frac{S_{\text{req}}^{0,-}}{S_1} \right\} / \ln \left(\frac{S_2}{S_1} \right) + 1 & \end{cases} \quad (\text{B.11})$$

This expression is then averaged over the galaxy disk, assuming a star formation rate proportional to the local gas density, to derive the mean escaping fraction for the entire galaxy. This must be done numerically

So far we have ignored radial variations in the density of the gas when computing the escaping fraction from a single OB association. This will cause us to underestimate $f_{\text{esc,gas}}$ at low values of S_m^0/S_2 where the size of the HII region around an OB association may become comparable to or larger than the disk radial scale length. A more accurate estimate for $f_{\text{esc,gas}}$ can be obtained in this regime as follows.

Consider photons being emitted by a source at position (r_0, z_0) into a small element of solid angle, $d\Omega$, along the direction (θ, ϕ) where θ and ϕ are the usual spherical polar

coordinates centred on (r_0, z_0) and with $\theta = 0$ coinciding with the z -axis. The total recombination rate in this cone is given by,

$$d\dot{n}_{\text{req}} = \int_0^\infty l^2 n^2(r, z) \alpha_{\text{H}}^{(2)} dl d\Omega, \quad (\text{B.12})$$

where l is measured along the line (θ, ϕ) and where $r^2 = r_0^2 + l^2 \sin^2 \theta - 2lr_0 \sin \theta \cos \phi$ and $z = z_0 + l \cos \theta$. An isotropically radiating source emitting S ionizing photons per second emits $S d\Omega/4\pi$ photons per second into this cone. The minimum source luminosity required for some photons to escape, S_{req} , is therefore,

$$S_{\text{req}}(r_0, z_0, \theta, \phi) = \frac{4S_{\text{m}}^0}{\pi^{1/2} h_z^3} \int_0^\infty x^2 [n(r, z)/n_0]^2 dx, \quad (\text{B.13})$$

where $x = l/r_{\text{disk}}$. The fraction of ionizing photons escaping the galaxy from this position is therefore given simply by,

$$f_{\text{esc}}(r_0, z_0, \theta, \phi) = \int_{\max(S_{\text{req}}, S_1)}^{S_2} (S - S_{\text{req}}) \frac{dN}{dS} dS \Big/ \int_{S_1}^{S_2} S \frac{dN}{dS} dS. \quad (\text{B.14})$$

For the specific case of $dn/dS \propto S^{-2}$ the escape fraction can be written as,

$$f_{\text{esc}}(r_0, z_0, \theta, \phi) = \begin{cases} 0 & \text{if } S_{\text{req}} \geq S_2 \\ \left[\frac{S_{\text{req}}}{S_2} - 1 - \ln \left(\frac{S_{\text{req}}}{S_2} \right) \right] / \ln \left(\frac{S_2}{S_1} \right) & \text{if } S_1 \leq S_{\text{req}} < S_2 \\ \left[\frac{S_{\text{req}}}{S_2} - \frac{S_{\text{req}}}{S_1} + \ln \left(\frac{S_2}{S_1} \right) \right] / \ln \left(\frac{S_2}{S_1} \right) & \text{if } S_{\text{req}} < S_1. \end{cases} \quad (\text{B.15})$$

For this model $f_{\text{esc}}(r_0, z_0, \theta, \phi)$ must be evaluated numerically. The resulting function is then averaged over all θ, ϕ, r_0 and z_0 , assuming that the star formation rate is proportional to the local gas density. The resulting escape fraction depends only upon S_{m}^0/S_2 for a given h_z . The average escape fraction for $h_z = 0.1$ is shown in Fig. B.1. As can be seen the results differ from the case where radial density variations are ignored only for low S_{m}^0/S_2 as expected.

B.1.3 Escaping Fraction in the DSGN98 Model

In the model of Devriendt et al. (1998, DSGN98) the stars producing the Lyman continuum photons are assumed to be uniformly mixed with the gas in the galaxy, which is distributed in an exponential disk. All of the hydrogen in the galaxy is assumed to be in the form of HI, allowing the optical depth for ionizing photons to be calculated.

We have calculated the escaping fraction in this model exactly, using the density profile given by eqn. (B.3). As in the case of the DS94 model with stars mixed uniformly with the gas, we begin by finding the escaping fraction as a function of position (r_0, z_0) and line of sight (θ, ϕ) . For the DSGN98 model we therefore find the total optical depth in neutral hydrogen along the line of sight, which is

$$\tau(r_0, z_0, \theta, \phi) = \sigma_{\text{HI}} \int_0^\infty n(r, z) dl, \quad (\text{B.16})$$

where $\sigma_{\text{HI}} = 6.3 \times 10^{-18} \text{cm}^{-2}$ is the cross-section for hydrogen ionization at 912\AA . The quantity $\exp(-\tau)$ is then averaged over all r_0, z_0, θ and ϕ , assuming that the star formation rate is proportional to the local gas density, to obtain the final escaping fraction. The average escape fraction for the entire galaxy depends only on the quantity $\tau_0 = n_0 r_{\text{disk}} \sigma_{\text{HI}}$ for a given value of h_z/r_{disk} , and is shown in Fig. B.2 for $h_z/r_{\text{disk}} = 0.1$.

B.1.4 Escaping Fraction in Starbursts

In the case of a burst of star formation triggered by a major merger, we use the same $f_{\text{esc, gas}}$ as for quiescent star formation in the the case where $f_{\text{esc, gas}}$ is assumed fixed, but in the DS94 and DSGN98 models we estimate the escape fraction by assuming the burst has an approximately spherical geometry, throughout which star formation proceeds uniformly. We assume a sphere of uniform hydrogen number density, n , given by

$$n = \frac{3M_{\text{gas}}}{4\pi r_{\text{burst}}^3 1.4m_{\text{H}}}, \quad (\text{B.17})$$

where M_{gas} is the mass of cold gas in the burst, r_{burst} is the radius of the region in which the burst occurs, and m_{H} is the mass of a hydrogen atom. The factor of 1.4 accounts for the presence of helium in the gas. We assume also that photons escape only from an outer shell of thickness l , within which the optical depth is less than 1. Therefore,

$$nl\sigma_{\text{HI}} \approx 1, \quad (\text{B.18})$$

where σ_{HI} is the cross section for hydrogen ionization. The escape fraction is simply the fraction of the sphere's volume in this shell, i.e.

$$f_{\text{esc, gas}} \approx \frac{4\pi r_{\text{burst}}^2 l}{4\pi/3 r_{\text{burst}}^3}. \quad (\text{B.19})$$

Substituting for l then gives

$$\begin{aligned} f_{\text{esc, gas}} &= \frac{3}{r_{\text{burst}} n \sigma_{\text{HI}}} \\ &= \frac{4\pi r_{\text{burst}}^2 1.4m_{\text{H}}}{M_{\text{gas}} \sigma_{\text{HI}}}. \end{aligned} \quad (\text{B.20})$$

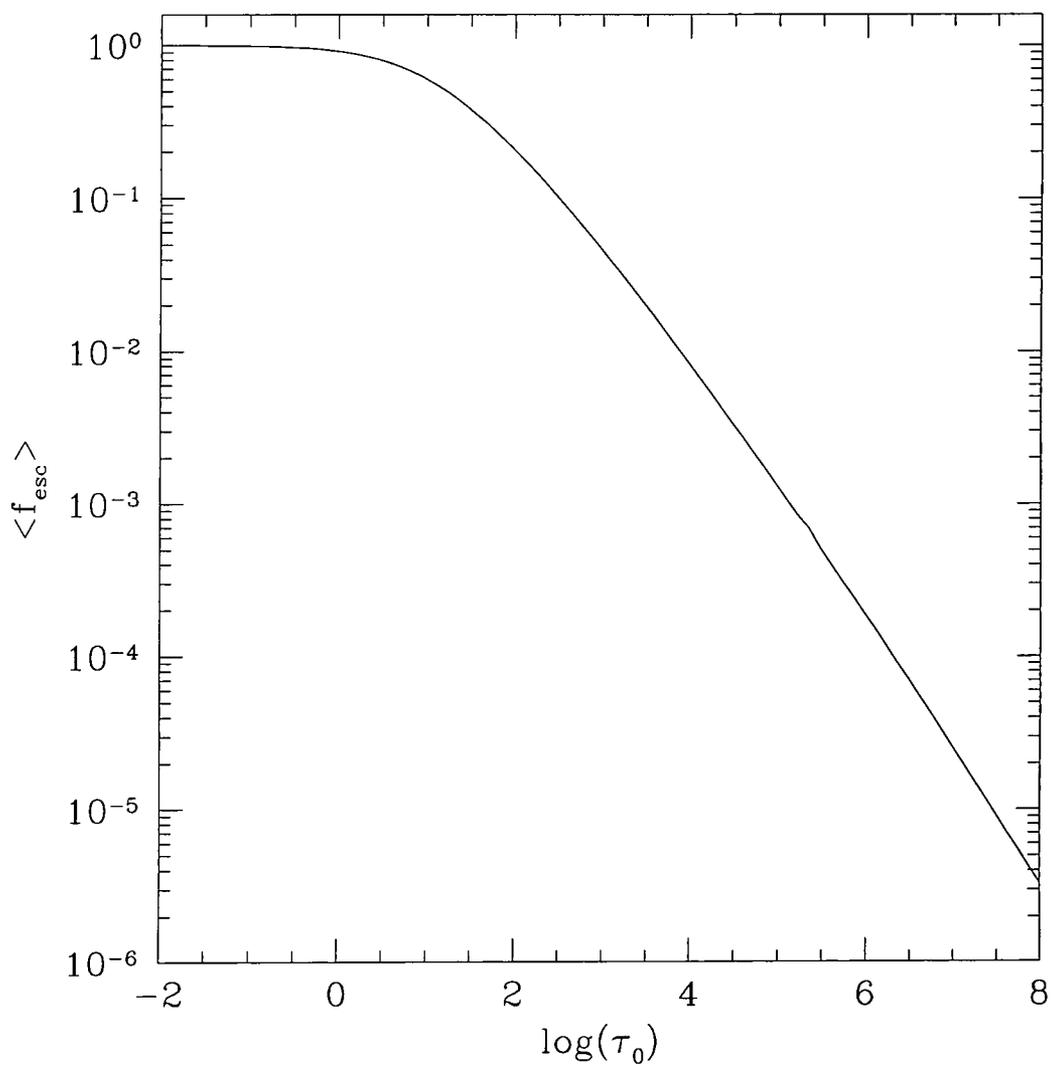


Figure B.2: Mean escape fraction for a galaxy disk in the DSGN98 model as a function of τ_0 .

We take r_{burst} to be equal to $0.1r_{\text{bulge}}$ where r_{bulge} is the half-mass radius of the bulge formed by the merger. This choice is motivated by observational fact which shows that starburst activity is usually confined to the nuclear region, the size of which is much smaller than that of the galaxy as a whole (e.g. Sanders & Mirabel 1996, and references therein). Ricotti & Shull (1999) have carried out more elaborate calculations of escaping fractions from spherical galaxies. However, their results are applicable to gas in hydrostatic equilibrium with an NFW dark matter profile and so are not well suited to the case of starbursts.

