

Evolution of Galaxies and the Intergalactic Medium at High Redshift

Girish Kulkarni
Harish-Chandra Research Institute

A thesis submitted to the
Board of Studies in Physical Sciences
Homi Bhabha National Institute

In partial fulfillment of requirements for the degree of
Doctor of Philosophy



July, 2011

Chhatnag Road, Jhansi, Allahabad 211019

“ ”

The universe is a procession
with measured and beautiful motion.
— Walt Whitman, *Leaves of Grass*, 1855.

Pandit, your thoughts are all untrue:
there is here no universe and no creator.
— Kabir, *Bijak*, 1515.

Contents

Declaration	xiv
Acknowledgements	xvi
Abstract	xviii
List of Publications	xxiv
1 Introduction	2
1.1 The Standard model	3
1.1.1 Linear perturbations in an expanding Universe	3
1.1.2 Formation and abundance of non-linear objects	7
1.1.3 Key observations	10
1.2 Galaxy formation	12
1.2.1 Gas infall and fragmentation	13
1.2.2 Galactic structure and evolution	17
1.2.3 Modelling and key observations	18
1.3 Inter-galactic medium	20
1.3.1 Absorption systems	20
1.3.2 Epoch of Reionization	24
1.3.3 Modelling the IGM	30
2 Metal enrichment and reionization constraints on early star formation	34
2.1 Observations	36
2.1.1 Metallicity of the IGM	37
2.1.2 CMB constraints on IGM reionization	38
2.2 Analytical Model	38
2.2.1 Reionization	39

2.2.2	Metal Enrichment	42
2.3	Results	44
2.3.1	Constraints on IMF evolution	46
2.3.2	Constraints on high mass star formation	47
2.4	Conclusions	49
3	Reionization and feedback in overdense regions at high redshift	52
3.1	Description of the Analytical model	55
3.1.1	Globally averaged reionization	55
3.1.2	Biased regions	59
3.1.3	Radiative feedback	63
3.2	Results	64
3.2.1	Effect of overdensity on reionization history	65
3.2.2	Effect of overdensity on luminosity function	65
3.2.3	Luminosity function as a probe of reionization	66
3.3	Discussion and Summary	67
A	Formation rate and survival probability of haloes in overdense regions	70
4	Formation rate of Dark Matter Haloes	74
4.1	Rate of Halo Formation	77
4.1.1	Sasaki prescription: Press-Schechter mass function	78
4.1.2	Sasaki prescription: Sheth-Tormen mass function	80
4.1.3	Excursion set approach to halo formation rates: Press-Schechter mass function	81
4.1.4	Excursion set approach to halo formation rates: Sheth-Tormen mass function	84
4.2	N-Body simulations	87
4.3	Results and Discussion	89
4.3.1	Halo destruction rate efficiency	89
4.3.2	Halo formation rate	94
4.3.3	Halo survival probability	101
4.3.4	Formation time distribution	101
4.3.5	Discussion	102
4.4	Conclusions	103
5	Formation of galactic nuclei with multiple supermassive black holes at high redshifts	106
5.1	Previous work	109
5.2	Formation of multiple-SMBH systems	111
5.2.1	Time scale of incoming SMBHs	111

5.2.2	Binary SMBH coalescence time scale	115
5.3	Evolution of multiple SMBHs	119
5.4	Simulations	120
5.5	Results	128
5.5.1	Dynamics of single and binary SMBHs	129
5.5.2	Evolution of nuclei with multiple SMBHs	137
5.5.3	Likelihood of nuclei with multiple SMBHs at high redshift . . .	139
5.5.4	Effects on the stellar distribution	140
5.6	Observational Signatures	142
5.7	Conclusions	144
6	Post-reionization cosmological H I distribution in a hierarchical galaxy formation model	148
6.1	Semi-analytic model for galaxy formation	151
6.1.1	N-body simulations	151
6.1.2	Galaxy formation	153
6.1.3	Calibration	162
6.2	Results	163
6.2.1	Comparison with local observations	164
6.2.2	Global HI distribution	165
6.2.3	Fluctuations in HI	167
6.3	Discussion and conclusions	172

List of Figures

1.1	Evolution of the 21 cm spin temperature.	27
1.2	Spectrum of the 21 cm brightness temperature contrast	28
1.3	Constraints on reionization	29
2.1	Fiducial dependence of Metal and UV emissivity on n_s and τ	40
2.2	Fiducial dependence of Metal emissivity on τ and σ_8	40
2.3	Fiducial dependence of UV emissivity on n_s and σ_8	41
2.4	Metal and UV emissivities for $M_{\text{low}} = 0.5 M_{\odot}$	43
2.5	Metal and UV emissivities for $Z_{\text{input}} = 0.001$	43
2.6	Metal and UV emissivities for $Z_{\text{input}} = 0.001$ and $M_{\text{low}} = 0.1 M_{\odot}$	44
2.7	Evolution of the IGM ionization fraction	45
2.8	Constrained ratio of emissivities in our model	48
3.1	IGM evolution in average and overdense regions	54
3.2	Luminosity function at $z = 6$ and 7	58
3.3	Luminosity functions with observations at $z = 8$	58
3.4	Evolution of threshold halo mass for star formation	60
3.5	Effect of reionization history on the luminosity function	63
3.6	Effect of overdensity on the luminosity function	64
4.1	Destruction rate efficiency for Press-Schechter mass function	83
4.2	Formation rate efficiency for Press-Schechter mass function	86
4.3	Destruction rates for $n = -0.5$: comparison with Sasaki formalism	90
4.4	Formation rates for $n = -0.5$: comparison with Sasaki formalism	91
4.5	Destruction rates for $n = -1.5$: comparison with Sasaki formalism	92
4.6	Formation rates for $n = -1.5$: comparison with Sasaki formalism	93
4.7	Destruction rates for the Λ CDM model	94
4.8	Formation rates for the Λ CDM model	95
4.9	An alternate method for formation rates	96

4.10	Comparison of the two methods	96
4.11	Formation rates for $\epsilon = 0.1$	97
4.12	Ratio of formation rates for $\epsilon = 0.1$	97
4.13	Halo survival probability	98
4.14	Halo formation epoch distribution	99
4.15	Less obvious differences between two approaches	100
5.1	Merger time scales for haloes and galaxies	113
5.2	Plausibility of systems with multiple SMBHs	116
5.3	Example merger tree from the Millennium simulation	120
5.4	Evolution of a single SMBH	123
5.5	Evolution of a binary SMBH	124
5.6	Evolution of number of SMBHs	129
5.7	SMBH ejection velocities	130
5.8	Evolution of the bulge escape velocity	131
5.9	Some sample runs of our simulation	132
5.10	Effect of binary SMBH on stellar density	133
5.11	Effect of multiple SMBHs on stellar density	134
5.12	Likelihood of occurrence of multiple SMBHs: low mass galaxies	135
5.13	Likelihood of occurrence of multiple SMBHs: high mass galaxies	136
5.14	Evolution of the bulge density profile	138
5.15	Bulge mass deficiency	141
6.1	B -band luminosity function in the model	152
6.2	Comparison of Tully-Fisher relation with observations	154
6.3	Evolution of the cosmic H I density	157
6.4	Star formation rate density	159
6.5	Cold gas mass of galaxies	161
6.6	Predicted circular velocity-cold gas mass relation at $z = 0$ and 1.	163
6.7	Cold gas fraction of haloes in our simulation	166
6.8	Power spectra in our model at $z = 3.34$	168
6.9	Scale-dependent H I bias	169
6.10	Stochasticity of H I bias	170
6.11	Evolution of redshift-space power spectra and linear bias	171

List of Tables

2.1	Various IMFs considered in our analytical model	36
4.1	Simulations used in this chapter: power law spectra	86
4.2	Simulations used in this chapter: Λ CDM spectra	87
5.1	Summary of results for $M_0 \sim 10^{14} M_\odot$	121
5.2	Summary of results for $M_0 \gtrsim 10^{15} M_\odot$	122
6.1	N-body simulations used for our semi-analytic model	154

Certificate

This is to certify that this Ph. D. thesis titled 'Evolution of Galaxies and the Intergalactic Medium at High Redshift,' submitted by Girish Kulkarni is a record of bona fide research work done under my supervision. It is further certified that this thesis represents independent work by the candidate and collaboration was necessitated by the nature and scope of the problems dealt with.

Date:

Prof. J. S. Bagla
Thesis Advisor

Declaration

This thesis is a presentation of my original research work. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions. The work is original and has not been submitted earlier as a whole or in part for a degree or diploma at this or any other Institution or University. This work was done under supervision of Prof. J. S. Bagla, at Harish-Chandra Research Institute, Allahabad.

Date:

Girish Kulkarni
Ph. D. Candidate

Acknowledgements

Not surprisingly, this thesis has benefitted immensely from support from diverse sources. This work was done under the supervision of Prof. J. S. Bagla at Harish-Chandra Research Institute (HRI). I thank him for the numerous helpful discussions and overall supervision. He patiently let me work in my own way and this thesis has gained much from that. His high-performance computing cluster at HRI has been used for most numerical calculations employed here. A portion of this work was done under the supervision of Prof. Abraham Loeb at the Institute of Theory and Computation (ITC), Harvard University. Avi was miraculously available for discussion any time I wanted, and inspired me with his creativity. Some work was also done in collaboration with Tirthankar Roy Choudhury, from whom I have learnt a lot of astrophysics. Tirth also taught me how to be rigorous in this subject and his advice on several matters has been valuable. I also thank Nishikanta Khandai, Jaswant Yadav, and Sourav Mitra for collaboration on parts of this research.

Less direct, but perhaps more crucial, has been the extraordinary support from individuals during my stay by the Ganga and by the Charles. In Allahabad, various discussions with Nighat Gandhi, Harvinder Kaur Jassal, L. Sriramkumar, Jaya Umadikar, Pinaki Majumdar, Dileep Jatkar, and Sumathi Rao bailed me out more than once during my work on this thesis. As did the hours spent talking with Rajarshi Tiwari on the art of computer programming, with Nishita Desai on the craft of research, and with all other friends on all kinds of conundrums. At Harvard, Michael Blake, Claire Shindler, Yue Shen, Jonathan Pritchard, Akshay Kulkarni, Pamela Petras and Gongjie Li did the same for me, for which I am grateful. Thanks are also due to the administrative staff at HRI, ITC and Harvard for various facilities. I thank HRI for financially supporting this work. Part of it was also supported by a Fulbright-Nehru fellowship from the US-India Educational Foundation. Lastly, at the risk of sounding trite, I thank my parents, Revati and Pramod Kulkarni, for their love and support, without which I would not have finished this thesis.

Abstract

Research presented in this thesis follows two threads in the broad area of cosmology: (1) properties of the intergalactic medium (IGM), and (2) formation of galaxies and their evolution.

Mainly from observations of absorption systems in high-redshift quasar spectra, the intergalactic medium has been inferred to be ionized, as well as chemically enriched, up to redshift of $z \simeq 6$. It is natural to imagine an epoch of reionization in which the post-recombination IGM was ionized due to radiation from star-forming galaxies, which also enriched the IGM. Reionization has been the focus of large amount of theoretical and observational research in the last decade, much of which has modelled it in various cosmological and structure formation scenarios. With this in mind, this thesis deals with the following issues: (1) the inverse problem of constraining early star formation from reionization-related observations; and (2) developing self-consistent models of reionization and suggesting new observables. In particular, we study reionization and reheating of the IGM in overdense regions to probe the role of overdensities in observations of luminosity function of high redshift galaxies.

In addition to IGM evolution, understanding how observed small scale structure, including galaxies, can emerge in the well-accepted Cosmological Constant-Cold Dark Matter (Λ CDM) cosmological model is an important open problem. First models of how galaxy formation occurs in an expanding universe were studied in the 1970s [White & Rees, 1978; Binney, 1977]. Such models are dealt with in the second thread of this thesis. We have developed a numerical code that implements a semi-analytic model of galaxy formation on top of a dark matter N-body simulation. This model includes mergers of dark matter haloes, mergers of galaxies, cooling of baryons, formation of satellite galaxies, and consistent population synthesis, apart from information about spatial distribution. This thesis includes (1) predicting the evolution of the neutral hydrogen content and its large scale distribution in the universe over a range of redshifts, and (2) a study of assembly of supermassive black holes in nuclei of high redshift galaxies.

Following sections describe above work in more detail. Publications included in the

thesis are listed later.

Evolution of the IGM

Several observations, primarily the absence of a Gunn-Peterson trough in spectra of objects up to $z \approx 6$, indicate that the IGM is devoid of neutral hydrogen in this redshift range [Fan, 2006]. Furthermore, intervening absorption systems with elements heavier than Helium, most commonly Carbon, Nitrogen, Silicon and Iron, but also others, have been observed up to this high redshift [Songaila, 2001]. The observed evolution of the ionization state and chemical enrichment of the IGM at high redshift is an important clue about the origin of ionizing radiation and metals. In order to study this, we build analytic models incorporating the relevant physics, and draw constraints from available observations.

IGM reionization and enrichment

We use a simplified approach for studying formation of stars in collapsed haloes and the resulting ionization and enrichment of the IGM. We consider a set of Λ CDM models allowed by observations of CMB temperature and polarization anisotropies for this study and constrain parameters related to star formation with the help of observations. We constrain subsets of these parameters independently by using the observed metallicity of the IGM at $z \approx 5$ and the requirement that the Thomson scattering optical depth due to an ionized IGM as determined for the model from CMB observations be reproduced. We consider a range of initial metallicities for star forming gas, and variations of the initial mass function (IMF) of stars.

We find that a ‘normal’ initial mass function (IMF) may satisfy these two constraints with a raised efficiency of star formation as compared to that seen in the local universe. We also find that observations require a significant fraction of metals to escape from haloes into the IGM. We can also place constraints on the ratio of escape fraction for metals and ionizing photons and find that this ratio is of order unity for most models. Ultra-high mass stars or AGNs may not simplify models of reionization in that these may produce more ionizing photons but do not contribute to production of metals and hence these help in reducing only the escape fraction for ionizing photons. However, suppression of very low mass stars is helpful in that it increases the production of metals as well as ionizing photons and hence leads to a reduction in both escape fractions. Such a change is also warranted by observations of metal poor halo stars in the Galaxy [Bagla, Kulkarni & Padmanabhan, 2009].

The most important conclusion of this work is that star formation without a significant evolution of the IMF is sufficient for satisfying the two constraints considered here. The escape fractions, and/or the star formation efficiency is required to be higher than

what we see in local galaxies. One can consider other sources of ionizing radiation, indeed at least some of these must be present. But as we have pointed out, these help in reducing only the escape fraction for ionizing radiation as none of the other potential sources help in transporting enriched material from the interstellar medium (ISM) to the IGM. This highlights the significance of the constraint arising from IGM enrichment for epoch of reionization studies.

Probing radiative feedback at high redshift

Observations of galaxy luminosity function at high redshifts typically focus on fields of view of limited size preferentially containing bright sources. These regions possibly are overdense and hence biased with respect to the globally averaged regions. We study the reionization and thermal history of the universe in overdense regions by refining our reionization model described above to include several physical processes, following earlier work [Choudhury & Ferrara, 2005].

We find that reionization proceeds differently in overdense regions. Overdense regions are ionised earlier because of enhanced number of sources and star formation. In addition, these regions have higher temperatures because of enhanced recombinations and hence the effect of radiative feedback is enhanced too. In particular, the shape of the galaxy luminosity function for biased regions is very different from that for average regions. There is a significant enhancement in the number of high-mass galaxies because of bias, while there is a reduction in low-mass galaxies resulting from enhanced radiative feedback. Also, because of the enhanced feedback, luminosity function in overdense regions is more sensitive to reionization history compared to average regions. The effect of radiative feedback shows up at absolute AB magnitudes $M_{AB} \gtrsim -17$ in these regions, while it occurs at much fainter magnitudes $M_{AB} \sim -12$ for average regions [Kulkarni & Choudhury, 2011]. This order of magnitude change, visible for absolute AB magnitude $M_{AB} \gtrsim -17$ at $z = 8$ in the overdense-region luminosity function should be detectable with the James Webb Space Telescope (JWST) in near future. This will serve as an additional probe of radiative feedback and hence reionization at high redshifts.

Galaxy formation and evolution

Formation of galaxies is a crucial ingredient of models of reionization. In the models described above, galaxy formation is implemented by calculating the mass function of dark matter haloes (number of haloes per unit volume as a function of halo mass) and their formation rates (number of haloes created per unit volume per unit time), and accounting for the baryonic processes like cooling and feedback to populate haloes with galaxies. In this thesis, we study the problem of calculating formation rate of

haloes for the Sheth-Tormen form of the mass function. We calculate this rate of haloes and compare it with N-body simulations. Next, in order to calculate various quantities beyond their global average, as is done in the above models, we implement a detailed semi-analytic model of galaxy formation on top of a dark matter N-body simulation. Finally, we also consider the problem of assembly of supermassive black holes in these galaxies.

Mass function and formation rates of dark matter haloes

We consider the issue of deriving analytic estimates of formation rates of dark matter haloes [Mitra et al., 2011]. The commonly used Sasaki prescription [Sasaki, 1994] gives unphysical results when applied to the Sheth-Tormen form of mass function. We develop a new prescription to calculate halo formation rate, using excursion set formalism but avoiding the assumption of scale invariance of halo destruction rate efficiency made by Sasaki. In our prescription, we introduce a parameter ϵ , the smallest fractional change in mass of a halo before we consider it as destruction of the old halo and formation of a new halo. We show that the halo destruction rate is not independent of mass even for power law models and hence the basis for the Sasaki ansatz does not hold. Two prominent features of the halo destruction rate are the rapid fall at large masses, and a pronounced peak close to the scale of non-linearity. The peak is more pronounced for smaller values of ϵ . Using the excursion set approach for the Sheth-Tormen mass function leads to positive halo formation rates, unlike the generalization of the Sasaki ansatz where formation rate at some mass scales is negative.

We compare the destruction rate and the halo formation rate computed using the excursion set approach with N-Body simulations. We find that our approach matches well with simulations for all models, at all redshifts, and also for different values of ϵ . In some cases there are deviations between the simulations and the theoretical estimate. However, these deviations are much smaller for the excursion set based method as compared to the Sasaki method. We also study sources of these small deviations.

Post-reionization H I distribution

In addition to considering globally averaged analytical models of the kind discussed above, we also perform semi-analytic simulations of galaxy formation that allow us to go beyond the average and consider, for example, the effect of clustering on galaxy properties.

We have developed a semi-analytic model of galaxy formation that takes many physical processes into account by using the output of dark matter N-body simulations. Merger trees of dark matter haloes are obtained from the simulations and baryonic physics is then implemented in each halo. This involves models of star formation and evolution, gas heating and cooling, supernova feedback, dynamical friction effects on

satellite galaxies, and major and minor mergers of galaxies. As a first application of this model, we study the distribution of neutral hydrogen (H I) in post-reionization universe. Whereas almost all of the matter in pre-reionization universe is in the form of neutral hydrogen, much of the hydrogen in the IGM is ionized during the epoch of reionization. As a result, after reionization, most neutral hydrogen is confined to the ISM of galaxies and in small-scale, partially ionized clumps in the IGM. This shows that a consistent model of galaxy formation is valuable in understanding distribution post-reionization neutral hydrogen. In turn, this will be important for future observations of the 21cm power spectrum with experiments like Low Frequency Array (LOFAR) and the interferometric array at Ooty Radio Telescope (ORT) which aim to observe fluctuations in this distribution. (Post-reionization observations with the 21cm line correspond roughly to observations at frequencies higher than 100 MHz.)

This problem has been tackled in the literature before [Bagla, Khandai & Datta, 2010]. However, neutral hydrogen has been prescriptively, rather than self-consistently, assigned to host dark matter haloes of galaxies. Thus most current work does not take key processes of baryonic evolution into account. Some work exists on distribution of cold gas at low redshifts ($z \lesssim 2$) using models of galaxy formation [Kim et al., 2011]. We extend these results to higher redshifts to understand what future observations of neutral hydrogen will imply for galaxy evolution.

Growth of SMBHs in galactic nuclei

Supermassive black holes (SMBHs) are known to exist in the bulge components of almost every massive galaxy in the local universe. These SMBHs have interesting correlations with the properties of the bulge that they inhabit. Moreover, observations of $z \approx 6$ quasars imply that SMBHs already existed at that high redshift. Still, the assembly of these SMBHs is an ill-understood subject.

We study the dynamical aspect of SMBH assembly by examining the formation of groups of multiple SMBHs in gas-poor galactic nuclei due to the high merger rate of galaxies at high redshifts. With simple estimates of the characteristic time scales, we show that systems with more than two SMBHs are generally expected to exist in the merger history of a Milky Way-mass galaxy for $z \gtrsim 1$. We then calculate the relative likelihood of binary, triple, and quadruple SMBH systems, by considering the timescales for relevant processes and combining merger trees with accurate direct summation N-body simulations for the dynamics of stars and SMBHs in galactic nuclei. We show that halos that have mass $\approx 10^{15} M_{\odot}$ at $z = 0$ will generally have more than two closely interacting SMBHs at around $z \approx 5$. We study the dynamics of these systems and show that many of them can survive for several Myr before slingshot effects and gravitational wave recoil deplete them. Most of these high mass galaxies are left with a single SMBH at $z = 0$. The existence of multiple SMBH systems leads to an enhanced rate of tidal disruption of stars, to modified gravitational wave signals compared to

isolated BH binaries, and to slingshot ejection of SMBHs from galaxies at high speeds.