The star formation rate in the burst is assumed to decline exponentially, with an e-folding time equal to f_{dyn} times the bulge dynamical time. Unless noted otherwise, we assume $f_{\text{dyn}} = 1$ in all models. As the burst proceeds the mass of cold gas present, M_{gas} , declines as it is turned into stars. The escape fraction, $f_{\text{esc,gas}}$, therefore increases during the burst, reaching unity as the amount of gas present drops to zero. However, as the star formation rate is declining exponentially during the burst only a small fraction of photons are produced whilst $f_{\text{esc,gas}}$ is high.

B.2 Calculation of Clumping Factor

To estimate the clumping factor of the photoionized IGM, we make the simplifying assumption that gas in the universe can be split into three components — that which has fallen into dark matter halos and is collisionally ionized or is part of a galaxy, that which has fallen into dark matter halos and is *not* collisionally ionized, and that which has remained outside halos and is smoothly distributed. The first component makes no contribution to the clumping factor. We define the clumping factor as

$$f_{\text{clump}} = \frac{\langle \rho_{\text{IGM}}^2 \rangle}{\bar{\rho}_{\text{IGM}}^2} = \frac{\langle \rho_{\text{IGM}}^2 \rangle}{f_{\text{IGM}}^2 \bar{\rho}^2}, \quad (\text{B.21})$$

where ρ_{IGM} is the IGM gas density at any point in the universe (i.e. it does *not* include contributions from collisionally ionized gas or galaxies), $\bar{\rho}_{\text{IGM}} = f_{\text{IGM}}\bar{\rho}$ is the mean density of gas in the IGM and $\bar{\rho}$ is the mean density of all gas in the universe (here f_{IGM} is the fraction of the total mass of gas in the universe which resides in the IGM, as defined in §5.2.2).

Let $f_{\text{m,clumped}}$ be the fraction of mass in halos above the Jeans halo mass, M_{J} , (the mass below which halos are unable to retain photoionized gas) as calculated from the Press-Schechter mass function for example. These halos occupy a fraction of the volume

of the universe given by $f_{v,\text{clumped}} = f_{m,\text{clumped}}/\Delta_{\text{vir}}$. Here, Δ_{vir} is the mean density within the virial radius of a halo in units of the mean density of the Universe. The smooth component of gas is assumed to uniformly fill the region outside halos with $M > M_J$, and so has density

$$\rho_{\text{smooth}} = \bar{\rho} f_{m,\text{smooth}}/f_{v,\text{smooth}}, \quad (\text{B.22})$$

where $f_{m,\text{smooth}} = 1 - f_{m,\text{clumped}}$ is the mass fraction of gas in this smooth component, and $f_{v,\text{smooth}} = 1 - f_{v,\text{clumped}}$ is the fraction of the volume of the universe that it occupies.

Consider next the non-collisionally ionized gas in a *single* dark matter halo. Averaging over the volume of this one halo we obtain

$$\langle \rho_{\text{clumped}}^2 \rangle = f_{\text{int}}(1 - f_{\text{gal}})^2(1 - x_{\text{H}})^2 \Delta_{\text{vir}}^2 \bar{\rho}^2, \quad (\text{B.23})$$

where f_{gal} is the fraction of the baryons which have become part of galaxies within the halo, x_{H} is the ionized fraction for the hydrogen in the halo gas assuming collisional ionization equilibrium (which we take from the calculations of Sutherland & Dopita 1993), and f_{int} is a factor of order unity which depends on the shape of the halo gas density profile and is given by

$$f_{\text{int}} = \frac{\int_0^{r_{\text{vir}}} \rho^2(r) r^2 dr}{\int_0^{r_{\text{vir}}} \bar{\rho}_{\text{int}}^2 r^2 dr}. \quad (\text{B.24})$$

Here r_{vir} is the virial radius of the halo, $\rho(r)$ is the density profile of the diffuse gas in the halo, and $\bar{\rho}_{\text{int}}$ is the mean density of this gas within the virial radius. We ignore any dependence of the density profile of the gas in the halo on the fraction which has cooled to form galaxies. Our results should be insensitive to this assumption, as $f_{\text{gal}} \ll 1$ in halos where x_{H} is significantly less than unity.

To find the contribution of gas in halos to the clumping factor, we integrate the above expression over all halos more massive than M_J , weighting by the volume for each halo. Adding the contribution from the smooth component, we then obtain

$$f_{\text{clump}} = \frac{f_{m,\text{smooth}}^2}{f_{v,\text{smooth}} f_{\text{IGM}}^2} + \frac{f_{\text{int}} \Delta_{\text{vir}}}{f_{\text{IGM}}^2} \int_{M_J}^{\infty} \langle (1 - f_{\text{gal}})^2 \rangle (1 - x_{\text{H}})^2 \frac{M_{\text{halo}}}{\rho_c \Omega_0} \frac{dn}{dM_{\text{halo}}} dM_{\text{halo}}, \quad (\text{B.25})$$

where we have used that fact that the comoving volume of a dark matter halo of mass M_{halo} is $M_{\text{halo}}/(\Delta_{\text{vir}} \Omega_0 \rho_c)$ (ρ_c being the critical density of the universe at $z = 0$). Here $\langle (1 - f_{\text{gal}})^2 \rangle$ is averaged over all halos of mass M_{halo} in our model of galaxy formation.

We determine M_J by finding the mass of a dark matter halo which has a potential well deep enough that it can just hold onto reionized gas. This gives us the minimum mass halo within which gas collects. For the halo to just retain its gas,

$$\frac{dP}{dr} = \frac{GM_J}{r_{\text{vir}}^2} \rho(r_{\text{vir}}), \quad (\text{B.26})$$

where r_{vir} is the virial radius of the halo and P is the gas pressure. We approximate this as

$$\frac{P}{r_{\text{vir}}} \approx \frac{GM_{\text{J}}}{r_{\text{vir}}^2} \rho(r_{\text{vir}}), \quad (\text{B.27})$$

and using the ideal gas law this becomes

$$\frac{k_{\text{B}}T}{\mu m_{\text{H}}} \approx \frac{GM_{\text{J}}}{r_{\text{vir}}} = \frac{4\pi}{3} G r_{\text{vir}}^2 \rho_{\text{c}} \Omega_0 \Delta_{\text{vir}} (1+z)^3, \quad (\text{B.28})$$

where we have used the relation $M_{\text{J}} = 4\pi \rho_{\text{c}} \Omega_0 (1+z)^3 \Delta_{\text{vir}} r_{\text{vir}}^3 / 3$. The virial radius is therefore

$$r_{\text{vir}} = \left(\frac{3}{4\pi G \mu m_{\text{H}} \rho_{\text{c}} \Omega_0 \Delta_{\text{vir}}} k_{\text{B}}T \right)^{1/2} (1+z)^{-3/2}, \quad (\text{B.29})$$

and the minimum halo mass in which gas is retained is

$$M_{\text{J}} = \frac{4\pi}{3} \rho_{\text{c}} (1+z)^3 \Omega_0 \Delta_{\text{vir}} r_{\text{vir}}^3. \quad (\text{B.30})$$

We evaluate f_{int} for the case of an isothermal profile with core radius r_{c} :

$$\rho_{\text{gas}}(r) \propto \frac{1}{r^2 + r_{\text{c}}^2}. \quad (\text{B.31})$$

The simulations of galaxy clusters by Navarro, Frenk & White (1995) and Eke, Navarro & Frenk (1998) show that the gas density profile is well described by this form. For a halo of virial radius r_{vir} and containing a mass of gas M_{gas} this becomes

$$\rho_{\text{gas}}(r) = \frac{M_{\text{gas}}}{4\pi r_{\text{c}}} \left[\frac{r_{\text{vir}}}{r_{\text{c}}} - \arctan \frac{r_{\text{vir}}}{r_{\text{c}}} \right]^{-1} (r^2 + r_{\text{c}}^2)^{-1}. \quad (\text{B.32})$$

Substituting the constant density core gas density profile in eqn. (B.24) we find,

$$\Delta_{\text{vir}}^2 f_{\text{int}} = \frac{\Delta_{\text{vir}} M_{\text{gas}}}{4\pi \bar{\rho}_{\text{int}} r_{\text{c}}^2} \left[\frac{r_{\text{vir}}}{r_{\text{c}}} - \arctan \frac{r_{\text{vir}}}{r_{\text{c}}} \right]^{-2} \int_0^{r_{\text{vir}}} \frac{r^2 dr}{r^2 + r_{\text{c}}^2}. \quad (\text{B.33})$$

Changing variables using $r = r_{\text{c}} \tan \theta$ this becomes,

$$\Delta_{\text{vir}}^2 f_{\text{int}} = \frac{\Delta_{\text{vir}} M_{\text{gas}}}{4\pi \bar{\rho}_{\text{gas}} r_{\text{c}}^3} \left[\frac{r_{\text{vir}}}{r_{\text{c}}} - \arctan \frac{r_{\text{vir}}}{r_{\text{c}}} \right]^{-2} \int_0^{\arctan r_{\text{vir}}/r_{\text{c}}} \sin^2 \theta d\theta \quad (\text{B.34})$$

$$= \frac{\Delta_{\text{vir}} M_{\text{gas}}}{8\pi \bar{\rho}_{\text{gas}} r_{\text{c}}^3} \left[\frac{r_{\text{vir}}}{r_{\text{c}}} - \arctan \frac{r_{\text{vir}}}{r_{\text{c}}} \right]^{-2} \int_0^{\arctan r_{\text{vir}}/r_{\text{c}}} (1 - \cos 2\theta) d\theta \quad (\text{B.35})$$

$$= \frac{\Delta_{\text{vir}} M_{\text{gas}}}{8\pi \bar{\rho}_{\text{gas}} r_{\text{c}}^3} \left[\frac{r_{\text{vir}}}{r_{\text{c}}} - \arctan \frac{r_{\text{vir}}}{r_{\text{c}}} \right]^{-2} \left[\arctan \frac{r_{\text{vir}}}{r_{\text{c}}} - \frac{1}{2} \sin 2 \arctan \frac{r_{\text{vir}}}{r_{\text{c}}} \right] \quad (\text{B.36})$$

Using the relation $M_{\text{gas}} = 4\pi \bar{\rho} \Delta_{\text{vir}} r_{\text{vir}}^3 / 3$, the above simplifies to

$$f_{\text{int}} = \frac{1}{6} \left(\frac{r_{\text{vir}}}{r_{\text{c}}} \right)^3 \left[\frac{r_{\text{vir}}}{r_{\text{c}}} - \arctan \frac{r_{\text{vir}}}{r_{\text{c}}} \right]^{-2} \left[\arctan \frac{r_{\text{vir}}}{r_{\text{c}}} - \frac{r_{\text{vir}}}{r_{\text{c}}} \left(1 + \frac{r_{\text{vir}}^2}{r_{\text{c}}^2} \right)^{-1} \right]. \quad (\text{B.37})$$

For a typical value of $r_{\text{vir}}/r_{\text{c}} = 10$, we therefore find $f_{\text{int}} = 3.14$.

B.3 The Spectrum of CMB Secondary Anisotropies

In Chapter 5, we concentrate on the kinematic Sunyaev-Zel'dovich effect which is induced by the peculiar motions (deviations from pure Hubble flow) of free electrons in ionized regions (Sunyaev & Zel'dovich 1980, Vishniac 1987). There exist other secondary sources of CMB anisotropies. However, on angular scales smaller than a few arc-minutes, the kinematic Sunyaev-Zel'dovich effect is likely to provide a dominant contribution. For example, it is known that the temperature anisotropies caused by non-linear growth of density perturbations, which are often referred to as the Rees-Sciama effect or integrated Sachs-Wolfe effect, are of order 10^{-7} or less (Seljak & Zaldarriaga 1996). These anisotropies depend on the the baryon bulk physical peculiar velocity, \mathbf{v} , and the number density of free electrons, n_e . In our calculations of the anisotropies we assume that the \mathbf{v} is equal to the bulk velocity of the dark matter and that n_e in ionized regions is proportional to the dark matter density.

The temperature anisotropy $\Theta(\gamma) = \frac{\Delta T}{T}$ observed in a given line of sight direction γ is (e.g. Hu 2000)

$$\Theta(\gamma, \eta_0) = - \int_{\eta_{\text{rec}}}^{\eta_0} \frac{d\eta}{(1+z)} \gamma_i v_{\text{B}}^i \dot{\tau}, \quad (\text{B.38})$$

where $\eta \equiv \int (1+z) dt$ is conformal time with its values at recombination and present denoted, respectively, by η_{rec} and η_0 . In eqn. (B.38) we have assumed an optically thin universe. In an optically thick universe these temperature fluctuations are damped by a factor $e^{-\tau}$, where the optical depth is $\tau = \int d\eta \sigma_{\text{T}} n_e / (1+z)$, where σ_{T} is the cross section for Thomson scattering. If the universe became instantaneously fully ionized after some redshift z_i , the relation between the optical depth $\tau(\eta_i, \eta_0)$ and z_i is approximately obtained as $z_i = 100\Omega_0 (0.025/\Omega_{\text{b}}h)^{2/3} \tau^{1/3}$. Therefore, if the reionization takes place at $z \ll 100\Omega_0 (0.025/\Omega_{\text{b}}h)^{2/3}$, as is the case in our reionization model, then the damping factor can be neglected.

The usual procedure to obtain the angular correlation function of temperature anisotropies in eqn. (B.38) is by means of Limber's equation in Fourier space (see for example Peebles 1980). However, in Chapter 5, we work in real space since we have the two point correlation functions of density and velocity fields directly measured in real space from N-body simulations.

The temperature angular correlation $C(\theta)$ can be written as

$$C(\theta) = \sigma_{\text{T}}^2 \int_{\eta_{\text{rec}}}^{\eta_0} d\eta \int_{\eta_{\text{rec}}}^{\eta_0} d\eta' \gamma_i \gamma_j' \langle v^i(\mathbf{x}, \eta) v^j(\mathbf{x}', \eta') n_e(\mathbf{x}, \eta) n_e(\mathbf{x}', \eta') \rangle, \quad (\text{B.39})$$

where $\gamma_i \gamma'^i = \cos \theta$, and, \mathbf{x} and \mathbf{x}' refer, respectively, to comoving coordinates in the past light geodesics in the directions γ and γ' at η and η' . We write n_e terms of density fluctuations δ as

$$n_e(\mathbf{x}, \eta) = \bar{n}_e(\eta) x_e(\mathbf{x}, \eta) [1 + \delta(\mathbf{x}, \eta)], \quad (\text{B.40})$$

where $\bar{n}_e(\eta)$ is the mean total (free and bound) electron number density at time η , and $x_e(\mathbf{x})$, the ionization fraction, is unity in ionized regions and zero otherwise. The correlation lengths of velocity and density fields are small compared to the Hubble radius so that we can approximate $n_e(\mathbf{x}', \eta') = n_e(\mathbf{x}', \eta)$ and similarly for the v , in eqn. (B.39).

$$\zeta = \left[\frac{x_e(1 + \delta)}{\langle x_e(1 + \delta) \rangle} - 1 \right] v_{\text{los}}, \quad (\text{B.41})$$

where $v_{\text{los}} = \gamma_i v^i$ is the velocity component in the direction γ . Therefore $C(\theta)$ can be written in terms of the velocity correlation function $\xi_{vv}(y) \equiv \langle v_{\text{los}}(\mathbf{x}) v_{\text{los}}(\mathbf{x} + \mathbf{y}) \rangle$ and the density-velocity correlation function $\xi_{\zeta\zeta}(y) \equiv \langle \zeta(\mathbf{x}) \zeta(\mathbf{x} + \mathbf{y}) \rangle$, both evaluated for fields at the same η .

$$C(\theta) = \sigma_T^2 \int_{\eta_{\text{rec}}}^{\eta_0} \frac{d\eta}{1+z} \int_{\eta_{\text{rec}}}^{\eta_0} \frac{d\eta'}{1+z'} \bar{n}_e(\eta) \bar{n}_e(\eta') \langle x_e(1 + \delta) \rangle^2 [\xi_{\zeta\zeta}(|\mathbf{x}' - \mathbf{x}|) + \xi_{vv}(|\mathbf{x}' - \mathbf{x}|) + \langle \zeta(\mathbf{x}) v_{\text{los}}(\mathbf{x}') + \zeta(\mathbf{x}') v_{\text{los}}(\mathbf{x}) \rangle]. \quad (\text{B.42})$$

The dominant contribution to $C(\theta)$ is from the term involving $\xi_{\zeta\zeta}$. The integration over ξ_{vv} yields to phase cancellation (Kaiser 1984b, Ostriker & Vishniac 1986, Vishniac 1987). The last term in the integrand also has negligible contribution*. We have checked that the dominant term produces at least an order of magnitude larger anisotropies than the other terms.

In flat space we use the triangle relation, $|\mathbf{x}' - \mathbf{x}|^2 = x^2 + x'^2 - 2xx' \cos \theta$, we first carry out the integration of eqn. (B.42) in terms of η' for fixed η and θ . We compute $\xi_{\zeta\zeta}$ at an average redshift \bar{z} given by $1/(1 + \bar{z}) = (1/(1 + z_1) + 1/(1 + z_2))/2$, which is an appropriate approximation if the correlation length is negligible relative to the horizon scale. It is straightforward to extend the calculation to an open geometry.

From the temperature angular correlation $C(\theta)$, we can obtain C_ℓ as

$$C_\ell = 2\pi \int_{-1}^1 d \cos \theta P_\ell(\cos \theta) C(\theta), \quad (\text{B.43})$$

where $P_\ell(\cos \theta)$ is the Legendre polynomial.

*It is interesting that the contribution from the integral over $\xi_{\zeta\zeta}$ is still dominant even if we approximate $\xi_{\zeta\zeta} = \xi_{\delta\delta} \xi_{vv}$, i.e. if we ignore any correlations between the density and velocity fields

Appendix C

Models for the X-ray Emission from Cooling Gas

Following White & Frenk (1991, WF91), we assume that cooling occurs within a static potential, and that the cooling gas remains isothermal as it flows to the halo centre. The assumption of isothermality allows considerable simplification in the determination of the X-ray luminosity of the flow since it is then directly related to the drop in potential energy (i.e. the internal energy of the gas remains constant). Detailed, fully self-consistent flow models, such as those of Bertschinger (1989) and Abadi, Bower & Navarro (2000), show that these assumptions are reasonable for the range of scales over which we wish to determine the flow solution.

Consider a unit mass of gas inflowing with the cooling flow. In a Lagrangian frame we have,

$$\frac{dq}{dt} = \frac{\partial u}{\partial t} + v_R \frac{\partial u}{\partial r} + \frac{\partial \phi}{\partial t} + v_R \frac{\partial \phi}{\partial r}, \quad (\text{C.1})$$

where q is the total energy, u is the internal (thermal) energy, ϕ is the gravitational energy per unit mass of gas, and v_R is the inflow velocity of the gas. We make the simplifying assumptions that both the hydrostatic cooling flow and the gravitational potential of the halo are constant in time. (This amounts to assuming that the inflowing gas has little effect on the halo potential and that any changes in the form of the cooling flow occur on timescales significantly longer than the time it takes a mass of gas to flow in; neither is exactly true, but should suffice as reasonable approximations.) The two partial derivatives with respect to time then vanish leaving

$$\frac{dq}{dt} = v_R \left(\frac{\partial u}{\partial r} + \frac{\partial \phi}{\partial r} \right) = -l_X, \quad (\text{C.2})$$

where l_X is the X-ray luminosity emitted per unit mass of gas (n.b. the negative sign

appears because the gas is losing energy as it emits X-rays). We now make the important assumption that the cooling flow is *isothermal*. Assuming that the gas behaves as a perfect gas (an excellent approximation at these temperatures and pressures) the internal energy per unit mass of gas is a function of temperature alone, namely

$$u = \frac{3kT}{2\mu m_{\text{H}}}, \quad (\text{C.3})$$

where k is Boltzmann's constant, T is the gas temperature, μ is the mean molecular weight of the gas and m_{H} is the mass of the hydrogen atom. Since the temperature is assumed to be constant with radius, the internal energy of the gas per unit mass is also constant with radius. Therefore,

$$l_{\text{X}} = -v_{\text{R}} \frac{\partial \phi}{\partial r}. \quad (\text{C.4})$$

To find the total energy emitted by this mass of gas as it flows inwards we need to know the energy dq' it emits in moving from a radius r to $r + dr$. Since it takes the gas a time $dt = dr/v_{\text{R}}(r)$ to do this we have

$$dq' = -\frac{l_{\text{X}}(r)}{v_{\text{R}}(r)} dr = \frac{\partial \phi}{\partial r} dr. \quad (\text{C.5})$$

In models such as those of Kauffmann, White & Guiderdoni (1993) and Cole et al. (1994) gas is assumed to flow from the cooling radius towards the centre of the halo. No mass is assumed to fall out of the flow until it reaches the central galaxy. We wish to determine the total X-ray energy emitted by a unit mass of gas as it travels from the cooling radius to some final radius, which we chose to be the optical radius of the galaxy. Integrating from r_{cool} to r_{optical} we find the total energy emitted by the mass of gas,

$$q' = \int_{r_{\text{optical}}}^{r_{\text{cool}}} \frac{\partial \phi}{\partial r} dr. \quad (\text{C.6})$$