High mass galaxies ($M \gtrsim 10^{14} M_{\odot}$ at $z = 0$) are generically expected to have had compact multiple SMBH systems in their nuclei during their assembly history. Although they undergo major mergers, galaxies comparable to the Milky Way in mass ($\approx 10^{12} M_{\odot}$) rarely have more than two SMBHs in their nuclei at any moment in their assembly history. Both categories of galaxies are rarely expected to retain more than two SMBHs in their nuclei at the present epoch. Within our model, SMBH coalescence is common and in galaxies with smaller mass, the subsequent recoil due to anisotropic gravitational wave emission often results in escaping SMBHs. Some of these SMBHs add to the wandering population of black holes in the galactic halo. In a few cases, this process also results in galactic nuclei with no SMBH near their centres. In larger galaxies, BH-BH interaction can also lead to escaper SMBHs due to the slingshot mechanism. Systems with more than two SMBHs seldom last for more than a relaxation time. These systems reduce to those with two, one, or zero black holes via slingshot escapes, coalescences and gravitational wave recoils. We estimate the signature of such systems on bulge profiles, left because of bulge heating and scouring.

Conclusion

Chief contribution of this thesis is (1) to state conditions under which observed metal enrichment of IGM at high redshift is consistent with accepted reionization scenarios; (2) to provide a new, independent observable to constrain reionization history and feedback at high redshift; (3) to develop a better analytical technique to calculate formation rate of dark matter haloes; (4) in using accurate N-body simulations to understand dynamical effects associated with SMBH assembly in galactic nuclei, and (5) to understand post-reionization distribution of HI using a consistent semi-analytic model of galaxy formation.

This work can be extended further in several ways. One possible extension is to incorporate outflows in our semi-analytic model of galaxy formation to study IGM enrichment. This will set earlier models of enrichment as a result of wind-blown bubbles [Furlanetto & Loeb, 2003; Samui, Subramanian & Srianand, 2008] in a simulation of large scale structure. This is likely to result in an understanding of effects of a multi-phase interstellar medium and clustering of galaxies. A second possible extension is to model luminosity functions of quasars and galaxies at high redshifts [Shen, 2009] simultaneously. This follows naturally from our work on galaxy luminosity functions [Kulkarni & Choudhury, 2011] but involves aspects like black hole assembly and accretion that we have not considered yet like properties of seed black holes, their merger rates, and dynamical effects like three-body interactions. The chief motivation behind this is the discovery of luminous quasars at redshifts $z > 6$ [Fan, 2006; Mortlock et al., 2011].

Publications

This thesis is mainly based on the following publications:

1. Bagla J. S., Girish Kulkarni, T. Padmanabhan; *Metal Enrichment and Reionization Constraints on Early Star Formation*, Monthly Notices of the Royal Astronomical Society 2009 **397**, 971–978.
2. Kulkarni Girish, T. Roy Choudhury; *Reionization and feedback in overdense regions at high redshift*, Monthly Notices of the Royal Astronomical Society 2011 **412**, 2781–2789.
3. Mitra Sourav, Girish Kulkari, J. S. Bagla, Jaswant K. Yadav; *Formation rates of dark matter haloes*, Bulletin of the Astronomical Society of India (Submitted; Preprint: arXiv:1103.5828).
4. Kulkarni Girish, Abraham Loeb; *Formation of galactic nuclei with multiple supermassive black holes at high redshift*, Monthly Notices of the Royal Astronomical Society (Submitted; Preprint: arXiv:1107.0517).
5. Kulkarni Girish, J. S. Bagla, Jaswant K. Yadav; *HI distribution in the post-reionization Universe: hierarchical galaxy formation models*, Manuscript under preparation.

Introduction

The problem of how galaxies and the inter-galactic medium (IGM) form and evolve has been described as the “final frontier of cosmology” [e.g. Bromm & Loeb, 2007]. Last two decades have seen significant advances in observational astronomy. Firstly, the geometry, and hence the energy content, of the universe is now well-constrained with observations of the cosmic microwave background (CMB) radiation [de Bernardis et al., 2000; Komatsu et al., 2011]. These constraints are further improved by observations of clustering of galaxies, high redshift supernovae, weak gravitational lensing and the Lyman- α forest [e.g. Seljak, Slosar & McDonald, 2006]. Secondly, increasingly accurate measurement of anisotropies in the CMB brightness temperature and polarization have confirmed the idea that the large scale structure in the universe formed as a result of evolution of small-amplitude density fluctuations at early times [Bennett et al., 1996; Larson et al., 2011]. Ever powerful surveys of bound objects—galaxies and X-ray clusters of galaxies—seem to suggest that abundance of such objects can be explained by suitably extrapolating the CMB anisotropies to small scales [Lilly et al., 1996; Adelman-McCarthy et al., 2008]. Thirdly, observations have probed conditions in the universe at increasingly higher redshift through the detection of galaxies and quasars. Current highest redshift quasar observation is at $z \sim 7$ [Mortlock et al., 2011] and the spectra of several $z \sim 6$ quasars have been well studied to constrain the IGM evolution [Fan, 2006]. The most distant γ -ray burst (GRB) candidate is reported at $z \sim 8.2$ [Tanvir et al., 2009; Salvaterra et al., 2009], while the highest redshift galaxy candidate is at $z \sim 10$ [Bouwens et al., 2011]¹.

This third development is expected to be particularly instrumental in helping us understand galaxy formation and evolution of the IGM. Crucial problems here are to understand when and how the IGM was reionized, how did it get chemically enriched

¹For comparison, until 1990, the most distant known quasars were at redshifts $z \lesssim 4$, galaxies were at $z \lesssim 2$ and GRBs were essentially only at $z \sim 0$. See Zhang [2009] for a timeline of redshift record-breakers.

and how do galaxies observed in the local universe result from the standard cosmological model. Research presented in this thesis follows these threads in the broad area of cosmology. In this chapter, we introduce this subject and place it in the context of the Lambda Cold Dark Matter (LCDM) cosmological model. We also review key observations in this field.

1.1 The Standard model

Our current understanding of the universe can be collected into a set of principles that has come to be called the Standard Model of cosmology [Peebles, 1993; Peacock, 1999]. It has gathered empirical confirmation from a variety of independent observations. In this picture, the universe had its origin in a singularity—the “Big Bang”—of infinite density and temperature about 13.7 billion years ago. This was immediately followed by a period of accelerated expansion called inflation that lasted for just about 10^{-34} s but increased the size of the universe by about 60 e-foldings. At the end of inflation, the universe was highly homogeneous on large scales, but had small-scale inhomogeneities that evolved out of quantum fluctuations. Its energy density was dominated by the contribution from relativistic matter (radiation and neutrinos). Radiation was in equilibrium with the non-relativistic matter via Compton and free-free scattering. The universe was also expanding. As a result, its temperature was falling with redshift z as $(1+z)$. When the temperature dropped to ~ 3000 K at $z \sim 1100$, protons and electrons combined to form hydrogen atoms, and radiation decoupled from matter. This gave rise to the CMB, which is said to have emerged from a “last scattering surface.” At this time, the energy density of the universe was dominated by non-relativistic matter, and the CMB carries an imprint of the inhomogeneities of the last-scattering surface. In this section, we begin by introducing elements of the Standard Model that deal with the evolution of these inhomogeneities.

1.1.1 Linear perturbations in an expanding Universe

Einstein argued theoretically that the distribution of matter and radiation in the universe should be isotropic and homogeneous at large scales. This is known as the cosmological principle [Peebles, 1993; Peacock, 1999]. Isotropy is now well-established thanks to observations of the CMB, and of optically-selected galaxies, the X-ray background, and faint radio sources [Wu, Lahav & Rees, 1999]. Although evidence for homogeneity is weaker, isotropic and inhomogeneous cosmological are observationally excluded [Goodman, 1995; Yadav et al., 2005; Bagla, Yadav & Seshadri, 2008; Sarkar et al., 2009].

Kinematics in the most general isotropic and homogeneous space is described by

the line element

$$ds^2 = dt^2 - a^2(t) \left[\frac{dR^2}{1 - kR^2} + R^2(d\theta^2 + \sin^2\theta d\phi^2) \right], \quad (1.1)$$

where $a(t)$ is called the cosmological scale factor and (R, θ, ϕ) are spherical comoving coordinates. The scale factor describes the expansion of the universe and the constant k determines its geometry. As a result, observers at rest remain at rest at fixed (R, θ, ϕ) while their physical separation changes with time in proportion to $a(t)$. A given observer sees a nearby observer at physical distance D receding at the Hubble velocity $H(t)D$, where $H(t) = \dot{a}(t)/a(t)$ is called the Hubble parameter. Light emitted by a source at time t is observed at $t = 0$ (“today”) with a redshift $z = 1/a(t) - 1$, where we have set $a(0) = 1$.

Dynamics of matter and radiation in the space-time described by Equation (1.1) is described by the Einstein field equations of general relativity, which for the metric in Equation (1.1) take the form

$$H^2(t) = \frac{8\pi G}{3}\rho - \frac{k}{a^2}. \quad (1.2)$$

This equation is known as the Friedmann equation [Freidmann, 1922; Weinberg, 1972]. It relates the expansion of the universe to its matter-energy content. For each component of the energy density ρ , with an equation of state $p(\rho)$, the density ρ varies with time to conserve energy

$$d(\rho R^3) = -pd(R^3). \quad (1.3)$$

Using this, we can now recast the Friedmann equation as

$$\frac{H(t)}{H_0} = \left[\frac{\Omega_m}{a^3} + \Omega_\Lambda + \frac{\Omega_r}{a^4} + \frac{\Omega_k}{a^2} \right]^{1/2}. \quad (1.4)$$

Here, for $i = m, \Lambda$, and r , we define $\Omega_i = \rho_i/\rho_c$, where $\rho_c = 2H^2(t)/8\pi G$ is a parameter called the critical density. The quantity Ω_i is called the cosmological density parameter of species i . We also define H_0 as the value of $H(t)$ at $t = 0$ and Ω_0 as the value of $\Omega_m + \Omega_\Lambda + \Omega_r$ at $t = 0$. The quantity Ω_k is given by

$$\Omega_k = 1 - \Omega_0. \quad (1.5)$$

A cosmological model with $\Omega_m = 1$ and $\Omega_\Lambda = \Omega_r = 0$ is particularly simple. It is known as the Einstein-de Sitter model. Although it is observationally ruled out, other models with non-zero Ω_Λ approach the Einstein-de Sitter model at high redshift.

The most-favoured cosmological model according to current observations has $\Omega_m = 0.27 \pm 0.01$, $\Omega_\Lambda = 0.726 \pm 0.015$ and $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$, where $h = 0.705 \pm 0.013$ [Komatsu et al., 2011].

As mentioned before, observations of the CMB show that the universe was extremely uniform at the last scattering surface, but had spatial fluctuations in the energy density of roughly one part in 10^5 . These fluctuations grow over time due to gravitational instability. In the standard model, formation of all structure in the universe—like galaxies and clusters of galaxies—is attributed to these fluctuations. In order to describe these fluctuations, we now distinguish between physical and comoving coordinates. In vector notation, the physical coordinate \mathbf{r} corresponds to the comoving coordinate $\mathbf{x} = \mathbf{r}/a$. We describe the matter content of the universe as an ideal pressure-less fluid of particles each of which is at fixed \mathbf{x} , expanding with the Hubble flow $\mathbf{v} = H(t)\mathbf{r}$ where $\mathbf{v} = d\mathbf{r}/dt$. Onto this uniform expansion, we impose small fluctuations, given by a relative density perturbation

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{r})}{\bar{\rho}} - 1, \quad (1.6)$$

where the mean fluid density is $\bar{\rho}$. Then the fluid is described by the continuity and Euler equations in comoving coordinates

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot [(1 + \delta)\mathbf{u}] = 0 \quad (1.7)$$

$$\frac{\partial \mathbf{u}}{\partial t} + H\mathbf{u} + \frac{1}{a}(\mathbf{u} \cdot \nabla)\mathbf{u} = -\frac{1}{a}\nabla\phi. \quad (1.8)$$

The gravitational potential ϕ used above is given by the Poisson equation, in terms of the density perturbation

$$\nabla^2\phi = 4\pi G\bar{\rho}a^2\delta. \quad (1.9)$$

Note that we are now working in the Newtonian approximation, which is valid since the relevant scales are much smaller than the Hubble scale. Observations imply that the non-relativistic matter in the universe is composed of cold, weakly-interacting massive dark matter particles in addition to the normal baryonic matter. Therefore, to understand the evolution of the perturbations introduced above, we have to solve the above three equations for baryons and dark matter separately. The fluid description is then valid only until different particle streams cross. This “shell crossing” typically occurs only after perturbations have grown to become non-linear. Similarly, baryons can be described as a pressure-less fluid only so long as their temperature is negligibly small.

For small perturbations $\delta \ll 1$, the fluid equations can be linearized and combined to yield

$$\frac{\partial^2 \delta}{\partial t^2} + 2H\frac{\partial \delta}{\partial t} = 4\pi G\bar{\rho}\delta. \quad (1.10)$$

This linear equation has two independent solutions in general, only one of which grows with time. Starting with random initial conditions, the “growing mode” comes to dominate the density evolution. As a result, the density perturbation maintains its shape in

comoving coordinates and grows in proportion to a growth factor $D(t)$. The decaying solution is given simply by the Hubble parameter $H(t)$. As a result, the growing mode can be obtained using the Wronskian [Heath, 1977]. It is given by

$$D(t) \propto \frac{(\Omega_\Lambda a^3 + \Omega_k a + \Omega_m)^{1/2}}{a^{3/2}} \int^a \frac{a^{3/2} da}{(\Omega_\Lambda a^3 + \Omega_k a + \Omega_m)^{3/2}}, \quad (1.11)$$

where we have neglected Ω_r [Peebles, 1980]. In the Einstein-de Sitter universe, the growth factor is proportional to the scale factor $a(t)$.

The density fluctuations $\delta(\mathbf{x})$ can be described in Fourier space, in terms of Fourier components

$$\delta_{\mathbf{k}} = \int d^3x \delta(\mathbf{x}) \exp(-i\mathbf{k} \cdot \mathbf{x}), \quad (1.12)$$

where \mathbf{k} is the comoving wavenumber. Fourier description is particularly useful because we assume that perturbations in the universe are generated by inflation [Kolb & Turner, 1990]. Inflation generates perturbations given by a Gaussian random field, in which \mathbf{k} -modes are statistically independent, each with a random phase. The statistical properties of the fluctuations are determined by the variance of the different \mathbf{k} -modes, and the variance is described in terms of the power spectrum $P(k)$ as follows

$$\langle \delta_{\mathbf{k}} \delta_{\mathbf{k}'}^* \rangle = (2\pi)^3 P(k) \delta^{(3)}(\mathbf{k} - \mathbf{k}'), \quad (1.13)$$

where $\delta^{(3)}$ is the three-dimensional Dirac delta function. In the Standard Model, inflation produces a primordial power-law spectrum $P(k) \propto k^n$ with $n \sim 1$. Growth of perturbations in the radiation-dominated and then the matter-dominated universe results in a modified final power spectrum, characterized by a turnover at a scale of order the horizon c/H at the matter-radiation equality, and a small-scale asymptotic shape of $P(k) \propto k^{n-4}$. On large scales the power spectrum evolves in proportion to the square of the growth factor and this simple evolution is termed linear evolution. On small scales, the power spectrum changes shape due to the additional non-linear gravitational growth of perturbations, yielding the full, non-linear power spectrum. The overall amplitude of the power spectrum is not specified by current models of inflation, and it is usually set observationally from CMB and observations of galaxy clustering.

Since density fluctuations exist on all scales, in order to determine the formation of objects of a given size or mass, it is useful to consider the statistical distribution of the *smoothed* density field. Using a window function $W(\mathbf{y})$ normalized so that $\int d^3\mathbf{y} W(\mathbf{y}) = 1$, the smoothed density perturbation field, $\int d^3\mathbf{y} \delta(\mathbf{x} + \mathbf{y}) W(\mathbf{y})$, itself follows a Gaussian distribution with zero mean. For the particular choice of a spherical top-hat, in which $W = 1$ in a sphere of radius R and is zero outside, the smoothed perturbation field measures the fluctuations in the mass in sphere of radius R . The normalization of the present power spectrum is often specified by the value of $\sigma_8 \equiv \sigma(R = 8h^{-1}\text{Mpc})$. For the top-hat, the smoothed perturbation field is denoted δ_R

or δ_M , where the mass M is related to the comoving radius R by $M = 4\pi\rho_m R^3/3$, in terms of the current mean density of matter ρ_m . The variance of $\langle\delta_m\rangle^2$ is

$$\sigma^2(M) = \sigma^2(R) = \int_0^\infty \frac{dk}{2\pi^2} k^2 P(k) \left[\frac{3j_1(kR)}{kR} \right]^2, \quad (1.14)$$

where $j_1(x) = (\sin(x) - x \cos(x))/x^2$. The function $\sigma(M)$ plays a crucial role in estimates of the abundance of collapsed objects. The current best-fit value of σ_8 is 0.812 ± 0.026 [Komatsu et al., 2011].

1.1.2 Formation and abundance of non-linear objects

The small density fluctuations seen in the CMB grow over time as described in the previous subsection, until the perturbation δ becomes of order unity, and the full non-linear gravitational problem must be considered. The dynamical collapse of a dark matter halo can be solved analytically only in cases of particular symmetry. If we consider a region which is much smaller than the horizon cH^{-1} , then the formation of a halo can be formulated as a problem in Newtonian gravity, in some cases with minor corrections coming from General Relativity. The simplest case is that of spherical symmetry, with an initial top-hat of uniform overdensity δ_i inside a sphere of radius R . Although this model is restricted in its direct applicability, the results of spherical collapse have turned out to be surprisingly useful in understanding the properties and distribution of haloes in models based on cold dark matter.

The collapse of a spherical top-hat is described by the Newtonian equation

$$\frac{d^2 r}{dt^2} = H_0^2 \Omega_\Lambda r - \frac{GM}{r^2}, \quad (1.15)$$

where r is a physical radius and M is the total mass enclosed within radius r . The enclosed δ grows initially as $\delta_L = \delta_i D(t)/D(t_i)$ in accordance with linear theory, but eventually δ grows faster than δ_L . If the mass shell at radius r is bound then it reaches a radius of maximum expansion and subsequently collapses. At the moment when the top-hat collapses to a point, the overdensity predicted by linear theory is $\delta_L = 1.686$ in the Einstein-de Sitter model, with only a weak dependence on Ω_m and Ω_Λ . Thus a top-hat collapses at redshift z if its linear overdensity extrapolated to the present day is

$$\delta_c = \frac{1.686}{D(z)}, \quad (1.16)$$

where we set $D(z=0) = 0$.

Even a slight violation of the exact symmetry of the initial perturbation can prevent the top-hat from collapsing to a point. Instead the halo reaches a state of virial equilibrium by violent relaxation. Using the virial theorem $U = -2K$ to relate the potential

energy U to kinetic energy K in the final state, the final overdensity relative to the average density of collapsed matter at the collapse redshift is $\Delta_c = 18\pi^2 \approx 178$ in the Einstein-de Sitter model. In the LCDM universe, we get

$$\Delta_c = 18\pi^2 + 82d - 39d^2, \quad (1.17)$$

where $d = \Omega_m^z - 1$ is evaluated at the collapse redshift, so that

$$\Omega_m^z = \frac{\Omega_m(1+z)^3}{\Omega_m(1+z)^3 + \Omega_\Lambda + \Omega_k(1+z)^2}. \quad (1.18)$$

Thus, a halo collapsing at redshift z has a physical virial radius [Barkana & Loeb, 2001]

$$r_{\text{vir}} = 0.784 \left(\frac{M}{10^8 h^{-1} M_\odot} \right)^{1/3} \left[\frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2} \right]^{-1/3} \left(\frac{1+z}{10} \right)^{-1} h^{-1} \text{kpc}, \quad (1.19)$$

and a corresponding circular velocity,

$$V_c = \left(\frac{GM}{r_{\text{vir}}} \right)^{1/2} = 23.4 \left(\frac{M}{10^8 h^{-1} M_\odot} \right)^{1/3} \left[\frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2} \right]^{1/6} \left(\frac{1+z}{10} \right)^{1/2} \text{km s}^{-1}. \quad (1.20)$$

We also define a virial temperature

$$T_{\text{vir}} = \frac{\mu m_p V_c^2}{2k_B} = 1.98 \times 10^4 \left(\frac{\mu}{0.6} \right) \left(\frac{M}{10^8 h^{-1} M_\odot} \right)^{2/3} \left[\frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2} \right]^{1/3} \left(\frac{1+z}{10} \right) \text{K}, \quad (1.21)$$

where μ is the mean molecular weight and m_p is the proton mass. Note that the value of μ depends on the ionization fraction of the gas; $\mu = 0.59$ for a fully ionized primordial gas, $\mu = 0.61$ for a gas with ionized hydrogen but only singly ionized Helium and $\mu = 1.22$ for neutral primordial gas. The binding energy of the halo is approximately

$$E_b = \frac{1}{2} \frac{GM^2}{r_{\text{vir}}} = 5.45 \times 10^{53} \left(\frac{M}{10^8 h^{-1} M_\odot} \right)^{5/3} \left[\frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2} \right]^{1/3} \left(\frac{1+z}{10} \right) h^{-1} \text{erg}. \quad (1.22)$$

The binding energy of baryons is smaller by a factor equal to the baryon fraction Ω_b/Ω_m .