If we now make the assumption that the mass cooling rate, \dot{M}_{cool} , is approximately constant over the time it takes the mass of gas to inflow, we can find the total luminosity of the cooling flow by multiplying q' by \dot{M}_{cool} , i.e. using the fact that

$$\frac{\partial \phi}{\partial r} = \frac{GM(r)}{r^2} = \frac{V_{\text{c}}^2(r)}{r}, \quad (\text{C.7})$$

where G is the gravitational constant, we find that

$$L_{\text{X}} = \dot{M}_{\text{cool}}(r_{\text{cool}}) \int_{r_{\text{optical}}}^{r_{\text{cool}}} \frac{V_{\text{c}}^2(r)}{r} dr. \quad (\text{C.8})$$

If the cooling rate is not constant throughout the flow it becomes necessary to consider a weighted sum of q' 's, i.e. we would need to integrate over $dq'(r)$ weighting by the local

value of $\dot{M}_{\text{cool}}(r)$. This would require us to follow the evolution of the cooling flow in a self-consistent way. The relevant equation for the total luminosity is, however, easily seen to be,

$$L_X = \int_{r_{\text{optical}}}^{r_{\text{cool}}} \dot{M}_{\text{cool}}(r) dq' = \int_{r_{\text{optical}}}^{r_{\text{cool}}} \dot{M}_{\text{cool}}(r) \frac{V_c^2(r)}{r} dr. \quad (\text{C.9})$$

The correct solution would then be to use for $\dot{M}_{\text{cool}}(r)$ the value $\dot{M}_{\text{cool}}(r_{\text{cool}}(t - t_{\text{inf}}))$ where t_{inf} is the time it has taken a mass element to flow from its initial position in the halo to its present position in the cooling flow.

We can also derive this relation by considering the specific entropy of the gas in the cooling flow,

$$\bar{d}q = T ds, \quad (\text{C.10})$$

where $\bar{d}q$ is the specific energy flowing into the gas, T its temperature and ds the change in specific entropy. Then for an isothermal cooling flow,

$$l_X = -v_R(r) \frac{\bar{d}q}{dr} \quad (\text{C.11})$$

$$l_X = -v_R(r) T \frac{ds}{dr}. \quad (\text{C.12})$$

The specific entropy is given by (Bower 1997)

$$s = C_V \ln \left(\frac{T}{\rho^{\gamma-1}} \right) + s_0, \quad (\text{C.13})$$

where $C_V = 3k/2\mu m_H$ and $\gamma = C_P/C_V = 5/3$ for a monatomic ideal gas (here C_V and C_P are the specific heats at constant volume and pressure respectively). Hence eqn. (C.12) becomes,

$$l_X = -v_R(r) (\gamma - 1) \frac{3}{2} \frac{kT}{\mu m_H} \left[-\frac{1}{\rho_g(r)} \frac{d\rho_g(r)}{dr} \right]. \quad (\text{C.14})$$

Using the hydrostatic equilibrium equation simplifies this to,

$$l_X = -v_R(r) (\gamma - 1) \frac{3}{2} \frac{kT}{\mu m_H} \frac{GM(r)}{r^2} \frac{\mu m_H}{kT} \quad (\text{C.15})$$

$$l_X = -v_R(r) \frac{GM(r)}{r^2}, \quad (\text{C.16})$$

which is equivalent to eqns. (C.4) and (C.7) as derived from energetics arguments.

C.1 Calculation of X-ray Luminosity in Specific Halo Potentials

We consider two possibilities for the structure of the dark matter halo. In §C.1.1 we consider an isothermal halo and in §C.1.3 we consider the dark matter profile described

by Navarro, Frenk & White (1997, hereafter NFW), in each case assuming that the gas initially traces the dark matter. In the case of the NFW halo also consider the effects of the central galaxy on the gravitational potential of the halo.

C.1.1 Cooling in an Isothermal Halo

We start by reviewing the simplest realistic halo model, the singular isothermal sphere (or just isothermal sphere for brevity). In this model, the dark matter density profile, $\rho(r) \propto r^{-2}$, where r is radial distance in the halo. This is the halo model adopted by WF91 who calculated the rate at which gas cools and so determined its bolometric X-ray luminosity.

The cooling radius, r_{cool} , in the halo is defined as the radius at which the cooling time of the gas equals the age of the halo, which in the cases of interest here is approximately equal to the age of the Universe. That is,

$$t_{\text{cool}}(r_{\text{cool}}) = t_0, \quad (\text{C.17})$$

where $t_{\text{cool}}(r)$ is the gas cooling time (i.e. the energy per unit volume of gas divided by the rate at which that gas is radiating energy) at radius r in the halo and t_0 is the age of the Universe.

Assuming a spherically symmetric halo in which the gas is initially distributed like the dark matter, the cooling time is given by (White & Frenk 1991)

$$t_{\text{cool}}(r) = \frac{192\pi\Omega_0 G m_{\text{H}}^2 r^2}{49 f_{\text{g}} \Omega_{\text{b}} \Lambda(\mu m_{\text{H}} V_{\text{c}}^2 / 2k)}, \quad (\text{C.18})$$

where Ω_{b} is the baryon density in units of the critical density, $\Lambda(T)$ is the cooling function of the gas, f_{g} is the fraction of the initial baryon density remaining in gaseous form and V_{c} is the circular velocity of the halo.

From eqns. (C.17) and (C.18), the cooling radius is given by,

$$r_{\text{cool}} = \left(\frac{49 f_{\text{g}} \Omega_{\text{b}} \Lambda(\mu m_{\text{H}} V_{\text{c}}^2 / 2k) t_0}{192\pi\Omega_0 G m_{\text{H}}^2} \right)^{1/2}. \quad (\text{C.19})$$

Here we have assumed that the gas temperature is constant as a function of radius and equal to the virial temperature. This structure follows from the assumption that the gas density profile parallels that of the dark matter and is in approximate agreement with the results of cosmological hydrodynamical simulations of the formation of galaxy clusters (e.g. Frenk et al. 1999).

$$3kT_{\text{vir}} = \frac{GM_{\text{vir}}\mu m_{\text{H}}}{r_{\text{vir}}}, \quad (\text{C.20})$$

where M_{vir} and r_{vir} are the virial mass and radius of the halo respectively.

The evolution of the cooling radius determines the mass of gas cooling per unit time:

$$\dot{M}_{\text{cool}}(r_{\text{cool}}) = 4\pi r_{\text{cool}}^2 \rho_{\text{g}}(r_{\text{cool}}) \frac{dr_{\text{cool}}}{dt}, \quad (\text{C.21})$$

where $\rho_{\text{g}}(r)$ is the gas density. WF91 show that in a $\Omega_0 = 1$ universe,

$$\dot{M}_{\text{cool}}(r_{\text{cool}}) = \frac{3}{4} \frac{f_{\text{g}} \Omega_{\text{b}}}{\Omega_0} H_0 (1+z)^{3/2} \frac{V_{\text{c}}^2(r_{\text{cool}}) r_{\text{cool}}(V_{\text{c,max}}, z)}{G}, \quad (\text{C.22})$$

where we take r_{cool} to be a function of $V_{\text{c,max}}$, the maximum circular velocity in the halo. Whilst $V_{\text{c}}(r)$ is constant for an isothermal halo, this is not the case for the NFW halos to be considered later. Defining r_{cool} in this way avoids any ambiguities. Generalising this relation to an arbitrary Friedman cosmology gives,

$$\dot{M}_{\text{cool}}(r_{\text{cool}}) = \frac{1}{2} \frac{f_{\text{g}} \Omega_{\text{b}}}{\Omega_0} \frac{V_{\text{c}}^2(r_{\text{cool}}) r_{\text{cool}}(V_{\text{c,max}}, z)}{G t_0(z)}, \quad (\text{C.23})$$

where $t_0(z)$ is the age of the universe at redshift z .

If we assume that the cooling rate of gas remains approximately constant throughout the cooling flow we can use eqn. (C.8) to estimate the total X-ray luminosity of the halo. In the case of an isothermal halo for which V_{c}^2 is constant with radius we find,

$$L_{\text{X}} = \dot{M}_{\text{cool}}(r_{\text{cool}}) V_{\text{c}}^2 \ln \left(\frac{r_{\text{cool}}}{r_{\text{optical}}} \right). \quad (\text{C.24})$$

Typically the cooling radius is an order of magnitude greater than the optical radius and so the logarithmic term in the above equation will be ~ 2.5 . This is then identical to the result used by WF91.

Since the gravitational potential steepens in the inner parts of the halo, that is where we would expect most of the X-ray emission to come from if the cooling flow actually extended to within r_{optical} . However, because of the logarithmic dependence on the integration limits in eqn. (C.24) a cooling flow extending down to $r_{\text{optical}}/2$ would only increase the predicted luminosity by a factor ~ 2 . Similarly a flow which terminated at $2r_{\text{optical}}$ would reduce the luminosity by less than half. However, the exact form of eqn. (C.24) depends on the structure of the cooling flow. In §6.3, we considered what would happen if the flow had a more complex, possibly multi-phase, structure.

C.1.2 Gas with a Constant Density Core

In the standard model just considered, the gas is initially assumed to follow the dark matter. We now study a model in which the gas has a constant density core with radius r_{core} . The density profile adopted in this case is

$$\rho_g(r) \propto \frac{1}{r^2 + r_{\text{core}}^2} \quad (\text{C.25})$$

which is the well-known “ β -model” with $\beta = 2/3$. Including such a core changes both the cooling radius (and hence \dot{M}_{cool}) and also the potential through which the gas falls (which results from the combined effect of the singular isothermal dark matter halo and the β -model gas). Straightforward calculation leads to the following expressions

$$r_{\text{cool}} = \left(\frac{49}{192\pi} \frac{f_g \Omega_b}{\Omega_0} \left(1 + \frac{r_{\text{core}}}{r_{\text{vir}}} \arctan \left[\frac{r_{\text{vir}}}{r_{\text{core}}} \right] \right)^{-1} \times \frac{\Lambda (\mu m_{\text{H}} V_c^2 / 2k) t_0}{G m_{\text{H}}^2} - r_{\text{core}}^2 \right)^{1/2} \quad (\text{C.26})$$

$$\dot{M}_{\text{cool}}(r_{\text{cool}}) = \frac{f_g \Omega_b}{2\Omega_0} \frac{V_c^2 r_{\text{cool}}}{G t_0} \left(1 + \frac{r_{\text{core}}}{r_{\text{vir}}} \arctan \left[\frac{r_{\text{vir}}}{r_{\text{core}}} \right] \right)^{-1} \quad (\text{C.27})$$

$$L_X = \dot{M}_{\text{cool}}(r_{\text{cool}}) V_c^2 \left[\left(1 - \frac{\Omega_b}{\Omega_0} \right) \ln \left(\frac{r_{\text{cool}}}{r_{\text{optical}}} \right) + \frac{\Omega_b}{\Omega_0} \left\{ 1 + \frac{r_{\text{core}}}{r_{\text{vir}}} \arctan \left(\frac{r_{\text{vir}}}{r_{\text{core}}} \right) \right\}^{-1} \left\{ \ln \left(\frac{r_{\text{cool}}}{r_{\text{optical}}} \right) + \int_{r_{\text{optical}}/r_{\text{core}}}^{r_{\text{cool}}/r_{\text{core}}} \arctan(y) \frac{dy}{y^2} \right\} \right]. \quad (\text{C.28})$$

If $r_{\text{core}} \ll r_{\text{cool}}$ the X-ray luminosity predicted by the β -model is very close to the predictions of the model in which the gas traces the dark matter.

C.1.3 Cooling in an NFW Halo

NFW found a simple formula, the NFW halo profile, that fits the density profile of dark matter halos in N-body simulations of hierarchical clustering. This formula provides a good approximation to the density profile of halos of all masses over at least two orders of magnitude in radius and is independent of the values of the cosmological parameters. This profile is characterised by two parameters: δ_c , a characteristic density and r_s , a characteristic radial scale in the halo. NFW find that δ_c and r_s are correlated so that the NFW profile provides a universal, one-parameter model of the halo.

The NFW profile is given by,

$$\rho(r) = \frac{\delta_c \rho_c}{(r/r_s)[1 + (r/r_s)]^2}. \quad (\text{C.29})$$

In order to calculate X-ray luminosities in an NFW halo it is necessary to obtain first the value of the parameter r_s , or equivalently the concentration parameter, c , defined as

$c = r_{200}/r_s$ (here r_{200} is the radius in the halo which encloses an overdensity of 200), which is related to δ_c as shown by NFW. The concentration, c , is related to δ_c by:

$$\delta_c = \frac{200}{3} \frac{c^3}{\ln(1+c) - c/(1+c)}. \quad (\text{C.30})$$

To find the (virial) mass of the halo from the observed circular velocity of the galaxy we must extrapolate from that circular velocity to the circular velocity at the virial radius, $V_{c,\text{vir}}$. We assume that the measured circular velocity of the galaxy corresponds to the maximum in the NFW rotation curve. The extrapolation then depends upon on the value of c . However, c itself must be determined from the virial mass and redshift of the halo. The solution to this circular problem which we have adopted is to make an initial guess for c which allows us to determine the mass of the halo (for the value of c chosen). We can then determine what value of c a halo of this mass should have according to the results of N-body simulations. This new value of c is then used to make a new estimate of the virial mass. The iteration continues until c converges to a given accuracy. The concentrations so found are in agreement with the results of Navarro, Frenk & White (1996).

The calculation of the virial mass proceeds as follows. Firstly, we find the critical density of the Universe at the redshift of the galaxy,

$$\rho_c = \frac{3H^2(z)}{8\pi G}, \quad (\text{C.31})$$

where $H(z)$ is as given in eqn. (A.5). In an NFW halo the maximum circular velocity is related to the circular velocity at r_{200} by,

$$V_{c,200} = V_{c,\text{max}} \left(\frac{2 \ln(1+c) - c/(1+c)}{c \ln 3 - 2/3} \right)^{1/2}, \quad (\text{C.32})$$

and so we are able to find r_{200} from the relation,

$$r_{200} = \left(\frac{3}{4\pi G 200 \rho_c} \right)^{1/2} V_{c,200}. \quad (\text{C.33})$$

Now, if $\rho_{\text{vir}} = \delta_{\text{vir}} \rho_c$ then the virial radius r_{vir} is defined by the relation,

$$\frac{4\pi}{3} \rho_{\text{vir}} r_{\text{vir}}^3 = \int_0^{r_{\text{vir}}} 4\pi \rho(r) r^2 dr \quad (\text{C.34})$$

$$\frac{r_{\text{vir}}}{r_s} = \left[\frac{3\delta_c}{\delta_{\text{vir}}} \left(\ln \left\{ 1 + \frac{r_{\text{vir}}}{r_s} \right\} - \frac{r_{\text{vir}}/r_s}{1 + r_{\text{vir}}/r_s} \right) \right]^{1/3}. \quad (\text{C.35})$$

Eqn. (C.35) can be solved numerically to find r_{vir}/r_s which is then simply scaled by r_{200}/c to find r_{vir} .

The density within the virial radius depends on the cosmology and may be estimated from the spherical top-hat model as described in §A.1.3. Finally, the virial mass is found from,

$$M_{\text{vir}} = \frac{4\pi}{3} r_{\text{vir}}^3 \rho_{\text{vir}}. \quad (\text{C.36})$$

This mass can then be used to calculate c as described by Navarro, Frenk & White (1996). This calculation requires an assumption about the power spectrum of density fluctuations. For the purposes of Chapter 6 we assume the standard CDM form with amplitude $\sigma_8 = 0.67$, which approximately produces the observed present-day abundance of galaxy clusters (White, Efstathiou & Frenk 1993). This value of c is then used to recalculate the virial mass, and this is used to find a new value of c .

The circular velocity profile corresponding to the NFW profile has the form,

$$V_c^2(r) = V_c^2(r_{200}) \frac{1}{x} \frac{\ln(1+cx) - cx/(1+cx)}{\ln(1+c) - c/(1+c)}, \quad (\text{C.37})$$

where $x = r/r_{200}$. Using the circular velocity profile of an NFW halo we can now follow through the calculation used above, noting that for an NFW profile the cooling radius is defined by the relation

$$t_0 = \frac{192\pi}{49} \frac{1}{f_g} \frac{\Omega_0}{\Omega_b} \frac{Gm_{\text{H}}^2 r_s^2}{\Lambda(T_{\text{vir}})} \frac{x_{\text{cool}}}{x_{\text{vir}}} (1+cx_{\text{cool}})^2 \times [\ln(1+cx_{\text{vir}}) - cx_{\text{vir}}/(1+cx_{\text{vir}})]^{-1}. \quad (\text{C.38})$$

The cooling rate is then

$$\begin{aligned} \dot{M}_{\text{cool}}(r_{\text{cool}}) &= \frac{f_g \Omega_b}{\Omega_0} \frac{V_{c,200}^2 r_{200}}{G t_0} \\ &\times \frac{c^2 x_{\text{cool}}^2}{(1+3cx_{\text{cool}})(1+cx_{\text{cool}})} \\ &\times [\ln(1+c) - c/(1+c)]^{-1}, \end{aligned} \quad (\text{C.39})$$

where $x = r/r_{200}$. We then obtain

$$\begin{aligned} L_X &= \dot{M}_{\text{cool}}(r_{\text{cool}}) V_c^2(r_{200}) \\ &\times \int_{x_{\text{optical}}}^{x_{\text{cool}}} x^{-2} \frac{\ln(1+cx) - cx/(1+cx)}{\ln(1+c) - c/(1+c)} dx, \end{aligned} \quad (\text{C.40})$$

where $x_{\text{cool}} = r_{\text{cool}}/r_{200}$ and $x_{\text{optical}} = r_{\text{optical}}/r_{200}$. This equation must be integrated numerically for any given c .

C.1.4 The Influence of the Central Galaxy

So far we have considered cooling in gravitational potentials which are due entirely to dark matter. Since each of the halos we study contains a galaxy at its centre the baryonic

matter of the galaxy will contribute to the gravitational potential, and furthermore will alter the distribution of dark matter in the halo. We include a central galaxy that grows slowly in the already collapsed dark matter halo. This allows the dark matter to respond adiabatically to the baryonic matter of the galaxy as described in §2.4.3. The mass of dark matter within radius r after the galaxy has grown satisfies eqn. (2.41). With the further assumption that shells of dark matter do not cross during this compression, this equation can be solved for the structure of the dark matter halo after the growth of the galaxy. It is then simple to determine the rotation curve (and hence the gravitational potential) for the galaxy accounting for contributions from dark matter plus the galactic disk and bulge. The cooling radius, rate and X-ray luminosity in a halo of this form can then be found (using eqns. (C.21) and (C.8) for the latter two). These calculations must be carried out numerically as the resulting halo potential has no simple analytic form.

Bibliography

The following abbreviations are used in this bibliography:

MNRAS: Monthly Notices of the Royal Astronomical Society

ApJ: The Astrophysical Journal

ApJS: The Astrophysical Journal Supplement

AJ: The Astronomical Journal

A&A: Astronomy and Astrophysics

ARA&A: Annual Reviews of Astronomy and Astrophysics

Abadi M., Bower R. G., Navarro J., 2000, MNRAS, 314, 759

Abel T., Haehnelt M. G., 1999, ApJ, 520, 13

Adelberger K. L., Steidel C. C., Giavalisco M., Pettini M., Kellogg M., 1998, ApJ, 505, 18

Aghanim N., Désert F. X., Puget J. L., Gispert R., 1996, A&A, 311, 1

Alimi J.-M., Valls-Gabaud D., Blanchard A., 1988, A&A, 206, L11

Arnaud K. A., 1996, in “Astronomical Data Analysis Software and Systems V”, eds. G. Jacoby and J. Barnes, ASP Conference Series volume 101, p. 17

Bagla J. S., 1998, MNRAS, 299, 417

Bahcall N. A., 1974, ApJ, 187, 439

Bahcall N. A., 1977, ARA&A, 15, 505

Ball R. S., 1901, “In The High Heavens”, Isbister, London

Bardeen J.M., Bond J.R., Kaiser N., Szalay A.S., 1986, ApJ, 304, 15

Barnes J., White S. D. M., 1984, MNRAS, 211, 753

- Barnes J. E., Efstathiou G. P., 1987, ApJ, 319, 575
- Barnes J. E., 1992, ApJ, 393, 484
- Baugh C. M., 1996, MNRAS, 280, 267
- Baugh C. M., Cole S., Frenk C. S., 1996a, MNRAS, 282, L27
- Baugh C. M., Cole S., Frenk C. S., 1996b, MNRAS, 283, 1361
- Baugh C. M., Cole S., Frenk C. S., Benson A. J., Lacey C. G., 1998a, in "Star Formation in Early-Type Galaxies", eds. P. Carral and J. Cepa, ASP Conference Series, Vol. 163, Astronomical Society of the Pacific, San Francisco, p. 227
- Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 1998b, ApJ, 498, 504
- Baugh C. M., Benson A. J., Cole S., Frenk C. S., Lacey C. G., 1999, MNRAS, 305, L21
- Bennett C. L., Banday A. J., Gorski K. M., Hinshaw G., Jackson P., Keegstra P., Kogut A., Smoot G. F., Wilkinson D. T., Wright E. L., 1996, ApJ, 464, 1
- Benoist C., Maurogordato S., da Costa L. N., Cappi A., Schaeffer R., 1996, ApJ, 472, 452
- de Bernardis P., et al., 2000, Nat., 404, 955
- Bertschinger E., 1989, ApJ, 340, 666
- Binney J. J., 1977, ApJ, 215, 483
- Binney J., Tremaine S., 1987, Princeton University Press
- Birkhoff G. D., 1923, "Relativity and Modern Physics", Cambridge, Mass., Harvard University Press
- Blanton M., Cen R., Ostriker J. P., Strauss M. A., Tegmark M., 2000, ApJ, 531, 1
- Blitz L., Spergel D. N., Teuben P. J., Hartmann D., Burton W. B., 1999, ApJ, 514, 818
- Blumenthal G. R., Faber S. M., Flores R., Primack J. P., 1986, ApJ, 301, 27
- Bond J. R., Efstathiou G., Silk J., 1980, Phys. Rev. Lett., 45, 1980
- Bond J. R., Efstathiou G., 1984, ApJ, 285, L45
- Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