Although spherical collapse captures some of the physics governing the formation of halos, structure formation in cold dark matter models proceeds hierarchically. At early times, haloes continuously accrete and merge to form high mass haloes. Numerical simulations of hierarchical halo formation indicate a roughly universal spherically-averaged density profile for the resulting haloes though with considerable scatter among different haloes [Navarro, Frenk & White, 1995; Bullock, Kravtsov & Weinberg, 2000]. The NFW profile has a form

$$\rho(r) = \frac{3H_0^2}{8\pi G} (1+z)^3 \frac{\Omega}{\Omega_m^z} \frac{\delta_c}{c_N x (1 + c_N x)^2}, \quad (1.23)$$

where $x = r/r_{\text{vir}}$, and the characteristic density δ_c is related to the concentration parameter c_N by

$$\delta_c = \frac{\Delta_c}{3} \frac{c_N^3}{\ln(1 + c_N) - c_N/(1 + c_N)}. \quad (1.24)$$

The concentration parameter itself depends on the halo mass M , at a given redshift z . Note that the dense, cuspy halo profile predicted by CDM models is not apparent in the mass distribution from measurements of rotation curves of dwarf galaxies [e.g. Salucci & Burkert, 2000]. We comment on this in detail later in this thesis.

In addition to characterising the properties of individual haloes, a critical prediction of any theory of structure formation is the abundance of haloes, i.e., the number density of haloes as a function of mass at any redshift. This prediction is an important step toward inferring the abundance of galaxies and galaxy clusters. While the number density of haloes can be measured for particular cosmologies in numerical simulations, an analytic model helps us gain physical understanding and can be used to explore the dependence of abundances on all the cosmological parameters.

A simple analytic model, which successfully matches most of the numerical simulations was developed by Press & Schechter [1974]. The model is based on the ideas of a Gaussian random field of density perturbations, linear gravitational growth, and spherical collapse. To determine the abundance of halos at a redshift z , we use δ_M , the density field smoothed on a mass scale M . Although the model is based on the initial conditions, it is usually expressed in terms of redshift-zero quantities. Thus, we use the linear density field at high redshift extrapolated to the present without including non-linear evolution. Since δ_M is distributed as a Gaussian variable with zero mean and standard deviation $\sigma(M)$, which depends only on the present power spectrum, the probability that δ_M is greater than some δ equals

$$\int_{\delta}^{\infty} d\delta_M \frac{1}{\sqrt{2\pi}\sigma(M)} \exp\left[-\frac{\delta_M^2}{2\sigma^2(M)}\right] = \frac{1}{2} \text{erfc}\left(\frac{\delta}{\sqrt{2}\sigma}\right). \quad (1.25)$$

The fundamental ansatz is to identify this probability with the fraction of dark matter particles which are part of collapsed haloes of mass greater than M at redshift z . There are two additional ingredients: first, the value used for δ is $\delta_c(z)$, which is the critical density of collapse found for a spherical top-hat; and second, the fraction of dark matter in haloes above M is multiplied by an additional factor of 2 in order to ensure that every particle ends up as part of some halo with $M > 0$. Thus, the final formula for the mass fraction in haloes above M at redshift z is

$$F(> M|z) = \text{erfc}\left(\frac{\delta_c(z)}{\sqrt{2}\sigma}\right). \quad (1.26)$$

This ad hoc factor of 2 is necessary, since otherwise only positive fluctuations of δ_M would be included. Bond et al. [1991] found an alternate derivation of this correction

factor, using a different ansatz. In their derivation, the factor of 2 has a more satisfactory origin, namely, the so-called cloud-in-cloud problem: for a given mass M , even if δ_M is smaller than $\delta_c(z)$, it is possible that the corresponding region lies inside a region of some larger mass $M_L > M$, with $\delta_{M_L} > \delta_c(z)$. In this case the original region should be counted as belonging to a halo of mass M_L . Thus, the fraction of particles which are part of collapsed haloes of mass greater than M is larger than the expression in Equation (1.25). Bond et al. [1991] showed that, under certain assumptions, the additional contribution results precisely in a factor of 2 correction.

Differentiating the fraction of dark matter in haloes above mass M yields the mass distribution. Letting dn be the comoving number density of haloes of mass between M and $M + dM$, we have

$$\frac{dn}{dM} = \sqrt{\frac{2}{\pi}} \frac{\rho_m}{M} \frac{-d(\ln \sigma)}{dM} \nu \exp(-\nu^2/2), \quad (1.27)$$

where $\nu = \delta_c(z)/\sigma(M)$ is the number of standard deviations which the critical collapse overdensity represents on mass scale M .

A better fit to the number density of haloes in simulations of gravitational clustering in the CDM models is given by Sheth & Tormen [1999].

$$N_{\text{ST}}(M, t) dM = A \sqrt{\frac{2}{\pi}} \frac{\rho_{\text{nr}}}{M} (a\nu)^{1/2} \left| \frac{d \ln \sigma}{dM} \right| [1 + (a\nu)^{-p}] \exp\left(-\frac{a\nu}{2}\right) dM, \quad (1.28)$$

where the parameters a , p , and A have best fit values of $a = 0.707$, $p = 0.3$ and $A = 0.322$ [Sheth & Tormen, 1999].

1.1.3 Key observations

Main observational evidence in support of the inflationary scenario comes from observations of the CMB, showing temperature anisotropies on large angular scales, which are interpreted as the results of the primordial density fluctuations at the decoupling epoch [e.g. Jarosik et al., 2011]. The same measurements confirm that the fluctuations are adiabatic and the value of the spectral index is the one predicted by the inflationary model, although there is an indication for a possible running spectral index, with $n > 1$ on large scales and $n < 1$ on small scales [Komatsu et al., 2011; Larson et al., 2011].

As we mentioned before, the amplitude of the power spectrum is not predicted by the model and has to be set observationally. Historically, this has been derived by comparing the observed CMB quadrupole anisotropies with theoretical ones or from abundance of galaxy clusters [Wright et al., 1992; Efstathiou, Bond & White, 1992]. This method has now been complemented by measuring the amplitude of the power spectrum from available large galaxy redshift surveys [Lahav et al., 2002]. The main problem related to the latter method is the attempt to estimate the distribution of matter using galaxies

as tracers. In fact, as the galaxies form preferentially in high density regions of the mass distribution, they are more strongly correlated than the underlying distribution and an additional factor, the bias, should be specified. It has been found from the 2dF Galaxy Redshift Survey that, at least on large scales ($5\text{--}30 h^{-1}$ Mpc), optically selected galaxies do indeed trace the underlying mass distribution [Verde et al., 2002].

An alternative method to measure $P(k)$ is based on observations of the high-redshift Ly α forest. The method is motivated by the physical picture that has emerged from hydrodynamical cosmological simulations and related semi-analytical models, in which typical Ly α forest lines arise in a diffuse IGM. The thermal state of this low-density gas is governed by simple physical processes, which lead to a tight correlation between the Ly α optical depth, τ and the underlying matter density [Bi & Davidsen, 1997; Hui, Gnedin & Zhang, 1997]. Thus, as the transmitted flux in a QSO spectrum is $F = e^{-\tau}$, one can extract information about the underlying mass density field from the observed flux distribution [Croft et al., 1998, 1999]. These works confirm a basic prediction of the inflationary CDM scenario: an approximately scale invariant spectrum of primeval fluctuations modulated by a transfer function that bends it toward k^{n-4} on small scales. A possible caveat on these applications is that radiative transfer effects are not included in current numerical simulations of the Ly α forest [Bolton, Meiksin & White, 2004]. Such effects might blur the polytropic temperature-density relation derived from the simulations and usually adopted, hence introducing a non-negligible error on the predicted cosmological spectrum.

A wealth of information about the initial density fluctuations is expected to come from the 21 cm line in absorption against the CMB, from $z = 200$ down to the epoch of the first structure formation. However, the main challenge for this kind of experiment is its observability. We discuss this method in more detail below. Also see the comprehensive review by Furlanetto, Oh & Briggs [2006].

To prove or disprove the existence of a universal density profile for dark matter haloes, accurate and extensive observations are needed. A method largely applied in the past to the study of the density profiles relies on H I rotation curves, although the spatial resolution of these observations is rarely good enough to set meaningful constraints. This is true both for the rotation curves of the low surface brightness galaxies and late-type dwarf galaxies, which are sometimes consistent with both a constant density core and a cusp, indicating that H I rotation curves do not have enough resolution to discriminate between the two models [van den Bosch et al., 2000; Borriello & Salucci, 2001; Kleya et al., 2003]. Only in a few cases it has been possible to derive a meaningful value of β (the inner logarithmic slope) giving $0.55 < \beta < 1.26$ at a 99.73% confidence level for a low surface brightness galaxy (LSBG), which should be consistent with the NFW slope, and a $\beta < 0.5$ for two dwarf galaxies at the same confidence level, more consistent with a flat core [van den Bosch et al., 2000]. H_α rotation curves have a higher spatial resolution and rise more steeply in the inner parts than the H I rotation

curves [Swaters, Madore & Trewhella, 2000]. Also, optical rotation curves have been used, indicating that the NFW profile provides a good fit to 66% of the 400 galaxies in the sample analyzed by Jimenez, Verde & Oh [2003], while 68% galaxies are well fitted by an isothermal profile with a core. Based on optical and radio rotation curves, it is confirmed that spiral galaxies have a universal rotation curve, characterised by one single free parameter, the I-band luminosity: low-luminosity spiral galaxies show rising rotation curves out to the optical radius, while high luminosity ones are flat or even decreasing [Persic, Salucci & Stel, 1996]. On the other hand, HST observations have revealed that elliptical galaxies have cusps which continue toward the center until the resolution limit. Bright elliptical galaxies have a shallow cuspy core with $0.5 < \beta < 1$, while faint ones have $\beta \approx 2$ [Merritt & Fridman, 1996].

Alternatively, the density profile of dark matter haloes can be investigated through gravitational lensing experiments. It has been argued that radially distorted, gravitationally lensed images of background sources in galaxy clusters, the so-called radial arcs, require a flat core in the cluster density profile. This would be consistent with the high resolution map of the $z = 0.39$ cluster 0024+1654 in which a very smooth, symmetric and non-singular core is observed [Tyson, Kochanski & dell'Antonio, 1998; Oguri, Taruya & Suto, 2001]. Nevertheless, it is also claimed that the NFW profile can produce radial arcs despite its central singularity, as indicated by some more recent observational data [Bartelmann, 1996]. But observations of radial arcs are so scarce that larger samples are needed.

Finally, as integral measures of weak gravitational lensing by dark matter haloes, like the aperture mass, are sensitive to the density profile, these can be used to discriminate between an isothermal and a NFW profile. In particular, as the halo mass range probed by the aperture mass is much wider for a NFW profile, counts of haloes with significant weak lensing signal are powerful discriminators [Bartelmann, King & Schneider, 2001].

1.2 Galaxy formation

Now that we have specified a cosmological model and can compute the abundance of dark matter halos, we are in a position to introduce models of galaxy formation. We can get a flavour of the subject by considering a simple galaxy formation model. Suppose we take each dark matter halo and assign to it a luminosity that scales linearly with the halo mass. Thus, each halo is given a fixed mass-to-light ratio. (We do not make any assumption about how this light is distributed among the galaxies within the halo.) We can then compare the prediction of this simple model with the luminosity function of galaxy groups—the observed abundance of galaxy groups as a function of their total luminosity, as measured in the 2dF galaxy redshift survey by Eke et al. [2004]. It is found that this simple prediction gives a poor match to the observed luminosity function

of groups [e.g. see Figure 6 of Baugh, 2006]. Thus if we are to retain the otherwise successful LCDM model, our prescription of a constant mass-to-light ratio is not valid.

We can reverse-engineer instead to find the mass-to-light ratio that will match the predictions with observations. It is seen that the ratio is lowest for haloes with mass $10^{12}h^{-1} M_{\odot}$ and rises by a factor of 6 for haloes with lower and higher mass [Yang, Mo & van den Bosch, 2003; Eke et al., 2004]. Thus, this simple consideration shows that galaxy formation is most efficient in haloes with a certain mass, namely $10^{12}h^{-1} M_{\odot}$. Furthermore, we learn two important properties of galaxy formation. Firstly, the efficiency of galaxy formation is low. Most baryons do not end up as stars. Observations suggest that galaxy formation is not particularly efficient at turning hot gas into cold gas and stars [Balogh et al., 2001]. It has been shown that only about 10% of baryons are in stars, depending on the stellar initial mass function (IMF) assumed [Cole et al., 2001]. An even smaller fraction is in the form of cold gas today [Zwaan et al., 2003]. Secondly, the efficiency of galaxy formation is not the same in haloes of different mass. Thus, the mass of the dark matter halo seems to play an important role in galaxy formation. Direct observational evidence for this is strong [Eke et al., 2006; Yang et al., 2005].

1.2.1 Gas infall and fragmentation

We begin by discussing the gravitational instability scenario, in which primordial density perturbations grow through gravitational Jeans instability to form the complex structures we observe today.

The Jeans length λ_J was originally defined in Newtonian gravity as the critical wavelength that separates oscillatory and exponentially growing density perturbations in an infinite, uniform and stationary distribution of gas. On scales l smaller than λ_J , the sound crossing time l/c_s is smaller than the gravitational free-fall time, $(G\rho)^{-1/2}$, allowing the build-up a pressure force that counteracts gravity. On larger scales, the pressure gradient force is too slow to react to a build-up of the attractive gravitational force. The Jeans mass is defined as the mass within a sphere of radius $\lambda_J/2$, $M_J = (4\pi/3)\rho(\lambda_J/2)^3$. In a perturbation with a mass greater than M_J , the self-gravity cannot be supported by the pressure gradient, and so the gas is unstable to gravitational collapse. This Newtonian derivation of the Jeans instability suffers from a conceptual inconsistency, however, as the unperturbed gravitational force of the uniform background must induce bulk motions [Binney & Tremaine, 2008]. This inconsistency has to be remedied when the analysis is done in an expanding universe.

The perturbative derivation of the Jeans instability criterion can be carried out in a cosmological setting by considering a sinusoidal perturbation superposed on a uniformly expanding background. Here, as in the Newtonian limit, there is a critical wavelength λ_J that separates oscillatory and growing modes. Although the expansion of the background slows down the exponential growth of the amplitude to a power-law growth, the fundamental concept of a minimum mass that can collapse at any given time remains

the same [Kolb & Turner, 1990; Peebles, 1993].

We can do this calculation by considering spherical fluctuations in the gas and dark matter densities in the form of a single spherical Fourier mode on a scale much smaller than the horizon:

$$\frac{\rho_{\text{dm}}(r, t) - \bar{\rho}_{\text{dm}}(t)}{\bar{\rho}_{\text{dm}}(t)} = \delta_{\text{dm}}(t) \frac{\sin(kr)}{kr}, \quad (1.29)$$

$$\frac{\rho_{\text{b}}(r, t) - \bar{\rho}_{\text{b}}(t)}{\bar{\rho}_{\text{b}}(t)} = \delta_{\text{b}}(t) \frac{\sin(kr)}{kr}, \quad (1.30)$$

where δ are the overdensity amplitudes and r is the comoving radial coordinate. We adopt an ideal gas equation of state with specific heat ratio $\gamma = 5/3$. Initially, the overdense amplitudes are small, gas temperature is uniform and we define the region inside the first zero of $\sin(kr)/kr$ as the collapsing object.

How does the gas temperature evolve? The temperature is determined by the coupling of its free electrons to the CMB through Compton scattering and by the adiabatic expansion. Thus T_b is generally between the CMB temperature $T_\gamma \propto (1+z)^{-1}$ and the adiabatically scaled temperature $T_{\text{ad}} \propto (1+z)^{-2}$. In the limit of tight coupling to the CMB, the gas temperature remains uniform, whereas in the adiabatic limit, it develops a gradient according to $T_b \propto \rho_b^{(\gamma-1)}$.

The evolution of the perturbation in dark matter is described in the linear regime by

$$\ddot{\delta}_{\text{dm}} + 2H\dot{\delta}_{\text{dm}} = \frac{3}{2}H^2(\Omega_b\delta_b + \Omega_{\text{dm}}\delta_{\text{dm}}), \quad (1.31)$$

whereas the evolution of the baryon overdensity evolves as [Kolb & Turner, 1990]

$$\ddot{\delta}_{\text{b}} + 2H\dot{\delta}_{\text{b}} = \frac{3}{2}H^2(\Omega_b\delta_b + \Omega_{\text{dm}}\delta_{\text{dm}}) - \frac{k_B T_i}{\mu m_p} \left(\frac{k}{a}\right)^2 \left(\frac{a_i}{a}\right)^{1+\beta} \left(\delta_b + \frac{2}{3}\beta[\delta_b - \delta_{b,i}]\right). \quad (1.32)$$

Here the parameter β distinguishes between the two limits of evolution of gas temperature: $\beta = 1$ in the adiabatic limit, and $\beta = 0$ in the strong coupling limit. The Jeans wavelength λ_J is obtained by setting the right hand side of Equation (1.32) to zero and solving for k_J . We see that this will be time dependent in general and also that perturbations with increasingly smaller initial wavelengths stop oscillating and start to grow.

Following recombination at $z \approx 10^3$, the residual ionization of the cosmic gas keeps its temperature locked to the CMB temperature via Compton scattering down to the redshift of [Peebles, 1993]

$$1 + z_t \approx 137(\Omega_b h^2 / 0.022)^{2/5}. \quad (1.33)$$

Thus, in the redshift range between recombination and z_t , $\beta = 0$, and we get

$$k_J = \left(\frac{2k_B T_\gamma(0)}{3\mu m_p}\right)^{-1/2} \sqrt{\Omega_m} H_0, \quad (1.34)$$

so that the Jeans mass obtains the value

$$M_J = \frac{4\pi}{3} \left(\frac{\lambda_J}{2}\right)^3 \bar{\rho}(0) = 1.35 \times 10^5 \left(\frac{\Omega_m h^2}{0.15}\right)^{-1/2} M_\odot. \quad (1.35)$$

On the other hand, for $z < z_t$, the gas temperature declines adiabatically and we have $\beta = 1$. The Jeans mass is then

$$M_J = 5.73 \times 10^3 \left(\frac{\Omega_m h^2}{.15}\right)^{-1/2} \left(\frac{\Omega_b h^2}{0.022}\right)^{-3/5} \left(\frac{1+z}{10}\right)^{3/2} M_\odot. \quad (1.36)$$

The Jeans scale is also referred to as the filtering scale. Note, however, that this expression for the Jeans mass is just a linear theory estimate and can only describe the initial phase of collapse. Indeed, it is not clear how the value of the Jeans mass is related to the mass of collapsed, bound objects. Thus the Jeans condition may only be a necessary but not sufficient condition. We will come to this point again below. The concept of Jeans filtering is explored in detail for galaxy formation in overdense regions in this thesis. We also note that alternate filtering models exist in the literature. For example, Gnedin & Hui [1998] showed that the Jeans mass is related to only the evolution of perturbations at a given time. When the Jeans mass itself varies with time, the overall suppression of growth depends on a time-averaged Jeans mass. See Choudhury & Ferrara [2005] for more discussion.

When perturbations grow to large amplitudes, we have to consider non-linear effects. The dark matter is cold and dominates gravity and so is unaffected by pressure effects. In order to estimate the minimum mass of baryonic objects, we have to consider the evolution of the accreted baryons. Assume that a dark matter halo with potential well $\phi(\mathbf{r})$ is formed at redshift z_{vir} . After gas settles down in this halo, it satisfies the hydrostatic equilibrium equation,

$$\nabla \rho_b = -\rho_b \nabla \phi. \quad (1.37)$$

At $z < 100$ gas temperature evolves adiabatically so we have $p_b \propto \rho_b^{5/3}$, which gives us

$$\frac{\rho_b}{\bar{\rho}_b} = \left(1 - \frac{2}{5} \frac{\mu m_p \phi}{k_B \bar{T}}\right)^{3/2}, \quad (1.38)$$

where \bar{T} is the background gas temperature. This mass estimate provides a better estimate than the Jeans mass since it incorporates the non-linear potential of the dark matter halo. If we define the virial temperature corresponding to the potential well as $T_{\text{vir}} = -(1/3)\mu m_p \phi / k_B$ then the overdensity of baryons can be written as

$$\delta_b = \left(1 + \frac{6}{5} \frac{T_{\text{vir}}}{\bar{T}}\right)^{3/2} - 1. \quad (1.39)$$

We may now say that a collapsed baryonic object—a protogalaxy—is formed when the value of δ_b exceeds 100. This gives a minimum mass

$$M_{\min} = 5.0 \times 10^3 \left(\frac{\Omega_m h^2}{0.15} \right)^{-1/2} \left(\frac{\Omega_b h^2}{0.022} \right)^{-3/5} \left(\frac{1+z}{10} \right)^{3/2} M_{\odot}. \quad (1.40)$$

This happens to be close to our earlier linear estimate, however it takes non-linear effects into account. Of course, when the first stars and galaxies form, this value of M_{\min} changes due to various feedback effects. We will consider these effects shortly.

Apart from gravitational instability, another process that greatly affects the condition of the gas is cooling. As we saw above, in objects with baryonic masses $3 \times 10^4 M_{\odot}$, gravity dominates and results in the bottom-up hierarchy of structure formation; at lower masses pressure delays the collapse. The first objects to collapse are those at the mass scale that separates the two regimes. Such objects can fragment only through cooling. Thus, there are two independent minimum mass thresholds for star formation: the Jeans mass and the cooling mass. The higher of these two decides the actual threshold. We elaborate more on this issue in Chapter 3. The primary molecule that acquires sufficient abundance to affect the thermal state of the gas is molecular Hydrogen H_2 . These molecules are produced through various processes in the early universe, but the dominant of them is the one where free electrons act as a catalyst. Cooling via H_2 forms objects with mass $\sim 10^4 M_{\odot}$, which are usually called minihaloes. Primordial H_2 abundance is small, about 10^{-7} relative to hydrogen by number. However, at redshift $z < 110$, CMB intensity becomes weak enough to allow H_2 formation.

Nonetheless, H_2 is fragile and can easily be dissociated by photons with energies of 11.26–13.6 eV, to which the IGM is transparent even before reionization. Haiman & Loeb [1997] showed that the UV flux necessary for dissociating H_2 throughout the collapsed environments is two orders of magnitudes lower than the amount required for IGM reionization. Thus H_2 is completely destroyed by trace amount of first generation stars. Further star formation is only possible via atomic cooling, which happens at $T_{\text{vir}} \geq 10^4$ K. Such objects correspond to mass of $\sim 10^8 M_{\odot}$. Note, however, that the radiative feedback on H_2 need not be negative alone. Indeed, production of free electrons can enhance the molecular Hydrogen production in dense environments. Haiman [2000] shows that this will have an important effect on the topology of the reionization.

Apart from star formation, black hole formation is expected to happen in the early stages of galaxy formation. This is important, because quasars are more effective than stars in ionizing the intergalactic hydrogen. Thus history of reionization may be greatly altered if quasars form early. The problem of formation of BHs is not a priori more complicated than that of formation of stars. However, little work exists on the formation of BHs and there are many open questions. We present a review of the literature on this topic in Chapter 5.

1.2.2 Galactic structure and evolution

After formation, several factors govern the evolution of galaxies. First models of how galaxy formation occurs with these factors were studied in the 1970s [White & Rees, 1978; Binney, 1977]. The first of these is feedback, which is also relevant to the formation of galaxies. Phenomenologically, the relevance for feedback processes was understood early on when it was found that the observed faint-end luminosity function was less steep than expectation [White & Rees, 1978]. Although feedback effects are thus quite important, they are also difficult to model. They can be broadly classified into three categories.

Mechanical feedback is the ejection of cold gas from a galaxy by a supernova driven wind [Dekel & Silk, 1986]. This reheated cold gas could be blown out to the hot gas halo, from which it may subsequently re-cools ('retention feedback'), or may be ejected from the halo altogether ('ejection feedback'). The distinction between these two modes of feedback can have a significant impact on the galaxy luminosity function [Kauffmann et al., 1999]. Radiative feedback on the other hand, modifies the rate at which gas cools, either by altering the density profile or entropy of the hot gas halo (following injection of energy into the halo due to mechanical feedback) or by reducing the fraction of baryons that fall into dark matter haloes and changing the cooling rate (photo-ionization suppression of star formation in low mass haloes).

Finally, chemical feedback changes the chemical composition of the protogalaxy. Formation of stars changes the metal content of the ISM. Firstly, the act of forming stars removes cold gas and associated metals from the ISM. Also, as stars evolve, they return gas to the ISM with an enhanced metallicity. The return mechanism is usually stellar winds and supernovae. This affects further star formation because: (1) the rate at which gas cools is a function of metallicity, higher metallicity results in a shorter cooling time; (2) metallicity of the stars has an effect on the luminosity and colours of the galaxy; and (3) the optical depth of a galaxy scales linearly with the metallicity of its cold gas. Type Ia supernovae dominate the production of iron whereas type II supernovae primarily produce the α -elements and also nitrogen and sodium. Thus, abundance ratios of metals in the ISM can contain clues to galaxy formation [Baugh, 2006].

In the standard model that we are discussing, haloes are assumed to grow through mergers and accretion. The halo resulting from a merger event is smooth and devoid of any substructure. The galaxies in progenitor haloes survive the merger of their parent haloes as a result of them being more concentrated than the dark matter, due to dissipative cooling of gas. This picture leads naturally to a scenario in which a dark halo contains a massive central galaxy surrounded by smaller satellite galaxies. Satellite galaxies lose their angular momentum due to dynamical friction and then merge into the central galaxy. In addition to providing an alternative mechanism to gas cooling for increasing the mass and luminosity of the central galaxy, the accretion of a satellite galaxy can have more dramatic consequences. The impact of a galaxy merger is usually

quantified by the ratio of mass of the accreted satellite galaxy to the mass of the central galaxy [Kauffmann et al., 1999; Li et al., 2007b].

All of above elements can be combined with our knowledge of stellar evolution to create predictions of the spectral energy distributions (SEDs) of galaxies, that is the amount of energy emitted by a galaxy as a function of wavelength or energy. This is usually achieved by using stellar population synthesis models, which provide a look-up table of the SED of a single-age population of stars as a function of the time elapsed since the stars were born. The stars are born with a distribution of masses set by an assumed initial mass function (IMF) and have a given metallicity. As a simple stellar population ages, hot massive stars evolve out of the main sequence most rapidly, with the result that the flux of UV photons declines with age. Stellar population synthesis models that encode this information are traditionally treated as trusted black boxes [Leitherer et al., 1999; Bruzual & Charlot, 2003; Conroy, Gunn & White, 2009].

1.2.3 Modelling and key observations

Semi-analytic models and direct simulations of gas and dark matter have been used to study galaxy formation.

There are two principle algorithms in common use to follow the hydrodynamics of gas in an expanding universe: particle-based Lagrangian schemes, which employ a technique called smoothed particle hydrodynamics (SPH) [Monaghan, 1992; Springel & Hernquist, 2003a; Springel, 2010], and grid-based Eulerian schemes [Cen & Ostriker, 1999; Ryu et al., 1993]. The SPH technique has traditionally achieved greater resolution compared to Eulerian schemes, due to its comoving nature. Instead, grid based codes deal much better with shocks and discontinuities. To improve the spatial resolution of these codes, much work has been done on adaptive mesh refinement (AMR) [see e.g. Nagai & Kravtsov, 2005]. While studying galaxy formation with such simulations, first a large representative volume of the universe is simulated. Next, a single halo is extracted from this large volume and re-simulated at a much higher resolution with gas. In the re-simulation, the region surrounding the high resolution volume is represented using higher mass particles, so that the tidal forces exerted on the high resolution structure is properly represented.

Various tests have been conducted to compare the performance of various gas dynamical codes. The best agreement between these codes was found for the dark matter distributions and the worst agreement for its X-ray luminosity [Baugh, 2006]. Quantities such as the gas temperature and the mass fraction of gas within the virial radius are found to agree within 10%.

Galaxy formation is also studied, including in this thesis, using semi-analytic prescriptions applied to dark matter halo merger trees drawn from N-body simulations. Virialised dark matter haloes are identified in N-body simulations using a friend-of-friend (FOF) algorithm. Merger trees describing the assembly of haloes can then be

extracted from these simulations by relating haloes across time as progenitors and descendants. (This construction is not completely straightforward, as spurious halo fragmentation can sometimes occur. We discuss this further in Chapter 4.) Adding these merger trees to semi-analytic models provides information about the spatial distribution of galaxies [e.g. Kauffmann et al., 1999].²

Semi-analytic models then populate high redshift haloes with hot gas at halo virial temperature and then calculate processes like gas cooling, star formation and feedback halo-by-halo. A galaxy is assigned to the most bound particle in the halo in which it first forms. When two haloes merge, the evolution of their constituent galaxies is computed using dynamical friction arguments. The calculation of cooling rate of gas is carried out under specialized conditions and depends upon a number of assumptions. The star formation rate of a halo is proportional to the cold gas available in that halo. Kereš et al. [2005] addressed the question of how galaxies acquire their gas in detail using SPH simulations. They characterise their results in terms of two cooling regimes: a cold mode in which gas is funneled down the filaments of large scale structure onto galaxies, and a hot mode in which gas cools from a quasi-static halo. The cold flow is found to dominate in low mass haloes ($M < 3 \times 10^{11} M_{\odot}$) and at high redshift ($z > 3$). Note that the metallicity of gas can have significant impact on gas cooling. Due to numerical limitations, early semi-analytic models used a fixed global metallicity with some ad hoc time evolution. However, as we do in this thesis, a self-consistent metallicity evolution can now be incorporated using population synthesis models.

Such models have been fairly successful in explaining observations of galaxy luminosity functions, various scaling relations like the Tully-Fisher relation and the fundamental plane, the abundance ratios of elliptical galaxies and galaxy downsizing. Continued evolution of the models is driven by datasets where model predictions are at odds with observations [Baugh, 2006].

As gas cools and condenses within dark matter potential wells, it emits copious radiation. The energy radiated is comparable to the gravitational binding energy of the baryons. Most of this cooling radiation is emitted by gas with $T < 20000$ K. As a result, roughly 50% of it emerges in the Ly- α line. Moreover, this radiation is less likely to be attenuated by dust since it will be emitted from the outer regions of the halo. This picture has been put forward as an explanation to explain observations of large (≈ 100 kpc) luminous ($L \approx 10^{44}$ erg s⁻¹) “blobs” of Ly- α emission found in narrow-band surveys of $z = 3$ protoclusters [Haiman & Loeb, 1999a; Steidel et al., 2000; Fardal et al., 2001].

In addition to the Ly- α radiation, observational probes of galaxy formation include

²Advent of large volume, high resolution simulations such as the Millennium [Springel et al., 2005] have allowed exceptionally detailed study of galaxy formation using semi-analytic models [Bower et al., 2006; De Lucia & Blaizot, 2007]. The Millennium Simulation is a pure dark matter simulation with a Λ CDM model with 2160^3 particles in a periodic cube $500 h^{-1}$ Mpc on a side. This corresponds to a particle mass of $8.6 \times 10^8 h^{-1} M_{\odot}$. This simulation has about twenty million dark matter haloes at $z = 0$.

cooling emission from H_2 molecules, metal poor stars in the Milky Way halo and gravitational waves from BH remnants of the first stars.

1.3 Inter-galactic medium

In the previous two sections, we saw how the hot ionized gas that formed as a result of Big Bang nucleosynthesis recombined, and then cooled and fragmented in dark matter haloes to form protogalaxies with stellar populations, an inter-stellar medium, and likely, a central massive black hole. We now turn to the baryonic component of the universe that is outside these galaxies, namely the inter-galactic medium.

Field [1959a] made the first attempt to detect the IGM by observing the hyperfine 21 cm absorption feature from hydrogen. This attempt did not succeed; it merely gave an upper bound on the baryon density parameter Ω_b . However, a much improved measurement on the IGM baryon density was available in the 1960s when Gunn & Peterson [1965] observed slight Ly- α absorption in the spectra of recently discovered quasars. They concluded that the neutral hydrogen density in the IGM was very small compared to the hydrogen density in stars. This meant that either galaxy formation was very efficient and that the IGM was largely empty, or that most hydrogen in the IGM was ionized. The latter hypothesis was strongly corroborated by evidence of a rise in Ly- α flux decrement in the spectra of $z \gtrsim 6$ quasars, and by other observations [Fan, Carilli & Keating, 2006].

In the 1960s, individual Ly- α absorption features were also identified in high resolution QSO spectra [Bahcall, Greenstein & Sargent, 1968]. These features are now collectively known as the Ly- α forest. From a systematic survey of such features, Sargent et al. [1980] argued that the Ly- α forest is of extra-galactic origin and showed that the measured widths of these features correspond to a temperature of 10^4 K. The neutral column densities of these absorption systems range from 10^{12} to 10^{22} cm $^{-2}$. The highest column-density systems are called damped Ly- α absorbers (DLAs). Many of these hydrogen absorption systems also show absorption lines from metals like carbon, silicon, nitrogen, oxygen, magnesium, iron, and others [e.g. Songaila, 2001]. The origin of these metals in absorption systems is an open problem.

1.3.1 Absorption systems

The IGM is detected through the absorption features it produces in the spectrum of a bright background source of light, like a quasar. Physics of these absorption features can be described using the equation of radiative transfer.

We define the specific intensity $I_\nu(\mathbf{r}, t, \hat{\mathbf{n}})$ as the rate at which energy crosses a unit area per unit solid angle per unit time as carried by photons of energy $h_P\nu$ traveling in the direction $\hat{\mathbf{n}}$. The equation of radiative transfer in an expanding universe can then be

written as [Abel, Norman & Madau, 1999; Choudhury, 2009]

$$\frac{\partial I_\nu}{\partial t} + c\hat{\mathbf{n}} \cdot \nabla I_\nu - H(t)\nu \frac{\partial I_\nu}{\partial \nu} + 3H(t)I_\nu = -c\alpha_\nu I_\nu + cj_\nu. \quad (1.41)$$

Here, $\alpha_\nu(\mathbf{r}, t, \hat{\mathbf{n}})$ is the attenuation coefficient of the medium and $j_\nu(\mathbf{r}, t, \hat{\mathbf{n}})$ is the emission coefficient, which describes the local specific luminosity. As before, $H(t)$ is the Hubble parameter. In general, attenuation is due to absorption of photons and their scattering out of the beam. We can then write the attenuation coefficient as

$$\alpha_\nu(\mathbf{r}, t, \hat{\mathbf{n}}) = \rho(\mathbf{r}, t)\kappa_\nu(\mathbf{r}, t, \hat{\mathbf{n}}) + n(\mathbf{r}, t)\sigma_\nu(\mathbf{r}, t, \hat{\mathbf{n}}), \quad (1.42)$$

where $\kappa_\nu(\mathbf{r}, t, \hat{\mathbf{n}})$ is the medium's opacity and $\sigma_\nu(\mathbf{r}, t, \hat{\mathbf{n}})$ is its scattering cross-section. As usual, $\rho(\mathbf{r}, t)$ is the mass density and $n(\mathbf{r}, t) = \rho(\mathbf{r}, t)/m$ is the number density of scattering particles of mass m .

Equation (1.41) relates the observed specific intensity to the ‘‘incoming’’ specific intensity and properties of the medium. Its formal solution is given by

$$I_\nu = I_{\nu_0}^{\text{in}}|_{s_0, t_0} \left[\frac{a(t_0)}{a(t)} \right]^3 \exp \left(- \int_{s_0}^s ds' \alpha_{\nu'}|_{s', t'} \right) + \int_{s_0}^s ds'' \left\{ j_{\nu''}|_{s'', t''} \left[\frac{a(t'')}{a(t)} \right]^3 \exp \left(- \int_{s''}^s ds' \alpha_{\nu'}|_{s', t'} \right) \right\}, \quad (1.43)$$

where $\nu_0 = \nu a(t)/a(t_0)$, $\nu' = \nu a(t)/a(t')$, and $\nu'' = \nu a(t)/a(t'')$.

Absorption features in QSO spectra are a result of resonant scattering of photons received from the QSO by atoms and molecules in the medium. Usually, this configuration does not have intermediate sources so we can set $j_\nu = 0$. For static atoms, the cross section of resonant scattering is given by the Lorentz profile. However, in general, atoms are not static; they may undergo both a random thermal motion and some secular motion. The secular motion does not modify the scattering cross-section; it may simply doppler shift the line center. On the other hand, random motion does modify the cross section by convolving the Lorentz profile with a Gaussian distribution of frequencies about the line center. The result is called the Voigt profile. The cross section can now be written as

$$\sigma_\nu = \left(\frac{\pi e^2}{m_e c} \right) \left[\frac{1}{4\pi\epsilon_0} \right] f_{\text{lu}} \phi_V(a, \nu), \quad (1.44)$$

where

$$\phi_V(a, \nu) = \frac{1}{\pi^{1/2} \Delta\nu_D} H(a, x) \quad (1.45)$$

is the normalised Voigt profile, that is

$$H(a, x) = \frac{a}{\pi} \int_{-\infty}^{\infty} dy \frac{e^{-y^2}}{(x-y)^2 + a^2}. \quad (1.46)$$

Here, $a = \Gamma_{ul}/4\pi\Delta\nu_D$ is the ratio of the Lorentz profile damping width to the thermal doppler width $\Delta\nu_D = \nu_{lu}(2k_B T/m_a)^{1/2}/c$. Also, f_{lu} is the oscillator strength of the resonant transition. For the Ly- α transition, $f_{lu} \approx 0.4162$.

We can now quantify the attenuation in the intensity of a background QSO using Equation (1.43). The intensity will be attenuated by a factor $e^{-\tau_\nu}$, where, from Equation (1.43),

$$\tau_\nu = \int_{s_0}^s ds' n(s', t') \sigma_{\nu'}. \quad (1.47)$$

This quantity is called the optical depth. If ν_{lu} is the resonance line frequency, then radiation emitted by the source at time t_0 and rest frame frequency $\nu_0 > \nu_{lu}$, will be scattered by the medium at time t' given by $\nu' = \nu_0 a(t_0)/a(t') = \nu_{lu}$. Thus the received spectrum will be attenuated at all observed frequencies ν in the range $\nu_{lu} > \nu > \nu_{lu} a(t_0)/a(t)$, or wavelengths $\lambda_{lu} < \lambda < \lambda_{lu} a(t)/a(t_0)$, where λ_{lu} is the wavelength corresponding to ν_{lu} . Further, we define the equivalent width of the absorption feature as

$$W = \int (1 - e^{-\tau_\nu}) d\nu \approx \int \tau_\nu d\nu, \quad (1.48)$$

where the latter inequality holds for small optical depth (“optically thin medium”). For optically thin systems in terms of the column density N of absorbers (total number of absorber per unit area column), the equivalent width can also be written as

$$W = N \int \sigma_\nu d\nu = N \sigma_0 \Delta\nu, \quad (1.49)$$

where σ_0 is the cross-section averaged over a bandwidth $\Delta\nu$.

Observations show that the IGM is in fact quite clumpy. These inhomogeneities result in discrete absorption lines in QSO spectra originating in distinct localised regions. The resulting collection of absorption features is known as the Ly- α forest. In this case the optical depth becomes $\tau_\nu = \sum_i \tau_\nu(i)$ where $\tau_\nu(i)$ is the optical depth corresponding to each region.

Absorption features comprising the Ly- α forest can be classified into three types based on the physical origin on features: (1) Ly- α forest systems, (2) DLAs, and (3) Lyman limit systems (LLSs). This classification is not strictly exclusive but Ly- α forest systems typically have column densities of $N \leq 10^{17} \text{ cm}^{-2}$, while those for LLSs are $10^{17} - 10^{19} \text{ cm}^{-2}$. DLAs are high column density systems with $N > 10^{19}$.

Furthermore, the number of absorption systems per unit redshift increases with increasing redshift. Part of this increase is expected simply because of the expanding universe. Thus if $n(z)$ is the proper number density of absorbers, and $\sigma(z)$ is their proper absorption cross-section, then the number density per unit proper length is $dN/dl = n(z)\sigma(z)$. Since the line element dl is given by $dl/dz = c/H(z)(1+z)$, it

can be shown that

$$\frac{dN}{dz} = (2100\text{Mpc})n_c(z)\sigma(z)(1+z)^{1/2} \left(1 + \frac{2.3}{(1+z)^3}\right)^{-1/2}, \quad (1.50)$$

where $n_c = n(1+z)^{-3}$ is the comoving number density of absorbers. Thus, if the comoving number density and absorption cross-section are constant, we expect an increase proportional to $dN/dz \propto (1+z)^{1/2}$. For $z < 1.5$, it is seen that this correctly describes the observed evolution of the number of absorption systems [Weymann et al., 1998; Penton, Shull & Stocke, 2000; Penton, Stocke & Shull, 2004; Danforth & Shull, 2008].

However, for redshifts $z > 1.5$, significant evolution in dN/dz is seen. For example, Kim et al. [2002] found $dN/dz = 6.1(1+z)^{2.47 \pm 0.18}$ for Ly- α forest systems at these redshifts. This implies a significant evolution in the number density of absorbers and/or the absorption cross-section. There is also a dependence on the column density: the number LLSs and DLAs grow more slowly than the Ly- α forest systems [Prochaska, Herbert-Fort & Wolfe, 2005; Janknecht et al., 2006]. These observations, combined with numerical simulations, suggest that this evolution can be attributed to (1) structural evolution of the IGM, (2) evolution in the ionizing UV background, and (3) decreasing gas density.³

Apart from hydrogen absorption systems mentioned above, deuterium and helium absorption systems are also seen. High density Ly- α absorption systems can be used to calculate the primordial deuterium abundance D/H, which can constrain cosmic baryon density parameter Ω_b . This observation is difficult, but several measurements have been made. The current best estimate for D/H is $\log_{10} \text{D/H} = -4.55 \pm 0.04$ [O'Meara et al., 2006]. This corresponds to a baryon density parameter $\Omega_b h^2 = 0.0213 \pm 0.0013$ and a nucleon-to-photon ratio of $\eta = 5.8 \pm 0.3 \pm 10^{-10}$ [Meiksin, 2009]. This is in agreement with estimates based on the WMAP data combined with other datasets [Komatsu et al., 2011]. The Helium Ly- α absorption has also been detected. At redshifts $z \lesssim 5$ almost all Helium in the IGM is expected to be in the form of He II or He III. High column densities of He II have been reported at $z \gtrsim 3$. For example, Heap et al. [2000] report $\tau > 4.8$ in the spectrum of a $z = 3.3$ quasar. On the other hand, for $z \lesssim 3$, very low optical depths have been reported. As an example, Davidsen, Kriss & Zheng [1996] observed $\tau \sim 1$ in the spectrum of a $z = 2.7$ quasar. This suggests that He II was ionized at $z \sim 3$. It is now known that the He II Ly- α optical depth is quite patchy, consistent with ongoing Helium reionization at $z \sim 3$.