- Bower R. G., 1991, MNRAS, 248, 332
- Bower R. G., 1997, MNRAS, 288, 355
- Branchini E., Teodoro L., Frenk C. S., Schmoldt I., Efstathiou G., White S. D. M., Saunders W., Rowan-Robinson M., Keeble O., Tadros H., Maddox S., Oliver S., Sutherland W., 1999, MNRAS, 308, 1
- Bressan A., Chiosi C., Fagotto F., 1994, ApJS, 94, 63
- Brighenti F., Matthews H., 1999, AJ, 117, 1056
- Bruscoli M., Ferrara A., Fabbri R., Ciardi B., 2000, astro-ph/9911467
- Bruzual A. G., Charlot S., 1993, ApJ, 405, 538
- Bruzual A. G., Charlot S., 2000, in preparation
- Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 1999, astro-ph/9908159
- Burles S., Tytler D., 1998, Sp. Sc. Rev., 84, 65
- Burstein D., Jones C., Forman W., Marston A. P., Marzke R. O., 1997, ApJS, 111, 163
- Carlberg R. G., Cowie L. L., Songaila A., Hu E. M., 1997, ApJ, 484, 538
- Cavaliere A., Fusco-Femiano R., 1976, A&A, 94, 137
- Chandrasekhar S., 1943a, Rev. Mod. Phys., 15, 2; reprinted, 1954, in "Selected Papers on Noise and Stochastic Processes", ed. Wax N., Dover, New York, p. 3
- Chandrasekhar S., 1943b, ApJ, 97, 255
- Charlot S., Worthey G., Bressan A., 1996, ApJ, 57, 625
- Chiu W. A., Ostriker J. P., 1999, submitted to ApJ (astro-ph/9907220)
- Christodoulou D. M., Shlosman I., Tohline J. E., 1995, ApJ, 443, 563
- Ciardi B., Ferrara A., Governato F., Jenkins A., 2000, MNRAS, 314, 611
- Cojazzi P., Bressan A., Lucchin F., Pantano O., 1999, submitted to MNRAS (astro-ph/9908030)

- Cole S., 1991, ApJ, 367, 45
- Cole S., Lacey C. G., 1993, MNRAS, 262, 627
- Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781
- Cole S., Lacey C. G., 1996, MNRAS, 281, 716
- Cole S., Baugh C. M., Frenk C. S., Lacey C. G., 2000, submitted to MNRAS
- Coles P., Lucchin F., 1995, "Cosmology: The Origin and Evolution of Cosmic Structure", Wiley, England.
- Colín P., Klypin A., Kravtsov A., Khokhlov A., 1999, ApJ, 523, 32
- Colless M., 1996, in proceedings of the Heron Island Conference, <http://msowww.anu.edu.au/heron/Colless/colless.html>
- Colpi M., Mayer L., Governato F., 1999, ApJ, 525, 720
- Copi C. J., Schramm D. N., Turner M. S., 1995, ApJ, 455, L95
- Couchman H. M. P., Rees M. J., 1986, MNRAS, 221, 53
- Davis M., Geller M. J., 1976, ApJ, 208, 13
- Davis M., Huchra J., Latham D. W., Tonry J., 1982, ApJ, 253, 423
- Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
- Davis M., Meiskin A., Strauss M. A., da Costa L. N., Yahil A., 1988, ApJ, 333, L9
- de Jong R. S., Lacey C. G., 1999, in "The Low Surface Brightness Universe", proceedings of IAU Colloquium 171
- de Vaucoulers G., 1979, ApJ, 227, 729
- Dekel A., Rees M. J., 1987, Nat., 326, 455
- Devriendt J. E. G., Sethi S. K., Guiderdoni B., Nath B. B., 1998, MNRAS, 298, 708 (DSGN98)

- Diaferio A., Kauffmann G., Colberg J. M., White S. D. M., 1999, MNRAS, 307, 537
- Dickey J. M., Lockman F. J., 1990, ARA&A, 28, 215
- Doroskevich A. G., 1970, Astrofisika, 6, 581
- Dove J. B., Shull J. M., 1994, ApJ, 430, 222 (DS94)
- Dove J. B., Shull J. M., Ferrera A., 2000, ApJ, 531, 846
- Dressler A., 1980, ApJ, 236, 351
- Efstathiou G., Lake G., Negroponte J., 1982, MNRAS, 199, 1069
- Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, MNRAS, 235, 715
- Efstathiou G., 1992, MNRAS, 256, 43
- Efstathiou G., Bond J. R., White S.D.M., 1992, MNRAS, 258, 1
- Efstathiou G., 2000, astro-ph/0002247 (submitted to MNRAS)
- Eke V. R., Cole S., Frenk C. S., 1996, MNRAS, 282, 263
- Eke V. R., Cole S., Frenk C. S., Henry J. P., 1998, MNRAS, 298, 1145
- Eke V. R., Navarro J. F., Frenk C. S., 1998, ApJ, 503, 569
- Evrard A. E., Summers F. J., Davis M., 1994, ApJ, 422, 11
- Fabbiano G., Juda J. Z., 1997, ApJ, 476, 666
- Faber S. M., Gallagher J. S., 1979, ARA&A, 17, 135
- Fabian A. C., ARA&A, 32, 227
- Ferrara A., Bianchi S., Cimatti A., Giovanardi C., 1999, ApJS, 123, 437
- Flores R., Primack J. R., Blumenthal G. R., Faber S. M., 1993, ApJ, 412, 443
- Forman W., Jones C., Tucker W., 1985, ApJ, 293, 102
- Frenk C. S., White S. D. M., Davis M., 1983, ApJ, 271, 417
- Frenk C. S., White S. D. M., Efstathiou G., Davis M. 1985, Nat, 327, 595.
- Frenk C. S., White S. D. M., Davis M., Efstathiou G., 1988, ApJ, 327, 507.

- Frenk C. S., Evrard A. E., White S. D. M., Summers F. J., 1996, ApJ, 472, 460
- Frenk C. S., White S. D. M., Bode P., Bond J. R., Bryan G. L., Cen R., Couchman H. M. P., Evrard A. E., Gnedin N., Jenkins A., Khokhlov A. M., Klypin A., Navarro J. F., Norman M. L., Ostriker J. P., Owen J. M., Pearce F. R., Pen U.-L., Steinmetz M., Thomas P. A., Villumsen J. V., Wadsley J. W., Warren M. S., Xu G., Yepes G., 1999, ApJ, 525, 554
- Gaitskell R., 1998, Abstracts of the 19th Texas Symposium on Relativistic Astrophysics and Cosmology, Eds. J. Paul, T. Montmerle and E. Aubourg (CEA Saclay)
- Gallego et al., 1995a, ApJ, 455, L1
- Gallego et al., 1995b, ApJ, 459, L43
- Gardner J. P., Sharples R. M., Frenk C. S., Carrasco B. E., 1997, ApJ, 480, L99
- Geller M. J., Diaferio A., Kurtz M. J., 1999, ApJ, 517, L23
- Giallongo E., Fontana A., Madau P., 1997, MNRAS, 289, 629
- Giavalisco M., Steidel C. C., Adelberger K. L., Dickinson M. E., Pettini M., Kellogg M., 1998, ApJ, 503, 543
- Gingold R.A., Monaghan J.J., 1977, MNRAS, 181, 375
- Giovanelli R., Haynes M. P., Chincarini G. L., 1986, ApJ, 300, 77
- Giroux M. L., Shapiro P. R., 1994, ApJS, 102, 191
- Glazebrook K., Ellis R., Colless M., Broadhurst T., Allington-Smith J., Tanvir N., 1995, MNRAS, 273, 157
- Gnedin N. Y., Ostriker J. P., 1997, ApJ, 486, 581
- Gnedin N. Y., Hui L., 1998, MNRAS, 296, 44
- Gnedin N. Y., 2000, astro-ph/0002151
- Godwin J. G., Metcalfe N., Peach J. V., 1983, MNRAS, 202, 113
- Governato F., Baugh C. M., Frenk C. S., Cole S., Lacey C. G., Quinn T., Stadel J., 1998, Nat, 392, 359

- Granato G. L., Lacey C. G., Silva L., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, astro-ph/0001308
- Gronwall C., in “After the Dark Ages: When Galaxies were Young (the Universe at $2 < z < 5$)”, eds. S. Holt and E. Smith, American Institute of Physics Press, 1999, p. 335
- Gross M. A. K., Somerville R. S., Primack J. R., Holtzman J., Klypin A., 1998, MNRAS, 301, 81
- Groth E. J., Peebles P. J. E., 1977, ApJ, 217, 38
- Gruzinov A., Hu W., 1998, ApJ, 508, 435
- Gunn J.E., Peterson B.A., 1965, ApJ, 142, 1633
- Gunn J. E., Weinberg D. H., 1995, in “Wide-Field Spectroscopy and the Distant Universe”, eds. S. J. Maddox and A. Aragón-Salamanca, proceedings of the 35th Hermonceux conference, World Scientific, Singapore, p. 3
- Guth A. H., 1981, Physical Review D, 23, 347
- Guzzo L., Bartlett J. G., Cappi A., Maurogordato S., Zucca E., Zamorani G., Balkowski C., Blanchard A., Cayatte V., Chincarini G., Collins C. A., Maccagni D., MacGillivray H., Merighi R., Mignoli M., Proust D., Ramella M., Scaramella R., Stirpe G. M., Vettolani G., 2000, A&A, 355, 1
- Haardt F., Madau P., 1996, ApJ, 461, 20
- Haehnelt M. G., Steinmetz M., Rauch M., 1998, ApJ, 495, 647
- Haiman Z., Loeb A., 1996, ApJ, 483, 21
- Haiman Z., Knox L., 1999, in “Microwave Foregrounds”, eds. de Oliveira-Costa & Tegmark, ASP, San Francisco,
- Hamilton A. J. S., 1988, ApJ, 331, L59
- Hamilton A. J. S., Kumar P., Lu E., Matthews A., 1991, ApJ, 374, L1
- Hamilton A. J. S., 1992, ApJ, 385, 5
- Hanany S., et al., 2000, astro-ph/0005123 (submitted to ApJ)