Finally, absorption systems for elements heavier than Helium have also been detected in the IGM. Most common metals are carbon, nitrogen, silicon and iron, although other metals like oxygen, magnesium, neon, and sulfur are also detected. These metal

³All observations mentioned here obtain dN/dz by fitting to systems with constant H I column density. At low redshifts, gas density can reduce so that fewer systems are found in this range.

absorption systems are useful probes of the temperature evolution of the IGM. Also, since they were likely produced in stars and transported to the IGM by galactic winds, metal systems also act as a probe of the star formation history. The IGM metallicity appears to be roughly constant till $z \sim 5$ with the C IV density parameter of a few times 10^{-8} [Songaila, 2001]. There is tentative evidence that the metallicity goes down at higher redshifts [Simcoe et al., 2011]. Metal absorption systems have also been used to rule out the variability of the constants of nature. A variation of less than 10^{-6} in the fine structure constant has been inferred [Chand et al., 2004; Srianand et al., 2004].

1.3.2 Epoch of Reionization

Reionization of the IGM is an important stage in cosmological structure formation. In later chapters of this thesis, we will be concerned with the epoch of reionization of hydrogen, although two more epochs of reionization can be considered, for helium.⁴ The epoch of hydrogen reionization has attracted much attention in the last decade following measurement of the IGM Gunn-Peterson optical depth and that of Thomson scattering of CMB radiation in the IGM.

The Ly- α optical depth of the IGM measured in the spectra of several $z > 5.8$ QSOs from the Sloan Digital Sky Survey (SDSS) show that the IGM was reionized at $z \sim 5.7$ [Fan et al., 2001]. The Ly- α optical depth rises rapidly at higher redshifts. By $z \sim 6$, optical depth measurement becomes difficult due to high absorption and only lower limits can be obtained. Although, this rapid rise indicates the epoch of hydrogen reionization, this interpretation is not straightforward. The reason behind this is that for flat universe the Ly- α optical depth is given by

$$\tau(z) \sim 4.6 \times 10^5 \frac{\Omega_{\text{H I}}(z)h}{\Omega_m^{1/2}(1+z)^{3/2}}, \quad (1.51)$$

where $\Omega_{\text{H I}}$ is the neutral hydrogen density parameter. As a result, only a small neutral hydrogen fraction (order of 10^{-4}) is required to obtain a large Ly- α optical depth. Another reason to doubt the above interpretation is because some $z > 6$ QSOs show transmitted flux at $z \sim 6$ [White et al., 2003, 2005].

Another indication of an epoch of hydrogen reionization at redshifts $z > 6$ is provided by the WMAP measurements of the IGM Thomson scattering optical depth

$$\tau_e = \int_{z_l}^{z_u} dz n_e c \sigma_T \left| \frac{dt}{dz} \right|, \quad (1.52)$$

where σ_T is the Thomson cross section and n_e is the comoving electron density. The five-year WMAP polarization data yield an optical depth of $\tau_e = 0.087 \pm 0.017$ up to

⁴The epoch of ionization of He I to He II is expected to coincide with that of hydrogen reionization since both ionization processes require similar energies. As we saw before, He II to He III reionization epoch is expected to be distinct, at around $z \sim 3$.

the epoch of reionization, which yields $z \sim 11.0 \pm 1.4$ for the redshift of reionization assuming that it was an instantaneous event throughout the universe [Dunkley et al., 2009]. The 2σ and 3σ lower limits are $z > 8.2$ and $z > 6.7$ respectively.

A third probe of the epoch of reionization comes from surveys of galaxies with strong Ly- α emission lines. These photometrically selected galaxies are also known as Ly- α emitters (LAEs). More than 100 LAEs have been observed at $z \sim 6.5$ and many have now been spectroscopically confirmed. These galaxies are expected to represent a significant fraction of star forming galaxies at high redshift. Properties of LAEs directly probe the IGM neutral fraction: in a largely neutral IGM the Ly- α emission is considerably attenuated. Thus, we expect that the Ly- α galaxy luminosity function will decrease sharply in an increasingly neutral IGM.

Malhotra & Rhoads [2004] and Stern et al. [2005] determined the luminosity function of these galaxies at $z = 6.5$ and $z = 5.7$. They found no evolution between these two redshifts, consistent with the interpretation that the IGM was largely reionized by $z \sim 6.5$. Interpreting high redshift luminosity functions is not straightforward and detailed modelling is required. For instance, local H II regions around these galaxies can affect luminosity function evolution [Cen, Haiman & Mesinger, 2005] and clustering of galaxies can enhance this effect [Cen, 2005].

An alternative method of observing the epoch of reionization is through the detection of the 21 cm signature from the neutral hydrogen in the IGM before and during the reionization process. We will consider the use of this method for studying the post-reionization IGM later in this thesis. We are interested in the two hyperfine levels of the ground energy state of H I. We denote the lower hyperfine level by 0 and higher hyperfine level by 1. Level 1 is a triplet and level 0 is a singlet. Energy difference between two levels is $E_{10} = 5.9 \times 10^{-6}$ eV, which corresponds to temperature $T_* = E_{10}/k_B = 0.068$ K. A transition between these states corresponds to rest frame frequency of $\nu_{10} = 1420.4057$ MHz (or rest wavelength of $\lambda_{10} = 21.10611$ cm) and is potentially observable in emission or absorption. The emission and absorption of 21 cm radiation from a neutral IGM is governed by the spin temperature T_S of hydrogen, defined by

$$\frac{n_1}{n_0} = 3 \exp\left(-\frac{T_*}{T_S}\right), \quad (1.53)$$

where n_0 and n_1 are the singlet and triplet $n = 1$ hyperfine levels. In the presence of the CMB alone, the spin temperature quickly comes into equilibrium with the CMB temperature, in about 10^5 yr. Thus there will be no detectable absorption or emission relative to the CMB. This is precisely what is expected to happen at redshifts $40 < z < 1100$. There are however, two mechanisms that can break the coupling with the CMB, namely (1) collisions between hydrogen atoms, and (2) the Wouthuysen-Field effect. Collisional coupling between the spin and kinetic temperatures is dominated by spin exchange between colliding hydrogen atoms. Electron-hydrogen and proton-hydrogen

collisions may also contribute. The Wouthuysen-Field effect mixes the hyperfine levels of neutral hydrogen in its ground state via an intermediate transition to the $2p$ state due to scattering by Ly- α photons. An atom initially in the $n = 1$ singlet state may absorb a Ly- α photon that puts it in an $n = 2$ state, allowing it to return to the triplet $n = 1$ state by spontaneous decay. Due to these two effects, a patchwork of 21 cm absorption or emission against the CMB will result. While the collisional coupling is significant only at high densities and temperatures, the Wouthuysen-Field coupling can be a dominant mechanism at detectable redshifts. Several large radio telescopes are trying to measure this signature, thereby revealing the transition from a neutral IGM to an ionized one.

Equilibrium implies that the spin temperature of neutral hydrogen is

$$T_S = \frac{T_{\text{CMB}} + y_\alpha T_\alpha + y_c T_K}{1 + y_\alpha + y_c}, \quad (1.54)$$

where

$$y_\alpha \equiv \frac{P_{10}}{A_{10}} \frac{T_*}{T_\alpha} \quad \text{and} \quad y_c \equiv \frac{C_{10}}{A_{10}} \frac{T_*}{T_K}, \quad (1.55)$$

are called the Ly- α and collisional pumping efficiencies respectively. The quantity T_K is the kinetic temperature of the hydrogen, while T_α is the colour temperature. Also, C_{10} is the collisional de-excitation rate, P_{10} is the rate of indirect de-excitation via the Wouthuysen-Field effect and A_{10} is the Einstein A coefficient.

The basic idea behind 21 cm observations can be understood by considering a patch of neutral hydrogen with spin temperature $T_S \neq T_{\text{CMB}}$, having angular size in the sky that is large compared with the beam-width, and radial velocity extent that is larger than the bandwidth due to Hubble expansion. Then this cloud will appear in emission or absorption against the CMB. We can calculate the inter-galactic optical depth as follows. For a two-level system, it is possible to define Einstein coefficients of stimulated and spontaneous emission (B_{10} and A_{10}) and stimulated absorption (B_{01}). When the system is in thermodynamic equilibrium, there is a definite relationship between these coefficients, known as Einstein relations. Once the Einstein coefficients for a system are known we can write down the emission and absorption coefficients that appear in the radiative transfer equation. In our case when the H I cloud is in thermal equilibrium, its absorption coefficient is given by

$$\alpha_\nu = \frac{h\nu}{4\pi} n_0 B_{01} \left[1 - \exp\left(-\frac{E_{10}}{k_B T}\right) \right] \phi(\nu), \quad (1.56)$$

where $\phi(\nu)$ is the line profile and n_0 is the number density of atom in the 0 level. The optical depth is then given by

$$\tau_\nu = \int ds \frac{h\nu}{4\pi} n_0 B_{01} \left[1 - \exp\left(-\frac{h\nu}{k_B T_s}\right) \right] \phi(\nu). \quad (1.57)$$

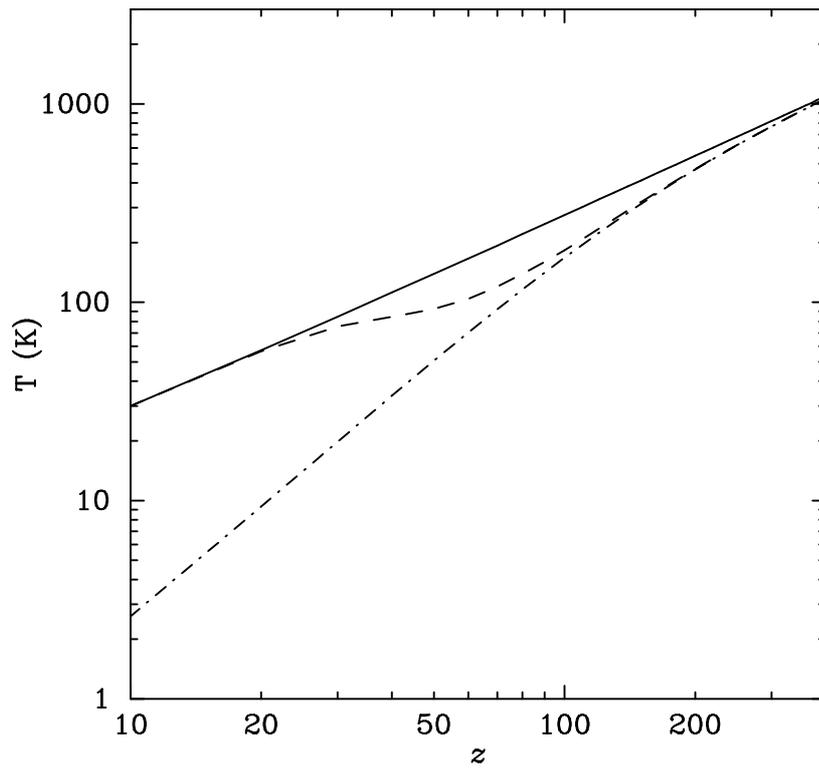


Figure 1.1: Evolution of the 21 cm spin temperature. Solid line is CMB temperature, dot-dashed line is matter temperature, and dashed line the 21 cm spin temperature. See text for details.

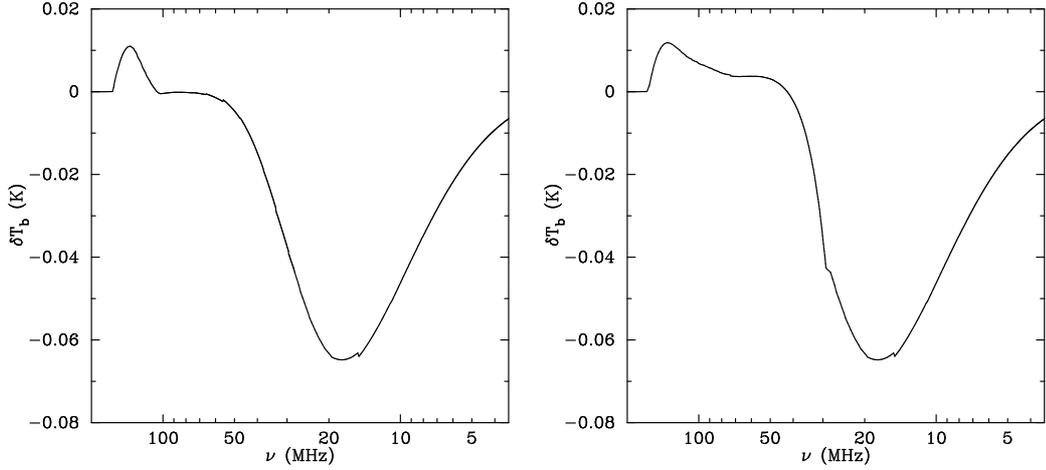


Figure 1.2: Spectrum of the 21 cm brightness temperature contrast in a observationally constrained reionization model. Effect of Ly- α pumping is clearly seen. Left hand side panel also shows the effect of X-rays on the spectrum. See text for details.

Using the Einstein relations $B_{01} = 3B_{10}$ and $B_{10} = c^2 A_{10}/2h\nu^3$, we can then write the optical depth as

$$\tau_\nu = \int ds \sigma_{01} \left[1 - \exp\left(-\frac{h\nu}{k_B T_s}\right) \right] \phi(\nu), \quad (1.58)$$

where

$$\sigma_{01} = \frac{3c^2 A_{10}}{8\pi\nu^2}. \quad (1.59)$$

The integral simplifies if T_s is sufficiently large, so that

$$\tau_\nu \approx \sigma_{01} \left(\frac{h\nu}{k_B T_s} \right) \left(\frac{N_{\text{HI}}}{4} \right) \phi(\nu), \quad (1.60)$$

where N_{HI} is the neutral hydrogen column density, and the factor 1/4 gives the fraction of atoms in the 0 level. In our case the line profile will include effect of natural, thermal and pressure broadening. This optical depth is typically much less than unity.

A different observational strategy is to cross-correlate measurements of patches separated in angle or frequency. This allows one to detect individual large H II regions or a globally statistically averaged signal (“the 21cm power signal”). The biggest problem in this approach is foreground subtraction. Discussion of feasibility of this technique is available in the literature [See e.g. Oh & Mack, 2003; Morales, Bowman & Hewitt, 2006].

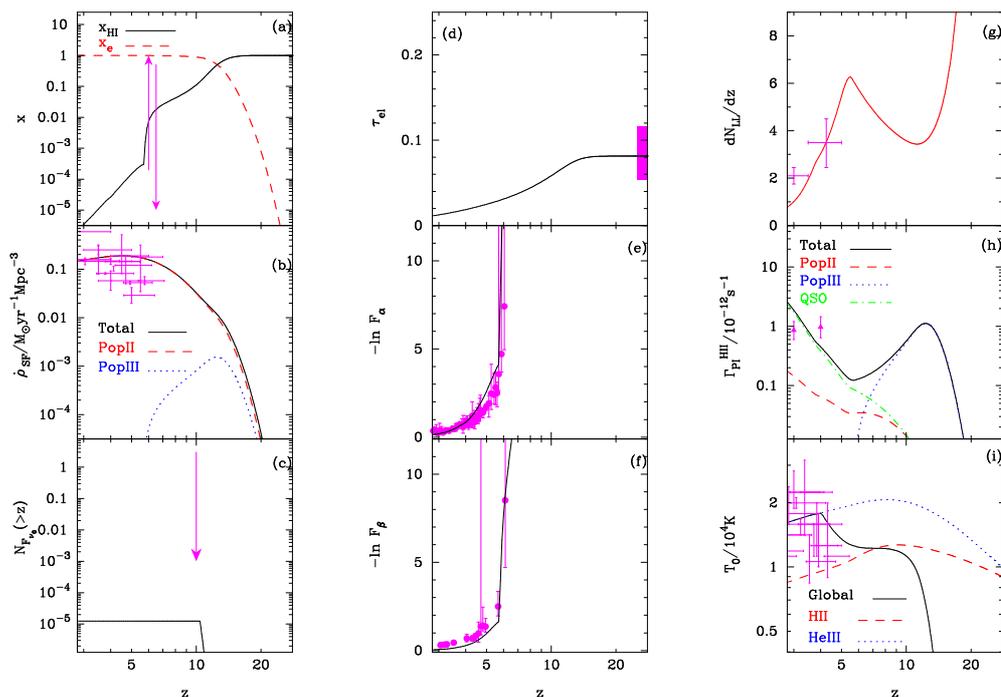


Figure 1.3: Comparison of model predictions from Choudhury & Ferrara [2006] to observations. Relevant panels indicate (a) the volume averaged H I fraction with the observational lower limit from QSO absorption lines at $z = 6$. The ionized fraction is shown with a dashed line. (b) the star formation rate density for different stellar populations, (d) the Thomson scattering optical depth, (e) Ly- α effective optical depth, (h) photoionization rate estimates and (i) temperature of the mean density IGM.

1.3.3 Modelling the IGM

After surveying our current knowledge of the IGM and the observations that have revealed it, we now discuss a basic theoretical framework required for modelling reionization of the IGM. We also discuss results from simulations of the IGM.

Photons with energies exceeding the ionization potential of a bound electron in a hydrogen atom will ionize the neutral hydrogen atoms in the IGM at the rate per neutral atom

$$\Gamma_{\text{H I}} = 4\pi \int_{\nu_{\text{H I}}}^{\infty} d\nu \frac{J_\nu}{h_P \nu} \sigma_{\text{H I}}(\nu), \quad (1.61)$$

where $\sigma_{\text{H I}}(\nu)$ is the photoelectric cross-section, $\nu_{\text{H I}}$ is the threshold frequency required to ionize hydrogen (the Lyman limit), and J_ν is the mean specific intensity of the ambient radiation field. The mean specific intensity is obtained by averaging the specific intensity I_ν over a large volume and over all directions

$$J_\nu(t) \equiv \int_V \frac{d^3x}{V} \int \frac{d\Omega}{4\pi} I_\nu(\mathbf{x}, t, \hat{\mathbf{n}}). \quad (1.62)$$

Free electrons are radiatively captured by protons at a rate per proton $n_r \alpha_{\text{H II}}(T)$, where $\alpha_{\text{H II}}(T)$ is the total rate coefficient for radiative capture, which is same as the case A radiative recombination coefficient $\alpha_A(T)$, the total rate coefficient for radiative coefficient summed over recombinations to all energy levels. As a result, in presence of ionizing radiation, the evolution of the mean neutral hydrogen density $n_{\text{H I}}$ is given by

$$\dot{n}_{\text{H I}} = -3H(t)n_{\text{H I}} - \Gamma_{\text{H I}}n_{\text{H I}} + \mathcal{C}\alpha_{\text{H II}}(T)n_{\text{H II}}(T)n_e. \quad (1.63)$$

Note that we have multiplied the recombination term by the quantity \mathcal{C} . This quantity is known as the clumping factor and is defined as

$$\mathcal{C} \equiv \frac{\langle n_{\text{H II}} n_e \rangle}{\langle n_{\text{H II}} \rangle \langle n_e \rangle} = \frac{\langle n_H^2 \rangle}{\langle n_H \rangle^2}, \quad (1.64)$$

where the angle brackets denote space average and the last equality holds when the IGM has hydrogen alone. In the presence of Helium, similar equations describe evolution of He I and He II number density. This ionization equation is usually supplemented by the evolution of the IGM temperature T , which is given by

$$\dot{E}_{\text{kin}} = -2H(t)E_{\text{kin}} + \Lambda, \quad (1.65)$$

where $E_{\text{kin}} = 3k_B T n_H$ is the gas kinetic energy and Λ is the net heating rate which takes into account all heating and cooling sources, like photoionization heating, Compton cooling and recombination cooling. The first term on the right hand side takes into account the adiabatic cooling of the gas because of cosmic expansion.

In order to model reionization, we have to solve Equation (1.63), which requires us to calculate the mean specific intensity of ionizing radiation $J_\nu(t)$. This can be done by solving the cosmological radiative transport equation, Equation (1.41), where the source term is obtained by modelling the sources of reionization. If we average Equation (1.41) over a large volume and over all angles, we get

$$\dot{J}_\nu = -3H(t)J_\nu - c\kappa_\nu J_\nu + \frac{c}{4\pi}\epsilon_\nu, \quad (1.66)$$

where the coefficients κ_ν and ϵ_ν are averaged over large volume.

We can define the volume filling factor of ionized regions, i.e. the fraction of total IGM volume occupied by ionized hydrogen, to be Q_{HII} . Reionization is said to be complete when $Q_{\text{HII}} = 1$. Further, we can write the number density of ionizing photons as

$$n_J(t) = \frac{4\pi}{c} \int_{\nu_{\text{HI}}}^{\infty} d\nu \frac{J_\nu}{h_p \nu}. \quad (1.67)$$

Since there is no ionizing flux within the neutral regions, the photoionization rate per hydrogen atom within the ionized regions is

$$\Gamma_{\text{HI}}^{\text{II}} = \frac{1}{Q_{\text{HII}}} 4\pi \int_{\nu_{\text{HI}}}^{\infty} d\nu \frac{J_\nu}{h_p \nu} \sigma_{\text{HI}}(\nu), \quad (1.68)$$

where the factor Q_{HII}^{-1} accounts for the fact that the radiation is limited to that fraction of the total volume. The emission rate of ionizing photon per unit volume from sources of emissivity ϵ_ν is

$$\dot{n}_{\text{ph}} = \int_{\nu_{\text{HI}}}^{\infty} d\nu \frac{\epsilon_\nu}{h_p \nu}. \quad (1.69)$$

The averaged radiative transport equation of Equation (1.66) can now be written as

$$\dot{n}_J = -3H(t)n_J - H(t) \frac{4\pi}{c} \frac{J_{\nu_{\text{HI}}}}{h_p} + \dot{n}_{\text{ph}} - n_{\text{HI}}^{\text{II}} Q_{\text{HII}} \Gamma_{\text{HI}}^{\text{II}} - n_{\text{HI}}^{\text{II}} \frac{dQ_{\text{HII}}}{dt}, \quad (1.70)$$

where n_i^{II} denotes the number density of species i in the H II regions. We now have to solve Equations (1.63) and (1.70) in order to get the evolution of the IGM ionization state. However, we can now reduce these two coupled differential equations into a single first-order equation by assuming (1) that the ionizing photons' mean free path is much smaller compared to the Hubble length, and (2) that the IGM is in photoionization equilibrium.

It can be shown that with the first assumption, the specific intensity J_ν can be related to the emissivity through a simple form

$$J_\nu(t) \approx \frac{\epsilon_\nu(t) \lambda_\nu(t)}{4\pi}, \quad (1.71)$$

where $\lambda_\nu = \kappa_\nu^{-1}$ is the photon mean free path. We can also ignore the terms containing J and n_J in Equation (1.70). This equation then becomes simply

$$\frac{dQ_{\text{HII}}}{dt} = \frac{\dot{n}_{\text{ph}}}{n_{\text{HII}}^{\text{II}}} - dQ_{\text{HII}}\Gamma_{\text{HI}}^{\text{II}}. \quad (1.72)$$

Now, our second assumption implies, from Equation (1.63), that

$$n_{\text{HII}}\Gamma_{\text{HI}}^{\text{II}} = \mathcal{C}\alpha(T)n_e^{\text{II}}. \quad (1.73)$$

This lets us combine Equations (1.63) and (1.70) into a single equation

$$\frac{dQ_{\text{HII}}}{dt} = \frac{\dot{n}_{\text{ph}}}{n_{\text{HII}}^{\text{II}}} - \mathcal{C}\alpha(T)n_e^{\text{II}}. \quad (1.74)$$

Figure 1.3 shows the result of solving this equation. We note in passing that the photo-heating rate can be calculated as

$$\Gamma_{\text{ph,HI}} = 4\pi \int_{\nu_{\text{HI}}}^{\infty} d\nu \frac{J_\nu}{h_p\nu} h_p(\nu - \nu_{\text{HI}})\sigma_{\text{HI}}(\nu). \quad (1.75)$$

The number of ionising photons \dot{n}_{ph} depends on the assumptions made regarding the sources. If we assume that hydrogen reionization is primarily driven by stellar sources, then $\dot{n}_{\text{ph}}(z)$ is essentially determined by the star formation rate (SFR) density $\dot{\rho}_*(z)$. The first step in this calculation is to evaluate the comoving number density $N(M, z, z_c)dMdz_c$ at redshift z of collapsed halos having mass in the range M and $M + dM$ and redshift of collapse in the range z_c and $z_c + dz_c$ [Sasaki, 1994]:

$$\begin{aligned} N(M, z, z_c)dMdz_c &= N(M, z_c)\nu^2(M, z_c)\frac{\dot{D}(z_c)}{D(z_c)} \\ &\times p_{\text{surv}}(z, z_c)\frac{dt}{dz_c}dz_c dM, \end{aligned} \quad (1.76)$$

where $N(M, z_c)dM$ is the comoving number density of collapsed halos with mass between M and $M + dM$, also known as the Press-Schechter (PS) mass function [Press & Schechter, 1974], and $p_{\text{surv}}(z, z_c)$ is the probability of a halo collapsed at redshift z_c surviving without merger till redshift z . This survival probability is simply given by

$$p_{\text{surv}}(z, z_c) = \frac{D(z_c)}{D(z)}, \quad (1.77)$$

where $D(z)$ is growth function of matter perturbations. Furthermore, $\nu(M, z_c)$ is given by $\delta_c/[D(z_c)\sigma(M)]$, where $\sigma(M)$ is the rms value of density fluctuations at the comoving scale corresponding to mass M and δ_c is the critical overdensity for collapse of the

halo. Next, we assume that the SFR of a halo of mass M that has collapsed at an earlier redshift z_c peaks around a dynamical time-scale of the halo and has the form

$$\dot{M}_*(M, z, z_c) = f_* \left(\frac{\Omega_b}{\Omega_m} M \right) \frac{t(z) - t(z_c)}{t_{\text{dyn}}^2(z_c)} \times \exp \left[-\frac{t(z) - t(z_c)}{t_{\text{dyn}}(z_c)} \right]. \quad (1.78)$$

where f_* denotes the fraction of the total baryonic mass of the halo that gets converted into stars. The global SFR density at redshift z is then

$$\dot{\rho}_*(z) = \int_z^\infty dz_c \int_{M_{\text{min}}(z_c)}^\infty dM \dot{M}_*(M, z, z_c) N(M, z, z_c), \quad (1.79)$$

where the lower limit of the mass integral, $M_{\text{min}}(z_c)$, incorporates the fact that low-mass halos do not form stars; its value is decided by different feedback processes. Here, we exclusively consider radiative feedback. For neutral regions, we assume that this quantity is determined by atomic cooling of gas within haloes (we neglect cooling via molecular hydrogen). Within ionised regions, photo-heating of the gas can result in a further suppression of star formation in low-mass haloes. We compute such (radiative) feedback self-consistently from the evolution of the thermal properties of the IGM, as discussed in Section 3.1.3.

We can then write the rate of emission of ionising photons per unit time per unit volume per unit frequency range, $\dot{n}_\nu(z)$, as

$$\dot{n}_\nu(z) = N_\gamma(\nu) f_{\text{esc}} \dot{\rho}_*(z), \quad (1.80)$$

where $N_\gamma(\nu)$ is the total number of ionising photons emitted per unit frequency range per unit stellar mass and f_{esc} is the escape fraction of photons from the halo. The quantity $N_\gamma(\nu)$ can be calculated using population synthesis, given the initial mass function and spectra of stars of different masses [Samui, Srianand & Subramanian, 2007].

Metal enrichment and reionization constraints on early star formation

The period of transition of the IGM from a completely neutral to a completely ionized state is known as the epoch of reionization [EoR; Loeb & Barkana, 2001]. The study of EoR has been an active area of research in recent years. Theoretical ideas about the reionization history have been constrained by a variety of observations [Choudhury & Ferrara, 2006; Fan, Carilli & Keating, 2006]. For example, observations of Gunn-Peterson troughs in AGN spectra at $z \sim 6$ [Becker et al., 2001; Fan et al., 2006] indicate that the process of reionization was nearly complete by that redshift. Bounds on luminosity function of $\text{Ly}\alpha$ galaxies at high redshifts [Malhotra & Rhoads, 2004; Stern et al., 2005; Bouwens et al., 2008, 2010a] also constrain the EoR. These bounds are consistent with the conclusion that the IGM was completely ionized by $z \sim 6$. Furthermore, Thomson scattering by free electrons in the IGM of the anisotropic photon distribution that constitutes the CMB leaves a signature in its temperature and polarization anisotropy. The optical depth due to this scattering [Spergel et al., 2003; Dunkley et al., 2009] is another probe of EoR. WMAP five-year data indicate a value of $\tau \sim 0.084 \pm 0.016$ that corresponds to $z \sim 10$ as the redshift for instantaneous reionization. Realistic scenarios, however, predict a protracted EoR where the process of ionization starts around $z \sim 20$ and ends by $z \sim 6$.

In this chapter, we consider the question of the sources of reionization. In the currently-favoured Λ CDM cosmological model, large scale structures in the universe like galaxies and clusters of galaxies are believed to have formed by gravitational amplification of small perturbations [Peebles, 1980; Padmanabhan, 2002; Bernardeau et al., 2002]. Much of the matter in galaxies and clusters of galaxies is the so called dark matter that is believed to be weakly interacting and non-relativistic [Trimble, 1987; Komatsu et al., 2009]. Dark matter responds mainly to gravitational forces, and by virtue of larger density than baryonic matter, assembly of matter into haloes and large scale

structure is driven by gravitational instability of initial perturbations. Galaxies are believed to form when gas in highly over-dense haloes cools and collapses to form stars in significant numbers [Hoyle, 1953; Rees & Ostriker, 1977; Silk, 1977; Binney, 1977]. The formation of first stars [McKee & Ostriker, 2007; Zinnecker & Yorke, 2007; Bromm & Larson, 2004] in turn leads to emission of UV radiation into the IGM. Due to this UV radiation, early star-forming galaxies are the most-favoured candidates as the sources of IGM reionization.

Although stellar sources are believed to be the most plausible candidates, many other sources of ionizing radiation have also been considered in the literature [Yan & Windhorst, 2004; Schneider et al., 2006; Choudhury & Ferrara, 2007]. With their hard spectra, active galactic nuclei (AGNs) of high redshift galaxies can be very effective in ionizing large regions of the IGM. However, the AGN density goes down more rapidly for $z > 3$ than the density of star-forming galaxies [Miralda-Escude & Ostriker, 1990; Madau, Haardt & Rees, 1999; Haehnelt et al., 2001] and therefore AGNs are not expected to contribute significantly to the ionizing radiation. X-rays from low mass quasars, X-ray binaries and supernova remnants are constrained by the soft X-ray background observed today [Dijkstra, Haiman & Loeb, 2004]. Particle decays can also play only a minor role in reionization [Bharadwaj & Sethi, 1998; Mapelli & Ferrara, 2005]. In this chapter, we assume that it was radiation from early stars that ionized the IGM and ignore other possibilities.

The total photon emissivity of early stars is poorly known. Studies of reionization typically use observations with semi-analytic models of heating and ionization of the IGM where efficiency of star formation, evolution of star formation rates, the number of ionizing photons emitted per baryon in stars, etc. are parameterized in some manner [Chiu & Ostriker, 2000; Choudhury & Ferrara, 2005]. Given the complexity of most of these approaches, and the number of parameters, it is often impractical to scan the parameter space. The main lesson we learn from these studies is that we may not require extraordinary physical processes in order to satisfy available observational constraints of reionization. In the approach that we take here, we make an attempt to simplify modeling of star formation and other astrophysical aspects of the problem. This allows us to reduce the number of free parameters in this sector while retaining many significant astrophysical relationships. Statements can then be made about quantities like the escape fraction of UV photons, $f_{\text{esc},\gamma}$ and their correlations with the cosmological parameters.

Large uncertainties exist in parameters related to early star formation, e.g. the escape fraction of UV photons, $f_{\text{esc},\gamma}$, the stellar initial mass function (IMF) and the efficiency of star formation, f_* [Bunker et al., 2004]. It has been suggested in the literature that a top-heavy IMF with very massive stars is not necessarily favored to satisfy the reionization and metal enrichment constraints [Daigne et al., 2006; Venkatesan & Truran, 2003].

In this work we assume that early star formation happens predominantly during

	IMF	M_{low}/M_{\odot}	$M_{\text{high}}/M_{\odot}$	Z_{input}	N_{γ}	p	$N_{\gamma}f_{\text{esc},\gamma}f_{*}$
1.	Kroupa	0.1	100	0.0004	6804	0.0123	50.0
2.	Kroupa	0.5	100	0.0004	9280	0.0167	50.0
3.	Kroupa	0.1	100	0.001	6297	0.0159	50.0
4.	Salpeter	1	100	0.001	11237	0.0283	50.0
5.	Kroupa	0.1	100	0.02	3996	0.0261	50.2

Table 2.1: Various IMFs used in this study are summarized here. M_{low} and M_{high} are the lower and upper mass cut-offs for the IMFs respectively. Z_{input} denotes the metallicity of the gas from which stars are formed. N_{γ} is the number of ionizing photons produced per baryon in stars and p is the metal yield per baryon in them. N_{γ} and p are obtained using population synthesis models. The last column lists the value of $N_{\gamma}f_{\text{esc},\gamma}f_{*}$ for the WMAP5 best-fit model with the corresponding IMF. This quantity is proportional to the number of photons escaping into the IGM for every baryon inside a collapse halo.

formation and major mergers of haloes, and occurs as a short lived burst. Photon emissivities and metal yields can then be calculated using population synthesis models for different IMFs [Leitherer et al., 1999; Bruzual & Charlot, 2003]. We then test if these scenarios generate enough photons to ionize the universe by $z \sim 6$ by comparing with the observed Thomson scattering optical depth. We also require the models to satisfy constraints arising from the observations of the metal content of the IGM. This constrains the amount of processed elements that escape from the ISM to the IGM. We can use models for outflows as a guide and put constraints on the efficiency of star formation, or use “reasonable” values of the efficiency of star formation to constrain the metal escape fraction. We can also combine the two constraints to scale out efficiency of star formation. As a result, we are able to constrain both the evolution of the universal stellar IMF during the EoR and its slope at the high mass end.

In summary, in this chapter we combine constraints of enrichment of the intergalactic medium with observations of reionization, and check whether the extra information can provide constraints on the initial mass function during early star formation. We also ask if the combined constraints be used to make useful statements with regard to other potential sources of ionization. Finally, we study any correlations between the parameters that describe star formation and cosmological parameters.

2.1 Observations

The IGM occupies most of the space and a substantial amount of matter in the Universe. Indeed, it is believed that at least half of the baryons in the universe are in the IGM. Thus it is not surprising that the observations of IGM dominate when we discuss constraints

on models of reionization. We introduce observations that are used for constraining models in this chapter.

Observational constraints that we use are essentially two: metallicity of the IGM as seen in quasar absorption systems, and the Thomson scattering optical depth from the cosmic microwave background. A third observation that can potentially be used is that of the cosmic stellar matter density, which we model in a manner explained below. We do not discuss constraints arising from this type of observations as the observational bounds on models are not very strong at present. It is not possible to use other observational constraints in the global averaged model discussed here [Choudhury & Ferrara, 2007].

2.1.1 Metallicity of the IGM

Observations of absorption systems in quasar spectra have been used to put constraints on the average metallicity of the IGM [Cowie et al., 1995; Songaila, 1997; Ellison et al., 2000; Simcoe, Sargent & Rauch, 2004; Becker, Rauch & Sargent, 2009; Ryan-Weber et al., 2009]. These observations indicate that the amount of C IV in the IGM does not evolve significantly between $2 \leq z \leq 5.5$ [Songaila, 2001; Becker, Rauch & Sargent, 2009; Ryan-Weber et al., 2009]. It is not yet clear whether the IGM has been contaminated by metals throughout, or if the enriched regions of the IGM are restricted to the neighborhood of galaxies and filaments. There are also issues related with understanding the ionization state of metals to map the absorption by a particular species to average metallicity [Schaye et al., 2003]. Observations also support a correlation between density and metallicity of the IGM, indicating that regions in proximity of galaxies are enriched to higher level than regions of IGM far away from galaxies [Schaye et al., 2003; Pieri, Schaye & Aguirre, 2006; Scannapieco et al., 2006]. We assume that the IGM is uniformly enriched at the level indicated by Songaila [2001] at $z \sim 5.5$.

Winds and outflows are expected to be the dominant processes that lead to ejection of metals from the inter-stellar medium (ISM) of galaxies. There is considerable evidence in favor of this mechanism as observations have detected outflows around almost all galaxies at high redshifts [Pettini et al., 2001; Frye, Broadhurst & Benítez, 2002]. The fraction of processed metals that can be deposited from the ISM to the IGM without disturbing the IGM in an observable manner is not known. Several authors often assume that around 1% of metals produced in galaxies can be ejected and deposited in the IGM. This may also be computed from first principles in detailed models [Daigne et al., 2004, 2006; Samui, Subramanian & Srianand, 2008]. In these models supernova-driven outflows are responsible for the IGM enrichment. The efficiency of these outflows depends on the star formation efficiency, the IMF and the efficiency of winds [Madau, Ferrara & Rees, 2001; Scannapieco, Ferrara & Madau, 2002; Scannapieco, 2005; Furlanetto & Loeb, 2003]. For a star formation efficiency of 10% the volume filling factor of the ejecta can be 20–30% and at $z \sim 3$ the IGM metallicity could be around $[-3]$ as detected

by Songaila [2001].

2.1.2 CMB constraints on IGM reionization

Finally, observations of the temperature and polarization anisotropies in the CMB provide a constraint on the EoR. Free electrons produced during reionization scatter the CMB photons and suppress temperature anisotropies on scales smaller than the Hubble radius at that time. The suppression is of the form $\mathcal{C}_l^{T'} = e^{-2\tau} \mathcal{C}_l^T$ where

$$\tau = \int n_e(t) \sigma_T c dt \quad (2.1)$$

is the Thomson scattering optical depth. Here $n_e(t)$ is the number density of free electrons and σ_T is the Thomson scattering cross-section. This damping is degenerate with the amplitude of the primordial power spectrum. The degeneracy is broken by detection of a polarization anisotropy for scales greater than the Hubble radius at the reionization redshift, an effect that dominates at scale of the Hubble radius at EoR and has an amplitude proportional to τ . Any model of reionization must reproduce the observed value of optical depth¹.

2.2 Analytical Model

The reionization history depends on the star formation history of the universe, which in the simplified models is closely related to the halo formation history. The IMF of stars and the escape fraction for ionizing photons then give us the number of ionizing photons that are available as a function of time. These can then be used to compute the evolution of the neutral or ionized fraction of gas in the universe. Here, we assume that star formation is triggered during formation of haloes. As most time scales of interest are longer than the dynamical time scale over which the bulk of star formation takes place, we assume star formation to be instantaneous in our model.

Observations of galaxies at high redshifts can be used to infer the density of matter in stars, ρ_* , at those redshifts. One way of estimating this quantity is to use the luminosity function of galaxies in various wavelength passbands and combine these with population synthesis models and an assumed IMF [Madau et al., 1996; Lilly et al., 1996; Bouwens et al., 2007]. We do not use these observations here as current observations do not provide sufficiently strong constraints. This is expected to change in coming years with better observations.

¹For a forecast on anticipated improvements of observational estimation of τ , please see Colombo & Pierpaoli [2009].

Instead, we assume that stars form only in virialised haloes. We take the minimum mass of star-forming haloes to be $10^8 M_\odot$, this is a proxy for haloes with a virial temperature of 10^4 K. We do not take into account star formation in haloes of around $10^6 M_\odot$ that is aided by molecular cooling [Tegmark et al., 1997]. We also ignore the effects of feedback that raise the Jeans mass to around $10^9 M_\odot^2$. We can then obtain metal and photon yields by using population synthesis models: we use the STARBURST99 code [Leitherer et al., 1999; Vázquez & Leitherer, 2005]. This assumption allows us to connect the rate of change of stellar mass to the rate of change of the total mass contained in massive haloes. We obtain this rate from the Press-Schechter formalism for a Gaussian PDF as

$$\dot{F}(m, z) = -\sqrt{\frac{2}{\pi}} \frac{\dot{z} \delta_c (1+z)^3}{\sigma(m) d_+(z)} \exp\left(\frac{-\delta_c^2}{2\sigma^2(m)}\right) \frac{\partial \log d_+}{\partial \log a}, \quad (2.2)$$

where an overdot denotes derivative with respect to the cosmic time, a prime denotes derivative with respect to the redshift, and F is the fraction of haloes with mass greater than m [Press & Schechter, 1974]. The critical density for spherical collapse is symbolized by δ_c , and $\sigma^2(m)$ is the variance in the initial density fluctuation field when smoothed with a top-hat filter of a scale corresponding to mass m . The rate of growth for perturbations in the linear theory is denoted by $d_+(z)$.

The total amount of baryons added to haloes of mass greater than $10^8 M_\odot$ is taken to be the amount of gas available for star formation. Gas already present in haloes is not considered for star formation. We do not consider the contribution of minihaloes as these do not contribute significantly to the total star formation due to radiative feedback [Trenti & Stiavelli, 2009]. We define the efficiency of star formation, f_* as the fraction of this gas that is converted into stars. The rate of change of stellar mass can now be written as

$$\dot{\rho}_*(z) = f_* \Omega_b \rho_c \dot{F}(10^8 M_\odot, z), \quad (2.3)$$

where ρ_c is the critical density and Ω_b is the density parameter for baryons. Here we have ignored mass lost by stars through winds, outflows, and supernovae. This can be taken into account using, for example, population synthesis models³. Equation (2.3) illustrates the small number of parameters and approximations that go into estimating $\dot{\rho}_*$ in our formulation.

2.2.1 Reionization

We consider a globally averaged evolution of ionized fraction instead of following evolution of HII regions around haloes, the approach used in most studies [Chiu & Ostriker,

²We have checked that including the effects of radiative feedback increases the required star formation efficiency by around 20%.

³A starburst with a Kroupa IMF and an initial metallicity of 0.02 loses about 10% of its mass to the ISM through these effects under normal assumptions.