- Hartwick F. D. A., Schade D., 1990, ARA&A, 28, 437
- Hawkins M. R. S., 1998, MNRAS, 296, L20
- Hoyle F., 1949, in "Problems in Cosmical Aerodynamics", Dayton, Ohio: Central Air Documents Office, p. 195
- Hoyle F., Baugh C. M., Shanks T., Ratcliffe A., 1999, MNRAS, 309, 659
- Hu W., 2000, ApJ, 529, 12
- Hubble E. P., 1929, ApJ, 69, 103
- Hubble E. P., Humason M. L., 1931. ApJ, 74, 43
- Huchra J. P., Davis M., Latham D. W., Tonry J., 1983, ApJS, 53, 89
- Huchtmeier W. K., Richter O.-G., 1988, A&A, 203, 237
- Hurwitz M., Jelinsky P., Dixon W. V. D., 1997, ApJ, 481, L31
- Infante L., Pritchett C. J., 1993, ApJ, 439, 565
- Iovino A., Giovanelli R., Haynes M., Chincarini G., Guzzo L., 1993, MNRAS, 265, 21
- Jaffe A. H., Kamionkowski M., 1998, astro-ph/9801022
- Jenkins A., Frenk C. S., Pearce F. R., Thomas P. A., Colberg J. M., White S. D. M., Couchman H. M. P., Peacock J. P., Efstathiou G., Nelson A. H., (The Virgo Consortium), 1998, ApJ, 499, 20
- Jenkins A., Frenk C. S., White, S. D. M., Colberg, J. M., Cole S., Evrard A. E., Yoshida N., 2000, submitted to MNRAS, (astro-ph/0005260)
- Jing Y. P., Mo H. J., Börner G., 1998, ApJ, 494, L1.
- Jing Y. P., 1998, ApJ, 503, L9
- Kaiser N., 1984a, ApJ, 282, 374
- Kaiser N., 1984b, ApJ, 284, L9
- Kaiser N., 1987, MNRAS, 227, 1
- Kaiser N., Wilson G., Luppino G., Kofman L., Gioia I., Metzger M., Dahle H., 1998, astro-ph/9809268

- Kaastra J. S., 1992, Internal SRON-Leiden Report, updated version 2.0
- Katz N., Gunn J. E., 1991, ApJ, 377, 365
- Katz N., Hernquist L., Weinberg D. H., 1992, ApJ, 399, 109
- Katz N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19
- Kauffmann G., White, S. D. M., Guiderdoni, B., 1993, MNRAS, 264, 201
- Kauffmann G., Guiderdoni B., White S. D. M., 1994, MNRAS, 267, 981
- Kauffmann G., 1995, MNRAS, 274, 153
- Kauffmann G., 1996a, MNRAS, 281, 475
- Kauffmann G., 1996b, MNRAS, 281, 487
- Kauffmann G., Charlot S., White S. D. M., 1996, MNRAS, 283, 117
- Kauffmann G., Nusser A., Steinmetz M., 1997, MNRAS, 286, 795
- Kauffmann G., Charlot S., 1998a, MNRAS, 294, 705
- Kauffmann G., Charlot S., 1998b, MNRAS, 297, 23
- Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999a, MNRAS, 303, 188
- Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999b, MNRAS, 307, 529
- Kauffmann G., Haehnelt M., 2000, MNRAS, 311, 576
- Kay S. T., Pearce F. R., Jenkins A., Frenk C. S., White S. D. M., Thomas P. A., Couchman H. M. P., 1999, astro-ph/9908107 (submitted to MNRAS)
- Kennefick J. D., Djorgovski S. G., de Carvalho R. R., 1995, AJ, 110, 2553
- Kennicutt R. C., 1983, ApJ, 272, 54
- Kennicutt R. C., Edgar B. K., Hodge P. W., 1989, ApJ, 337, 761
- Kennicutt R. C., 1989, ApJ, 344, 685
- Kennicutt R. C., 1997, in "Starbursts: Triggers, Nature and Evolution", eds. B. Guiderdoni and A. Kembhavi, Editions de Physique/Springer-Verlag

- Kennicutt R.C., 1998a, ApJ, 498, 541
- Kennicutt R. C., 1998b, ARA&A, 36, 189
- Kent S. M., 1987, AJ, 93, 816
- Kerr F. J., Lynden-Bell D., 1986, MNRAS, 221, 1023
- Kim D.-W., Fabbiano G., Trinchieri G., 1992, ApJS, 80, 645
- Knox L., Scoccamarro R., Dodelson S., 1998, astro-ph/9805012
- Kolatt T. S., Bullock J. S., Somerville R. S. Sigad Y., Jonsson P., Klypin A. A., Primack J. R., Faber S. M., Dekel A., 1999, ApJ, 523, L109
- Koo D. C., Kron R. G., 1992, ARA&A, 30, 613
- Kravtsov A. V., Klypin A. A., Bullock J. S., Primack J. R., 1998, ApJ, 1998, 502
- Kunth D., Mass-Hesse J.M., Terlevich E., Terlevich R., Lequeux J., Fall M. S., 1998, A&A, 334, 11
- Lacey C. G., Silk J., 1991, ApJ, 381, 14
- Lacey C. G., Cole S., 1993, MNRAS, 262, 627
- Lacey C. G., Guiderdoni B., Rocca-Volmerange B., Silk J., 1993, ApJ, 402, 15
- Lacey C. G., Cole S., 1994, MNRAS, 271, 676
- Lahav O., Rees M. J., Lilje P. B., Primack J. R., 1991, MNRAS, 251, 128
- Landy S. D., Szalay A. S., Koo D. C., 1996, ApJ, 460, 94
- Lanzetta K. M., Wolfe A. M., Tunshek D. A., 1995, ApJ, 440, 435
- Le Fèvre O., Hudon D., Lilly S. J., Crampton D., Hammer F., Tresse L., 1996, ApJ, 461, 534
- Leitherer C., Ferguson H., Heckman T. M., Lowenthal J. D., 1995, ApJ, 454, 19
- Lemson G., Kauffmann G., 1999, MNRAS, 302, 111
- Lifshitz E. M., 1946, Journal of Physics, 10, 116
- Lilly et al., 1996, ApJ, 460, L1

- Loveday J., Peterson B. A., Efstathiou G., Maddox S. J., 1992, ApJ, 390, 338
- Loveday J., Maddox S. J., Efstathiou G., Peterson B. A., 1995, ApJ, 442, 457
- Loveday J., 1996, MNRAS, 278, 1025
- Lucy L., 1977, AJ, 82, 1013
- Lumb D., Eggel K., Laine R., Peacock A., 1996, Proceedings of the International Society for Optical Engineering (SPIE), 2808, 326
- Lynden-Bell D., 1967, MNRAS, 136, 101
- Madau P., 1995, ApJ, 441, 18
- Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, MNRAS, 283, 1388
- Madau P., Haardt F., Rees M. J., 1999, ApJ, 514, 648
- Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990, MNRAS, 286, 115
- Marigo P., Bressan A. G., Chiosi C., 1996, A&A, 315, 545
- Martin C., 1999, ApJ, 513, 156
- Marzke R. O., Geller M. J., da Costa L. N., Huchra J. P., 1995, AJ, 110, 477
- Marzke R. O., da Costa L. N., Pelligrini P. S., Willmer C. N. A., Geller M. J., 1998, ApJ, 503, 617
- Mathews W. G., Brighenti F., 1998, ApJ, 503, 15
- Mathewson D. S., Ford V. L., Buchhorn M., 1992, ApJS, 81, 413
- Mewe R., Gronenschild E. H. B. M., van den Oord G. H. J., 1985, A&AS, 62, 197
- Mewe R., Lemen J. R., van den Oord G. H. J., 1986, A&AS, 65, 511
- Miralda-Escudé J., Ostriker J. P., 1990, ApJ, 350, 1
- Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, ApJ, 530, 1
- Mittaz J. P. D., Carrera F. J., Romero-Colmenero E., Mason K. O., Hasinger G., McMahon R., Andernach H., Bower R., Burgos-Martin J., Gonzalez-Serrano J. I., Wonnacott D., 1999, MNRAS, 308, 233

- Mo H. J., Jing Y. P., Börner G., 1993, MNRAS, 264, 825
- Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
- Mo H. J., Jing Y. P., Börner G., 1997, MNRAS, 286, 979
- Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319
- Mobasher B., Sharples R. M., Ellis R. S., 1993, MNRAS, 263, 560
- Monaghan, J. J. 1992, ARA&A, 30, 543
- Moore B., Frenk C. S., White S. D. M., 1993, MNRAS, 261, 827
- Moore B., Frenk C. S., Efstathiou G., Saunders W., 1994, MNRAS, 269, 742
- Moore B., Davis M., 1994, MNRAS, 270, 209
- Moore B., Governato F., Quinn T., Stadel J., Lake G., 1998, ApJ, 499, L5
- Moore B., Quinn T., Governato F., Stadel J., Lake G., 1999, MNRAS, 310, 1147
- Moore B., Gelato S., Jenkins A., Pearce F. R., Quilis V., 2000, ApJ, 535, 21
- Morrison R., McCammon D., 1983, ApJ, 270, 119
- Mould J. R., Huchra J. P., Freedman W. L., Kennicutt R. C., Jr., Ferrarese L., Ford H. C., Gibson B. K., Graham J. A. Hughes S. M. G., Illingworth G. D., Kelson D. D., Macri L. M., Madore B. F., Sakai S., Sebo K. M., Silbermann N. A., Stetson P. B., 2000, ApJ, 529, 786
- Mulchaey J. S., Davis D. S., Mushotzky R. F., Burstein D., 1996, ApJ, 456, 80
- Mushotzky R. F., Szymkowiak A. E., 1988, in "Cooling Flows in Clusters and Galaxies", ed. A. C. Fabian, Kluwer, Dordrecht, p. 53
- Navarro J. F., White S. D. M., 1993, MNRAS, 265, 271
- Navarro J. F., Frenk C. S., White S. D. M., 1995, MNRAS, 275, 720
- Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
- Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
- Navarro J. F., Steinmetz M., 1997, ApJ, 478, 13

- Navarro J. F., Steinmetz M., 1999, ApJ, 513, 555
- Navarro J. F., 2000, astro-ph/9807084
- Nulsen P. E. J., Fabian A. C., 1995, MNRAS, 277, 561
- Nulsen P. E. J., Fabian A. C., 1997, MNRAS, 291, 425
- Oemler A., 1973, PhD thesis, California Institute of Technology
- Ostriker J., Vishniac, E.T. 1986, ApJ, 306, L51
- Osterbrock D. E., 1974, "Astrophysics of Gaseous Nebulae", Freeman, San Francisco, USA
- Owen J. M., Weinberg D. H., Evrard A. E., Hernquist L., Katz N., 1998, ApJ, 503, 160
- Pagel B. E. J., Patchett B. E., 1985, MNRAS, 172, 13
- Park C., Vogeley M. S., Geller M. J., Huchra J. P., 1994, ApJ, 431, 569
- Peacock J. A., Dodds S. J., 1996, MNRAS, 280, L19
- Peacock J. A., Smith R. E., 2000, astro-ph/0005010 (submitted to MNRAS)
- Pearce F. R., Couchman H. M. P., 1997, *New Astronomy*, 2, 411
- Pearce F. R., Jenkins A., Frenk C. S., Colberg J. M., White S. D. M., Thomas P. A., Couchman H. M. P., Peacock J. A., Efstathiou G. (The VIRGO Consortium), 1999, ApJ, 512, L99
- Pearce F. R. et al., 2000, in preparation
- Peebles P. J. E., 1967, ApJ, 147, 859
- Peebles P. J. E., 1968, ApJ, 153, 1
- Peebles P. J. E., 1969, ApJ, 155, 393
- Peebles P. J. E., 1980, "The Large-Scale Structure of the Universe", Princeton University Press, New Jersey, USA
- Peebles P. J. E., 1984, ApJ, 284, 439
- Peebles P. J. E., Juskiewicz R., 1998, ApJ, 509, 483