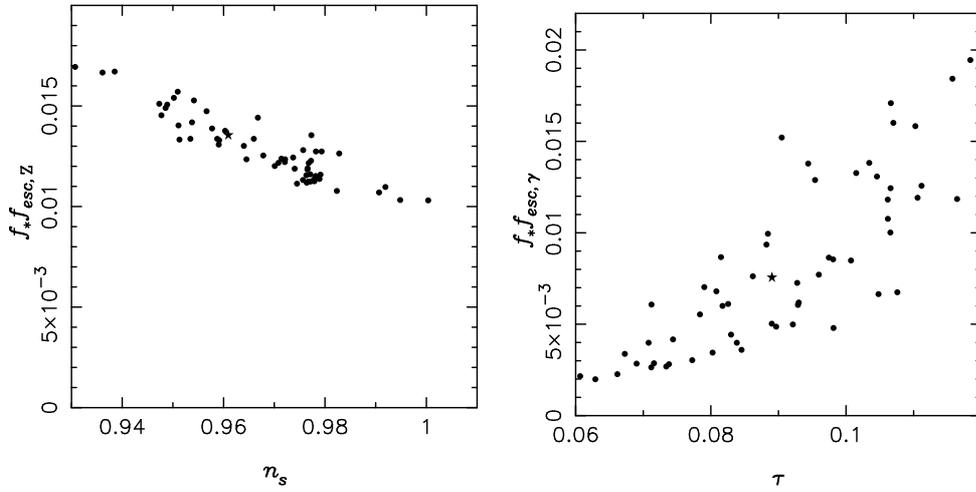


Figure 2.1: The products $f_* f_{\text{esc},Z}$ and $f_* f_{\text{esc},\gamma}$ against cosmological parameters n_s and τ for our fiducial model — model 1 of Table (2.1). The star symbol denotes the WMAP5 best fit model. These points also represent lower bounds on $f_{\text{esc},Z}$ and $f_{\text{esc},\gamma}$. Please see text for details.

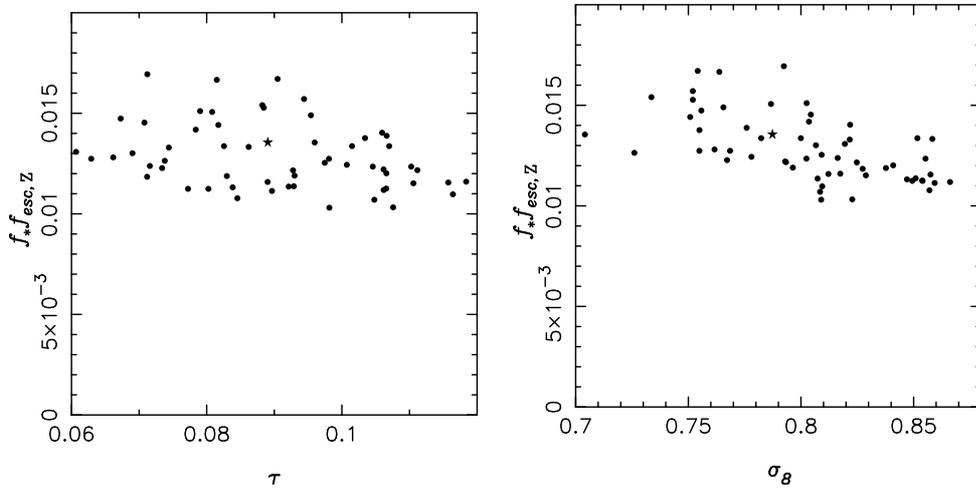


Figure 2.2: The product $f_* f_{\text{esc},Z}$ against cosmological parameters τ and σ_8 for model 1 of Table (2.1). The star symbol denotes the WMAP5 best fit model.

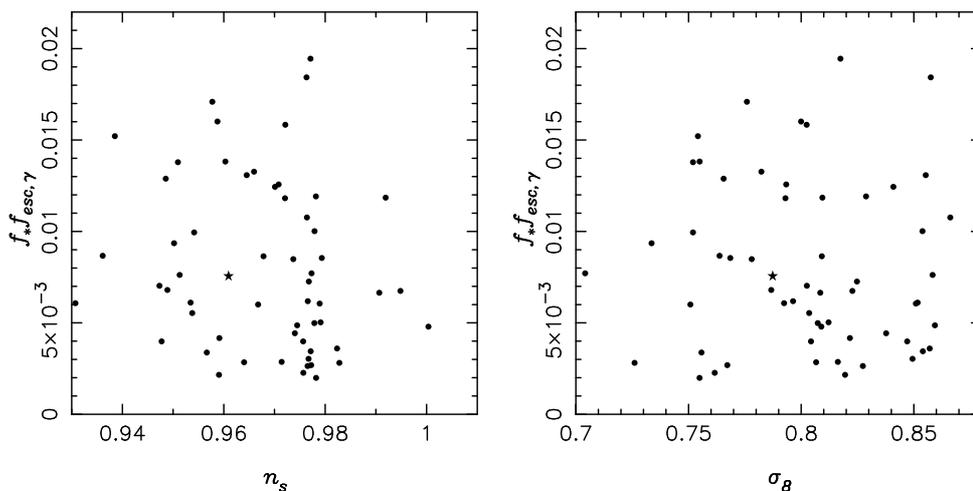


Figure 2.3: The product $f_* f_{\text{esc}, \gamma}$ against cosmological parameters n_s and σ_8 for model 1 of Table (2.1). The star symbol denotes the WMAP5 best fit model.

2000; Sethi, 2005]. Further, we assume that during reionization, a region is either neutral or completely ionized. With these assumptions, the evolution of the ionized fraction evolves as:

$$\dot{x} = -\alpha_B \mathcal{C}^2 n_H x + \sigma_p y n_H c (1 - x), \quad (2.4)$$

$$\dot{y} = -\sigma_p y n_H c (1 - x) + m_p f_* f_{\text{esc}, \gamma} \dot{F} N_\gamma, \quad (2.5)$$

where x is the fractional volume that is ionized, and y is the number of ionizing photons per baryon. σ_p denotes the effective cross-section of photoionization, α_B is the recombination coefficient for all levels except the ground state of neutral hydrogen, and m_p denotes the mass of a proton.

The first term on the right hand side of equation (2.4) describes recombination. \mathcal{C} is the clumping factor defined as $\mathcal{C}^2 = \langle n_H^2 \rangle / \langle n_H \rangle^2$. This term usually involves square of the ionized fraction but in our model we assume that the ionized fraction is either unity or zero. This, when used in volume averaging over the universe with an additional assumption that the clumping is the same in ionized and neutral regions, leads to a linear dependence. In the process of averaging, the meaning of x changes from the ionized fraction to the volume filling fraction of the ionized regions. We can express this in terms of equations:

$$\frac{1}{V} \int n_H^2 x^2 dV = \frac{\langle n_H^2 \rangle}{V} \int x^2 dV = \frac{\langle n_H^2 \rangle}{V} \int x dV = \langle n_H^2 \rangle x. \quad (2.6)$$

We have assumed that the clumping factor is the same in all parts of the universe, this allows us to take $\langle n_H^2 \rangle$ outside the integral. The third equality in equation (2.6) follows

from the definition of x as a filling fraction. We also expect \mathcal{C} to change with redshift due to the evolution of clustering. We take this dependence to be of the form

$$\mathcal{C}^2 \simeq 26.2917 \exp[-0.1822z + 0.003505z^2], \quad (2.7)$$

as obtained from high-resolution simulations by Iliev et al. [2007].

Sources of ionizing radiation are represented in the last term of equation (2.5), \dot{F} being related to the formation rate of collapsed haloes. This is obtained from the Press-Schechter formalism as described above. N_γ denotes the number of photons produced per unit mass of star formation. Ionization of neutral hydrogen is described in the last term on the right hand side of equation (2.4). This term occurs in both equations. We neglect the contribution of collisional ionization.

2.2.2 Metal Enrichment

The amount of metals produced per baryon in stars can also be computed once we fix the initial mass function (IMF) of stars and metallicity of the star-forming gas [Leitherer et al., 1999; Vázquez & Leitherer, 2005]. We can write

$$n_Z = f_* f_{\text{esc},Z} \Omega_b \frac{\rho_c}{m_p} \dot{F}(10^8 M_\odot, z) p \quad (2.8)$$

for the number density of metals that reaches the IGM. Here $f_{\text{esc},Z}$ is the fraction of total metals produced that is deposited in the IGM and p is the metal yield of the stars per baryon. Note that we assume the same escape fraction for metals from galaxies of different masses, whereas it is far more likely that low mass galaxies lose nearly all the metals from the ISM and more massive galaxies lose very little [Dekel & Silk, 1986].

Having modeled the ionization and metal enrichment of the IGM, we solve these equations numerically for different cosmological models. The system of equations (2.4) and (2.5) is “stiff,” since $(x = 0, y = 0)$ is a stable point and time scales for evolution of x and y are very different. Further, x is bounded from above (by unity) while y is not. Thus the usual forward differencing methods do not give accurate solutions easily. We bypass this problem by noting that during the process of reionization almost every ionizing photon will be immediately absorbed by the medium⁴. This means that the two terms in the right hand side of the second equation are of the same order till x becomes nearly equal to 1, whereas the left hand side is much smaller and may be assumed to be zero. This reduces the system of equations to a single equation, which can now be solved using forward differencing methods. Note that this approximation is not valid

⁴There are two approximations being discussed here: one approximation is that all UV photons are available for ionizing atoms. This is not really true as ionized regions can be expected to host an ionizing background. The other approximation is in solving the equations that we arrive at with the first approximation.

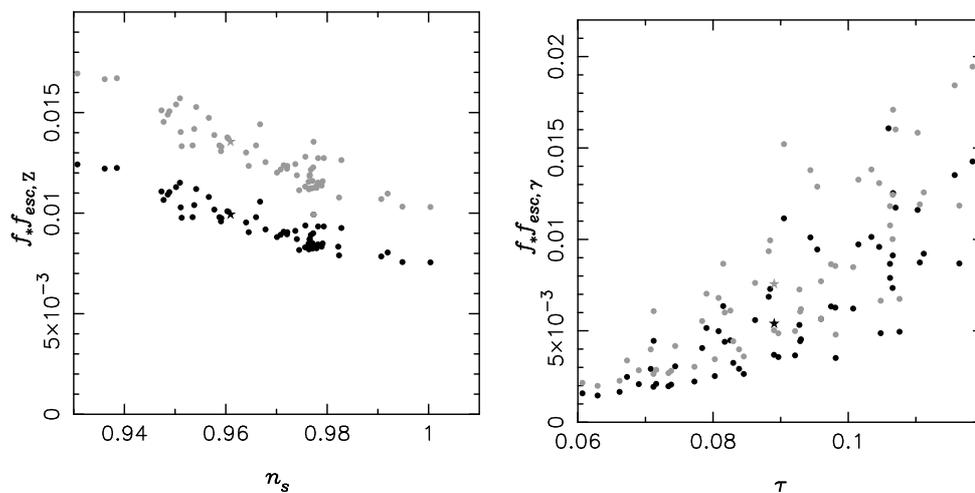


Figure 2.4: The products $f_* f_{\text{esc},Z}$ and $f_* f_{\text{esc},\gamma}$ for model 2 of Table (2.1). Corresponding plots from Figure 2.1) are superimposed in grey. Filled stars denote values for the WMAP5 best-fit models.

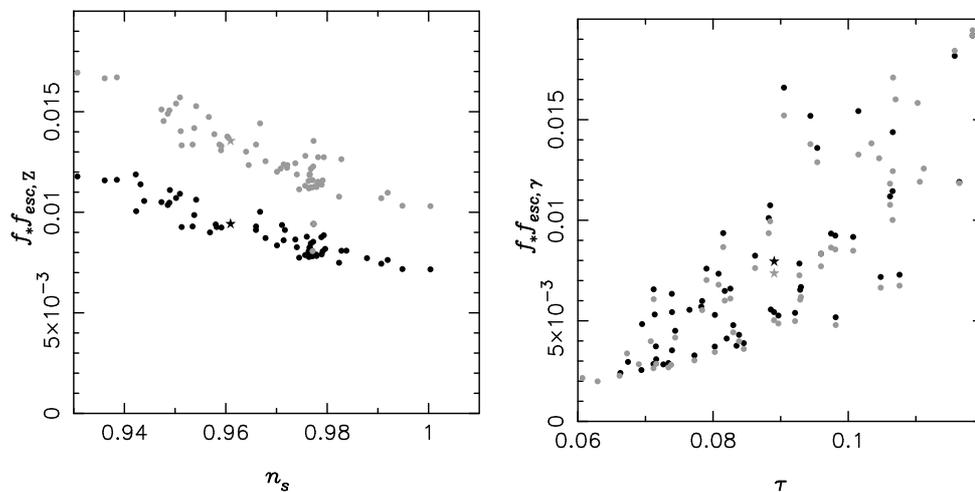


Figure 2.5: The products $f_* f_{\text{esc},Z}$ and $f_* f_{\text{esc},\gamma}$ for model 3 of Table (2.1). Corresponding plots from Figure (2.1) are superimposed in grey. Filled stars denote values for the WMAP5 best-fit models.

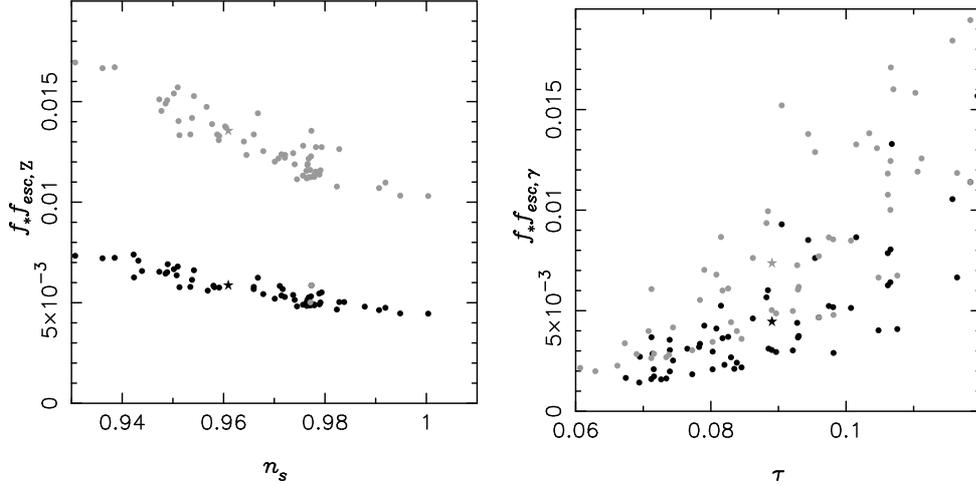


Figure 2.6: The products $f_* f_{\text{esc},Z}$ and $f_* f_{\text{esc},\gamma}$ for model 4 of Table (2.1). Corresponding plots from Figure (2.1) are superimposed in grey. Filled stars denote values for the WMAP5 best-fit models.

when x approaches 1, although in practice the approximate solution is fairly accurate up to $x \sim 0.9$. Indeed, if we use the approximation up to $x = 1.0$ then we make an error in estimation of τ of less than 5%.

We take $\alpha_B = 1.0 \times 10^{-13} \text{ cm}^3 \text{ sec}^{-1}$, ignoring its dependence on temperature. This dependence is fairly weak at temperatures of interest. We use $\sigma_p = 6.30 \times 10^{-18} \text{ cm}^2$. We thus assume that most of the ionizing radiation is around the Lyman limit. The number of ionizing photons released per baryon of stars formed, denoted by N_γ , depends on the initial mass function (IMF) of the stars. We obtain this number from the STARBURST99 stellar population synthesis code⁵ [Leitherer et al., 1999; Vázquez & Leitherer, 2005].

2.3 Results

Figure 2.7 shows evolution of ionization fraction in our model for two values of (f_*, f_{esc}) when a constant clumping factor is used. Left panel shows result for $(f_*, f_{\text{esc}}) = (0.1, 0.1)$ and $(f_*, f_{\text{esc}}) = (0.3, 0.5)$. Higher values of these parameters increase the emissivity of ionizing radiation, thereby causing early reionization. Three curves in each panel correspond to three different values of the clumping factor, given by $\mathcal{C} = 1$

⁵We consider instantaneous starbursts with a fixed total stellar mass of $10^6 M_\odot$. We use the Geneva evolutionary tracks for models with metallicity 0.001 and the Padova tracks with AGB stars for models with metallicity 0.0004.

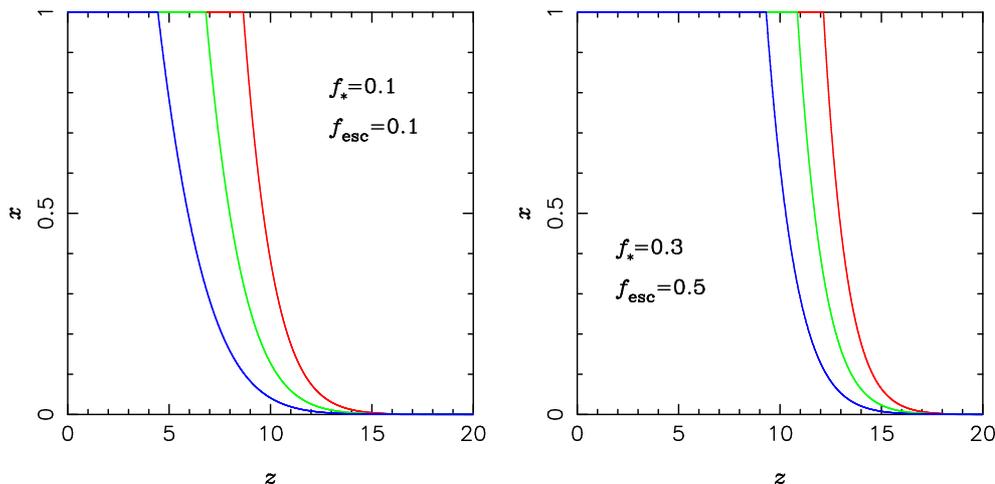


Figure 2.7: Result of our model for two values of (f_*, f_{esc}) when a constant clumping factor is used. Left panel shows result for $(f_*, f_{\text{esc}}) = (0.1, 0.1)$ and $(f_*, f_{\text{esc}}) = (0.3, 0.5)$. Higher values of these parameters increase the emissivity of ionizing radiation, thereby causing early reionization. Three curves in each panel correspond to three different values of the clumping factor, given by $\mathcal{C} = 1$ (red), 5 (green), 10 (blue).

(red), 5 (green), 10 (blue). Clearly, larger values of the clumping factor result in high rate of recombination, thereby slowing down reionization.

Our aim now is to study a variety of models with varying cosmological parameters as well as parameters related to star formation and enrichment. We consider a random subset of flat Λ CDM models allowed by WMAP 5-year data [Komatsu et al., 2009; Dunkley et al., 2009]. We do not consider models with massive neutrinos or a non-vanishing tensor component, or models where the primordial power spectrum deviates from a pure power law. We use only WMAP constraints for limiting cosmological parameters. We used the MCMC chains made available by the WMAP team. We considered a random subset of all models allowed with a confidence level of 68% from the MCMC chains. We studied a handful of models for parameters related to star formation; details of these are given in Table (2.1). The table lists the IMFs used in our study⁶. We have also listed the amount of ionizing photons produced per baryon in stars, and the total metal yield per baryon for these IMFs. These numbers also depend on the metallicity of gas from which stars form and this is listed in the table as well. It is interesting to note that for a given IMF, as the metallicity increases, the production of ionizing photons per

⁶We should note that there is considerable uncertainty in the shape of the IMF in the local neighborhood [Kroupa, 2002; Conroy, Gunn & White, 2009]. This can easily have a significant impact on our conclusions. The uncertainties introduced by other assumptions and approximations should be seen with the uncertainty in the IMF as the reference.

baryon comes down but the amount of enriched material returned to the ISM increases. Gas that forms the first stars is likely to have primordial abundance [Olive, Steigman & Walker, 2000]. In our analysis of all the models, we keep metallicity fixed and hence it is appropriate to use low values of input metallicity.

Metallicity of the IGM constrains the product $f_* f_{\text{esc},Z}$ for a given model. Similarly we constrain the product $f_* f_{\text{esc},\gamma}$ with the optical depth due to reionization for the CMB. Figures 2.1, 2.2 and 2.3 shows these products for all cosmological models studied here, when the star formation parameters for model 1 in Table (2.1) are used. Given that the efficiency of star formation can at best be 100%, i.e., $f_* \leq 1$, the points in these figures also represent lower bounds on $f_{\text{esc},Z}$ and $f_{\text{esc},\gamma}$. These are shown as a function of the slope of the primordial power spectrum (n_s), optical depth due to reionization (τ_{CMB}), and, amplitude of clustering at the scale of $8 \text{ h}^{-1} \text{ Mpc}$ (σ_8). The best fit WMAP5 model is marked in each panel as a star. We find that there is some correlation between the lower bound on $f_* f_{\text{esc},\gamma}$ and τ_{CMB} , and also between $f_* f_{\text{esc},Z}$ and n_s as highlighted in Figure 2.1. There are weak correlations with other cosmological parameters, as seen in Figures 2.2 and 2.3 but nothing as remarkable as the two mentioned above.

2.3.1 Constraints on IMF evolution

The efficiency of star formation is likely to be much less than unity in any realistic scenario. Indeed, if we try to keep the different efficiencies and escape fractions at the same order then we require these to be around 0.1–0.15, or 10–15%. While the escape fraction for ionizing photons and star formation efficiency we have obtained are comparable to those found in other studies, these are higher than the values seen in local galaxies. In particular, it is not clear if it is possible to expel 10% of the metals from the ISM to distant parts of the IGM using known physical mechanisms. Before commenting on the numbers, let us consider the sensitivity of the result to our assumptions by constraining the evolution of the IMF.