- Pelligrini S., Ciotti L., 1998, A&A, 333, 433
- Perlmutter S., et al. (The Supernova Cosmology Project), 1999, ApJ, 517, 565
- Phillipps S., Shanks T., 1987, MNRAS, 229, 621
- Ponman T. J., Allan D. J., Jones L. R., Merrifield M., McHardy I. M., Lehto H. J., Luppino G. A., 1994, Nat., 369, 462
- Portinari L., Chiosi C., Bressan A. G., 1998, A&A, 334, 50
- Press W. H., Schechter P., 1974, ApJ, 187, 425
- Ratcliffe A., Shanks T., Parker Q. A., Fong R., 1998a, MNRAS, 293, 197
- Ratcliffe A., Shanks T., Parker Q. A., Fong R., 1998b, MNRAS, 296, 173
- Ratcliffe A., Shanks T., Parker Q. A., Fong R., 1998c, MNRAS, 296, 191
- Raymond J. C., Cox D. P., Smith B. W., 1976, ApJ, 204, 290
- Read A. M., Ponman T. J., Strickland D. K., 1997, MNRAS, 286, 626
- Read A. M., Ponman T. J., 1998, MNRAS, 297, 143
- Rees M. J., 1977, in "Evolution of galaxies and stellar populations", eds. R. B. Larson & B. M. Tinsley, Yale University Observatory Publications, p. 399
- Rees M. J., Ostriker J. P., 1977, MNRAS, 179, 541
- Renzini A., Voli M., 1981, A&A, 94, 175
- Ricotti M., Shull J. M., 1999, astro-ph/9912006 (submitted to ApJ)
- Ritchie B. W., Thomas P. A., 2000, astro-ph/0005357 (submitted to MNRAS)
- Robertson H. P., 1935, ApJ, 82, 284
- Rood H. J., 1974, ApJ, 194, 27
- Roukema B. F., Peterson B. A., Quinn P. J., Rocca-Volmerange B., 1997, MNRAS, 292, 835
- Sage L. J., 1993, A&A, 272, 123
- Sandage A., 1961, ApJ, 134, 916

- Salpeter E. E., 1955, ApJ, 121, 61
- Santiago B. X., da Costa L. N., 1990, ApJ, 362, 386
- Sanders D. B., Mirabel I. F., 1996, ARA&A, 34, 749
- Scalo J. M., 1998, in "The Stellar Initial Mass Function", eds. G. Gilmore & D. Howell, ASP conference Series Vol. 142, p. 201
- Schmidt M., Schnieder D. P., Gunn J. E., 1995, AJ, 110, 68
- Schneider D.P., Schmidt M., Gunn J. E., 1991, AJ, 101, 2004
- Schramm D. N., Turner M. S., 1998, Rev. Mod. Phys., 70, 303
- Seager S., Sasselov D. D., Scott D., 1999, ApJ, 523, L1
- Seljak U., 1996, ApJ, 460, 549
- Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437
- Sellwood J. A., in "Astrophysical Discs — An EC Summer School", Astronomical Society of the Pacific, Conference series Vol. 160, eds. J. A. Sellwood and Jeremy Goodman, 1999, p. 327
- Searle L., Sargeant W. L. W., 1972, ApJ, 173, 25
- Shapiro P.R., Giroux M.L., 1987, ApJ, 321, L107
- Shapley H., 1919, ApJ, 49, 249
- Shapley H., Curtis H. D., 1921, Bulletin of the National Research Council, 2, 217
- Shectman S. A., Landy S. D., Oemler A., Tucker D. L., Lin H., Kirshner R. P., Schechter P. L., 1996, ApJ, 470, 172
- Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
- Sheth R. K., Mo H. J., Tormen G., 1999, astro-ph/9907024
- Silk J. I., 1977, ApJ, 211, 638
- Simien F., de Vaucouleurs G., 1986, ApJ, 302, 564
- Snowden S. L., McCammon D., Burrows D. N., Mendenhall J. A., 1994, ApJ, 424, 714

- Somerville R. S., Primack J. R., Nolthenius R., 1997, ApJ, 479, 606
- Somerville R. S., Davis M., Primack J. R., 1997, ApJ, 479, 616
- Somerville R. S., Lemson G., Kolatt T. S., Dekel A., 1998, astro-ph/9807277 (submitted to MNRAS)
- Somerville & Primack, 1999, MNRAS, 310, 1087
- Spinrad H., Stern D., Bunker A., Dey A., Lanzetta K., Yahil A., Pascarelle S., Fernandez-Soto A., 1998, astro-ph/9809145
- Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelburger K. L., 1996, ApJ, 462, 17
- Steinmetz M., Müller E., 1994, A&A, 281, 97
- Steinmetz M., Müller E., 1995, MNRAS, 276, 549
- Sullivan M., Treyer M. A., Ellis R. S., Bridges T. J., Milliard B., Donas J., 2000, MNRAS, 312, 442
- Sunyaev R. A., Zel'dovich Y. B., 1980, MNRAS, 190, 413
- Sutherland R., Dopita M., 1993, ApJS, 88, 253
- Syer D., Mao S., Mo H. J., 1999, MNRAS, 305, 357
- Tadros H., Efstathiou G., 1996, MNRAS, 282, 1381
- Tegmark M., Peebles P. J. E., 1998, ApJ, 500, 79
- Thacker R. J., Tittley E. R., Pearce F. R., Couchman H. M. P., Thomas P. A., 1999, astro-ph/9809221
- Thomas P. A., Nulsen P. E. J., Fabian A. C., 1987, MNRAS, 228, 973
- Thoul A. A., Weinberg D. H., 1996, ApJ, 465, 608
- Tinsley B. M., 1972, A&A, 20, 383
- Tinsley B. M., 1980, Fundamentals of Cosmic Physics, 5, 287
- Tittley E. R., Couchman H. M. P., Pearce F. R., 1999, astro-ph/9911017 (submitted to MNRAS)

- Tormen G., 1997, MNRAS, 290, 411
- Tormen G., 1998, MNRAS, 297, 648
- Tresse L., Maddox S. J., 1998, ApJ, 495, 691
- Trinchieri G., Kim D. W., Fabbiano G., 1994, ApJ, 428, 555
- Trinchieri G., Fabbiano G., Kim D. W., 1997, A&A, 318, 361
- Tucker D. L., Oemler A., Kirshner R. P., Lin H., Shectman S. A., Landy S. D., Schechter P. L., Muller V., Gottlober S., Einasto J., 1997, MNRAS, 285, L5
- Tully R. B., Fouqué P., 1985, ApJS, 58, 67
- Tumlinson J., Shull J. M., 2000, ApJ, 528, 65
- Valageas P., Silk J., 1999, A&A, 347, 1
- Viana P. T. P., Liddle A. R., 1996, MNRAS, 281, 323
- Vishniac E. T., 1987, ApJ, 322, 597
- Vogel S. N., Weymann R., Rauch M., Hamilton T., 1995, ApJ, 441, 162
- Walker T. P., Steigman G., Kang H., Schramm D. M., Olive K. A., 1991, ApJ, 376, 51
- Warren M. S., Quinn P. J., Salmon J. K., Zurek W. H., 1992, ApJ, 399, 405
- Warren S. J., Hewett P. C., Osmer P. S., 1994, ApJ, 421, 412
- Weil M. L., Eke V. R., Efstathiou G., 1998, MNRAS, 300, 773
- White M., Gelmini G., Silk J., 1995, Phys. Rev. D, 51, 2669
- White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
- White S. D. M., Frenk C. S., Davis M., 1983, ApJ, 274, L1
- White S. D. M., 1984, ApJ, 286, 38
- White S. D. M., Tully R. B., Davis M., 1988, ApJ, 333, 45
- White S. D. M., 1989, in "The Epoch of Galaxy Formation", eds. C. S. Frenk et al., Dordrecht: Kluwer
- White S. D. M., Frenk C. S., 1991, ApJ, 379, 52 (WF91)

- White, S. D. M., Efstathiou, G., Frenk, C. S., 1993, MNRAS, 262, 1023
- White S. D. M., Navarro J. F., Evrard A. E., Frenk C. S., 1993, Nat., 366, 429
- Willmer C. N. A., da Costa L. N., Pellegrini P. S., 1998, AJ, 115, 869
- Wolfire M. G., Mckee C. F., Hollenbach D., Tielens A. G. G. M., 1995, ApJ, 453, 673
- Wood K., Loeb A., 1999, astro-ph/9911316
- Woosley S. E., Weaver T. A., 1995, ApJS, 101, 181
- Wu K. K. S., Nulsen P. E. J., Fabian A. C., 1999, MNRAS, submitted
- Yahil A., Lanzetta K.M., Ferna'ndez-Soto A., 1998, "Galaxies at High Redshifts" in "Large Scale Structure: Tracks and Traces", eds. V. Müller, S. Gottlöber, J. P. Mücke, J. Wambsganss, Proceedings of the 12th Potsdam Cosmology Workshop, Singapore: World Scientific (astro-ph/9803049)
- Young J. S., Allen L., Kenny J. D. P., Lesser A., Rownd B., 1996, AJ, 112, 1903
- Zaritsky D., Kennicutt R. C., Huchra J. P., 1994, ApJ, 420, 87
- Zaritsky D., Smith R., Frenk C. S., White S. D. M., 1997, ApJ, 487, 39
- Zucca E., Zamorani G., Vettolani G., Cappi A., Merighi R., Mignoli M., Stirpe G. M., MacGillivray H., Collins C., Balkowski C., Cayatte V., Maurogordato S., Proust D., Chincarini G., Guzzo L., Maccagni D., Scaramella R., Blanchard A., Ramella M., 1997, A&A, 326, 477
- Zurek W., Quinn P. J., Salmon T. K., Warren M. S., 1994, ApJ, 431, 559
- Zwaan M. A., Briggs F. H., Sprayberry D., Sorar E., 1997, ApJ, 490, 173
- Zwicky F., 1933, Helvetica Phys. Acta., 6, 110