If the IMF has a low mass cutoff that is higher than the $0.1 M_{\odot}$ used for model 1 from Table 2.1, then a larger fraction of mass goes into high mass stars that produce the ionizing photon flux and the enriched material. This can lower the required $f_{\text{esc},Z}$ by a significant amount. The product $f_* f_{\text{esc},Z}$ for model 2 from Table 2.1 is shown in Figure 2.4. There is some evidence that there are more intermediate mass stars in the population of metal poor stars in the halo of the Galaxy as compared to metal rich stars, if we normalize the two distributions at low stellar masses [Tumlinson, 2007; Komiya et al., 2007]. Thus a higher cutoff for M_{low} may be required for explaining other observations. Also, our analysis assumes that the metallicity of gas that forms stars is fixed. If we do a self-consistent analysis where this is allowed to evolve, the gas metallicity gradually increases. It is then clear from Table 2.1 that later generations of stars will enrich the ISM faster. As an illustration, we can see the results of analysis with higher fixed input metallicity for models 2–4 in Figures 2.5 and 2.6. Our estimates show

that $f_{\text{esc},Z}$ required to satisfy observations can come down by a few tens of percents due to this. There is a corresponding increase in $f_{\text{esc},\gamma}$ due to stars with higher metallicity producing fewer ionizing photons.

Apart from the IMF, we have assumed that haloes with mass above $10^8 M_{\odot}$ can form stars. The standard approach is to assume that radiative feedback during the EoR increase this by about an order of magnitude in regions that have been photo-ionized [Efsthathiou, 1992]. We do not take this into account as it has been pointed out that the actual effect may only be to reduce the efficiency of star formation in lower mass haloes [Mesinger & Dijkstra, 2008]. If we do consider the effect of radiative feedback as disabling star formation then it leads to a reduction in total gas available for star formation, and hence requires slightly higher efficiencies and escape fractions. We find that this effect requires an increase in the two products by about 20%. Furthermore, we have assumed that the universe is enriched uniformly. It may very well happen that the enrichment process is effective only in the vicinity of galaxies. In such a case only overdense regions are enriched. The escape fraction of metals required can be lowered by as much as a factor of two if this is the case.

Thus we may require only around 5% of the ISM to be ejected to the IGM on an average in models with $f_* \simeq 0.2$. This is comparable with semi-analytic *ab initio* models of early star formation, outflows and IGM enrichment that have been studied in the literature. We have also assumed the same loss fraction for ISM for galaxies over the entire range of masses. This, of course, is not true. We expect that the low mass galaxies can potentially disperse a large fraction of the ISM in supernova explosions but larger galaxies can retain most of their ISM [Larson, 1974; Dekel & Silk, 1986]. If most of the IGM enrichment is done by metals that form in dwarf galaxies then the constraint is not very stringent. In most models, the fraction of mass in galaxies with a halo mass of less than $10^{10} M_{\odot}$ is larger than 10% even at $z \simeq 6$. If these galaxies lose a significant fraction of the ISM on an average and heavier galaxies lose very little mass then we can comfortably satisfy the constraints from enrichment of the IGM.

2.3.2 Constraints on high mass star formation

We now turn our attention to the ratio of escape fractions for photons and metals in order to draw constraints on high mass star formation. Figure 2.8 shows the ratio $f_{\text{esc},Z}/f_{\text{esc},\gamma}$ for the first IMF listed in Table 2.1. It is interesting to note that the ratio $f_{\text{esc},Z}/f_{\text{esc},\gamma}$ is of order unity, differing from unity by at most a factor of a few. Thus the fraction of ionizing photons that escape galaxies is broadly of the same order as the fraction of metals that must leak into the IGM in order to explain the observed enrichment of the IGM. We have not plotted this ratio for other IMFs in Table 2.1 as the expected change can be seen from other figures. Indeed, the change in the ratio is less than a factor two as we consider IMFs listed in rows 2–4 of the Table 2.1. The last row in Table 2.1 is more appropriate for late time star formation and we need not discuss that here.

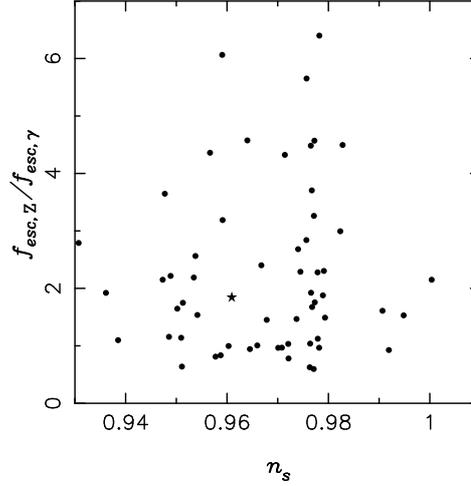


Figure 2.8: Ratio $f_{\text{esc},Z}/f_{\text{esc},\gamma}$ for our fiducial model. See text for details.

Most studies of reionization have tended to focus on the escape of ionizing photons. In order to satisfy observations of τ or the luminosity function of high redshift galaxies, these often invoke a top heavy IMF for the first generation of stars [Cen, 2003; Haiman & Holder, 2003; Wyithe & Loeb, 2003c; Bromm, 2004]. Other sources of ionizing radiation like AGNs, magnetic fields and decaying dark matter particles have also been studied [Pierpaoli, 2004; Schleicher, Banerjee & Klessen, 2008]. It is interesting to note that all such modifications lead to an enhanced production of ionizing photons without affecting the production of metals in a significant manner. This is because the very massive stars are expected to implode and do not enrich the ISM with the products of nuclear fusion that takes place in the core. The ratio $f_{\text{esc},Z}/f_{\text{esc},\gamma}$ changes on addition of extra sources of ionizing radiation. In view of the arguments presented above, all modifications that have been discussed so far lead to a smaller $f_{\text{esc},\gamma}$. In other words, the ratio plotted in Figure 2.8 should be thought of as a lower bound.

An important implication of this is that the constraints from enrichment of the IGM require certain amount of star formation, and this requirement needs to be satisfied even when we invoke other sources of ionizing radiation during the epoch of reionization. That is, adding new potential sources of ionizing radiation can be helpful only in lowering the escape fraction of ionizing photons and not in lowering the amount of star formation⁷. We may even end up with a scenario where a much larger fraction of processed elements need to be transferred from the ISM to the IGM as compared to the fraction of UV photons escaping from galaxies.

⁷An exception is the scenario where the universe is not enriched throughout. In such a case even $f_{\text{esc},Z}$ can be reduced by a significant amount.

2.4 Conclusions

In this chapter, we have compared simple models for star formation in the early universe with two observational constraints. The simplicity of the model allows us to consider variation in cosmological parameters as well. We present a summary of our results here:

- The product of star formation efficiency and escape fraction of ionizing photons, $f_* f_{\text{esc},\gamma}$, is correlated with the optical depth due to reionization.
- The product of star formation efficiency and escape fraction of ISM, $f_* f_{\text{esc},Z}$, is anti-correlated with the index of the primordial power spectrum.
- These are weak correlations, in the sense that the values for the two products do not change strongly for small changes in the cosmological parameter in question.
- We do not find any other correlation amongst parameters of star formation and cosmological parameters.
- We are able to satisfy observational constraints with the standard initial mass function for stars observed in the local universe [Kroupa, 2002], and with reasonable values for star formation efficiency and escape fractions for photons and ISM. Given that the local IMF itself is somewhat ill constrained, this implies that we do not require a significant evolution of the IMF in order to explain observations considered here.
- Small variations in the IMF, indicated by observations of metal poor stars in the Galaxy, reduce the efficiency of star formation and the escape fractions required for the standard IMF.
- Approximations used by us in the model do not change the overall numbers by more than 10–20%. Indeed, different approximations change numbers in different directions so we can consider the overall results to be fairly robust.
- Our model allows us to estimate the ratio of the two escape fractions. We find that the two escape fractions are of the same order.
- If we consider other potential sources of ionizing photons then the required escape fraction for photons can come down, however the escape fraction for processed elements does not change. Indirectly, the required amount of star formation is required to remain the same unless there is some very efficient mechanism for transporting processed elements into the IGM while keeping the escape fraction of ionizing photons low.

The most important conclusion of this chapter is that star formation without a significant evolution of the IMF is sufficient for satisfying the two constraints considered here. The escape fractions, and/or the star formation efficiency is required to be higher than we see in local galaxies. One can consider other sources of ionizing radiation, indeed at least some of these must be present. But as we have pointed out, these help in reducing only the escape fraction for ionizing radiation as none of the other potential sources help in transporting enriched material from the inter-stellar medium to the inter-galactic medium. This highlights the significance of the constraint arising from enrichment of the IGM for epoch of reionization studies.

Finally, we have neglected the presence of other sources of reionization, e.g., metal-free stars, minihaloes, and so on. It is expected that these sources would be too faint to affect the luminosity function in the ranges we are considering. Recently, such structures have also been shown to supersonic coherent flows of baryons relative to the underlying potential wells created by the dark matter [Tseliakhovich & Hirata, 2010]. However, these sources may affect the thermal history of the medium, e.g, the metal-free stars would produce higher temperatures because of harder spectra. In such cases, it is most likely that feedback would occur at magnitude brighter than what we have indicated and hence would possibly be easier to detect.

Reionization and feedback in overdense regions at high redshift

Deep surveys have now discovered galaxies at redshifts close to the end of reionization [Bouwens & Illingworth, 2006; Iye et al., 2006; Bouwens et al., 2007; Henry et al., 2007; Stark et al., 2007; Bouwens et al., 2008; Bradley et al., 2008; Henry et al., 2008; Ota et al., 2008; Richard et al., 2008; Bunker et al., 2010; Bouwens et al., 2009; Henry et al., 2009; McLure et al., 2009; Oesch et al., 2009; Ouchi et al., 2009b; Bouwens et al., 2011; Ouchi et al., 2009a; Sobral et al., 2009; Zheng et al., 2009; Oesch et al., 2010; Castellano et al., 2010; Bouwens et al., 2010a; Hickey et al., 2010; McLure et al., 2010]. Luminosity function of these galaxies, and its evolution, can answer important questions about reionization. Indeed, much work has been done on constructing self-consistent models of structure formation and the evolution of ionization and thermal state of the IGM that explain these observations [Choudhury & Ferrara, 2005; Haiman & Cen, 2005; Wyithe & Loeb, 2005; Choudhury & Ferrara, 2006; Dijkstra, Wyithe & Haiman, 2007; Samui, Srianand & Subramanian, 2007; Iliev et al., 2008; Samui, Srianand & Subramanian, 2009]. Studies of the Gunn-Peterson trough [Gunn & Peterson, 1965] at $z \geq 6$ have established that the mean neutral hydrogen fraction is higher than 10^{-4} (e. g. Fan et al. 2006) and it is most likely that the IGM is still highly ionized at these redshifts [Gallerani et al., 2008a,b]. Furthermore, CMB observations indicate the electron scattering optical depth to the last scattering surface to be $\tau_e = 0.088 \pm 0.015$ based on the WMAP seven year data. A combination of high redshift luminosity function data with the data from these absorption systems and CMB observations favour an extended epoch of reionization that begins at $z \approx 20$ and ends at $z \approx 6$ [Choudhury & Ferrara, 2006]. In this chapter, we study the luminosity function in order to find what we can learn from its observation. We will see that the finiteness of field of view of any luminosity function observation can be an independent probe of the reionization history of the IGM.

Interpreting high redshift luminosity functions is not straightforward and detailed modeling is required. For instance, local HII regions around these galaxies can affect luminosity function evolution [Cen, Haiman & Mesinger, 2005] and clustering of galaxies can enhance this effect [Cen, 2005]. Another complication is because of the fact that these surveys can detect only the brightest galaxies at these high redshifts ($z \gtrsim 6$). Such galaxies can form first in highly overdense regions and therefore the surveyed volume is far from average. An important question in that case is whether reionization proceeds differently in such regions [Wyithe & Loeb, 2007].

Galaxy formation is enhanced in overdense regions because of a positive bias in abundance of dark matter haloes. The enhancement in the number of galaxies is proportional to the mass overdensity in the region, with the constant of proportionality ('bias') related to halo masses and collapse redshifts [Cooray & Sheth, 2002]. This increases the number density of sources of ionising radiation and aids reionization of the intergalactic medium (IGM) in overdense regions. However, an increase in the IGM density also adds to radiative recombination. Furthermore, reionization is accompanied by radiative feedback [Thoul & Weinberg, 1996]. Radiative feedback heats the IGM and suppresses formation of low mass galaxies. This increase in radiative recombinations and feedback works against the process of reionization and the two effects need not cancel out. Relative significance of these negative and positive contributions will determine how differently reionization evolves in overdense regions.

Recently, Kim et al. [2009] studied a sample of i_{775} -dropout candidates identified in five Hubble Advanced Camera for Surveys (ACS) fields centred on Sloan Digital Sky Survey (SDSS) QSOs at redshifts $z \approx 6$. They compared results with those from equally deep Great Observatory Origins Deep Survey (GOODS) observations of the same fields in order to find an enhancement or suppression in source counts in ACS fields. An enhancement would imply that bias wins over negative feedback in these overdense regions. They found the ACS populations to be overdense in two fields, underdense in two field, and equally dense as the GOODS populations in one field. Somewhat surprisingly, they did not find a clear correlation between density of i_{775} dropouts and the region's overdensity.

We pursue this line of inquiry further in this chapter by using semi-analytic models to study reionization within overdense regions. The main aim is to quantify the effects of enhancement in the number of sources and radiative feedback within such regions and explore the possibility of whether the galaxy luminosity function in overdense regions can be used as a probe of feedback and reionization history. It is known that if ionization feedback is the main contributor to the suppression of star formation in low mass haloes then one can distinguish between early and late reionization histories by constraining the epoch at which feedback-related low-luminosity flattening occurs in the galactic luminosity function. The effect of reionization feedback on the high redshift galaxy luminosity function was first demonstrated using semi-analytic models by

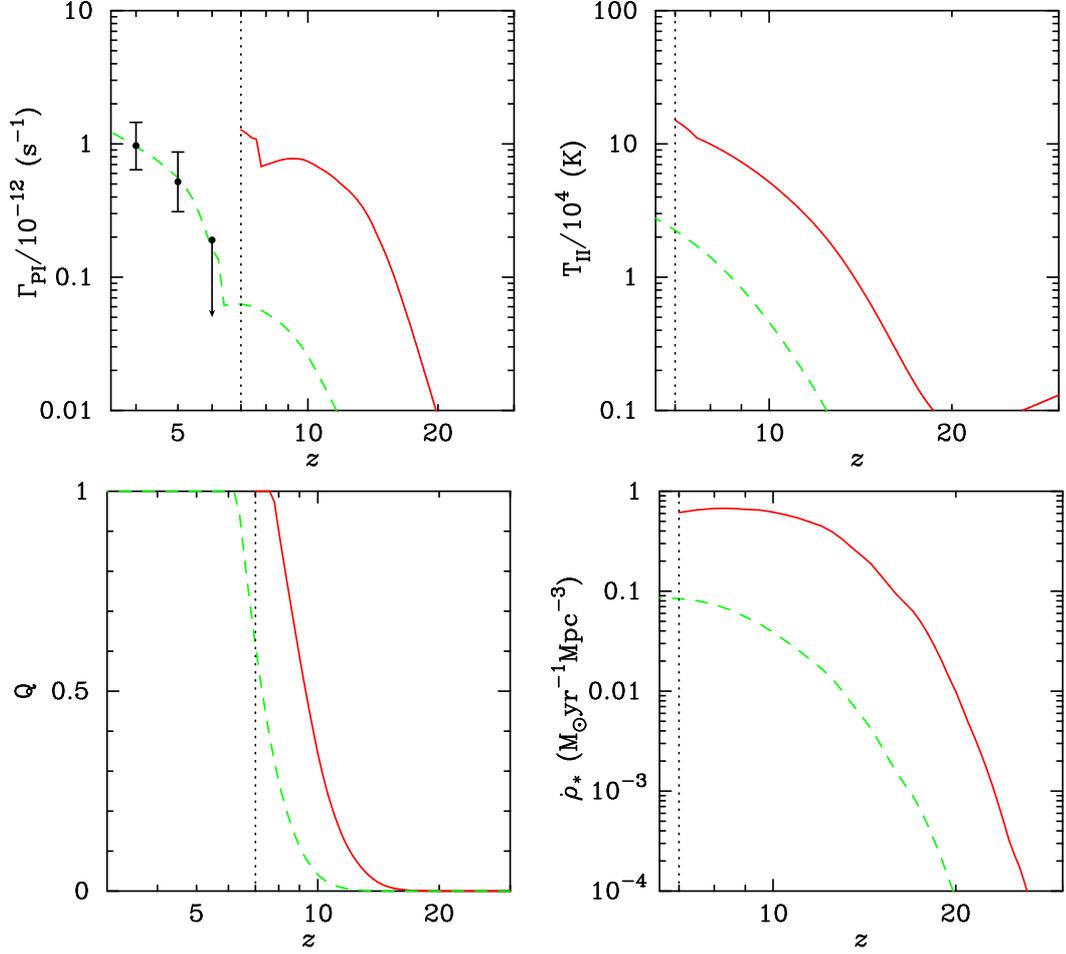


Figure 3.1: Behaviour of various quantities in our fiducial model, in the average and overdense regions are shown by dashed and solid lines respectively. The top left panel shows the photoionisation rate, with data points taken from [Bolton & Haehnelt, 2007]. The top right panel shows the mass-averaged temperature for ionised regions, which essentially determine the radiative feedback. The bottom left panel is for the volume filling factor of ionised regions. The bottom right panel shows the cosmic star formation rate. Note that the overdense region that we consider here collapses at $z = 6.8$. The vertical dotted line in the top left panel highlights this. We cannot evolve our reionization model for smaller redshifts.

Samui, Srianand & Subramanian [2007]. We apply their method to study the luminosity function in overdense regions.

3.1 Description of the Analytical model

In this section, we first summarise the basic features of the semi-analytic model used for studying the globally averaged reionization history. We then describe in detail the modifications made to this model in order to study reionization in biased regions.

3.1.1 Globally averaged reionization

Our model for reionization and thermal history of the average IGM is essentially that developed in Choudhury & Ferrara 2005 (CF05). The main features of this model are as follows.

The model accounts for IGM inhomogeneities by adopting a lognormal distribution with the evolution of volume filling factor of ionized hydrogen (H II) regions $Q_{\text{HII}}(z)$ being calculated according to the method outlined in Miralda-Escudé, Haehnelt & Rees [2000]; reionization is said to be complete once all the low-density regions (say, with overdensities $\Delta < \Delta_{\text{crit}} \sim 60$) are ionised. We follow the ionization and thermal histories of neutral and H II regions simultaneously and self-consistently, treating the IGM as a multi-phase medium. Here, we do not consider the reionization of singly ionised helium as it occurs much later ($z \sim 3$) than redshifts of our interest.

The number of ionising photons depends on the assumptions made regarding the sources. In this chapter, similar to the previous one, we have assumed that reionization of hydrogen is driven by stellar sources. The rate of ionising photons injected into the IGM per unit time per unit volume at redshift z is denoted by $\dot{n}_{\text{ph}}(z)$ and is essentially determined by the star formation rate (SFR) density $\dot{\rho}_*(z)$. The first step in this calculation is to evaluate the comoving number density $N(M, z, z_c)dMdz_c$ at redshift z of collapsed halos having mass in the range M and $M + dM$ and redshift of collapse in the range z_c and $z_c + dz_c$ [Sasaki, 1994]:

$$N(M, z, z_c)dMdz_c = N(M, z_c)\nu^2(M, z_c)\frac{\dot{D}(z_c)}{D(z_c)} \times p_{\text{surv}}(z, z_c)\frac{dt}{dz_c}dz_c dM, \quad (3.1)$$

where $N(M, z_c)dM$ is the comoving number density of collapsed halos with mass between M and $M + dM$, also known as the Press-Schechter (PS) mass function [Press & Schechter, 1974], and $p_{\text{surv}}(z, z_c)$ is the probability of a halo collapsed at redshift z_c

surviving without merger till redshift z . This survival probability is simply given by

$$p_{\text{surv}}(z, z_c) = \frac{D(z_c)}{D(z)}, \quad (3.2)$$

where $D(z)$ is growth function of matter perturbations. Furthermore, $\nu(M, z_c)$ is given by $\delta_c/[D(z_c)\sigma(M)]$, where $\sigma(M)$ is the rms value of density fluctuations at the comoving scale corresponding to mass M and δ_c is the critical overdensity for collapse of the halo. Next, we assume that the SFR of a halo of mass M that has collapsed at an earlier redshift z_c peaks around a dynamical time-scale of the halo and has the form

$$\begin{aligned} \dot{M}_*(M, z, z_c) = f_* \left(\frac{\Omega_b}{\Omega_m} M \right) \frac{t(z) - t(z_c)}{t_{\text{dyn}}^2(z_c)} \\ \times \exp \left[-\frac{t(z) - t(z_c)}{t_{\text{dyn}}(z_c)} \right]. \end{aligned} \quad (3.3)$$

where f_* denotes the fraction of the total baryonic mass of the halo that gets converted into stars. The global SFR density at redshift z is then

$$\dot{\rho}_*(z) = \int_z^\infty dz_c \int_{M_{\text{min}}(z_c)}^\infty dM \dot{M}_*(M, z, z_c) N(M, z, z_c), \quad (3.4)$$

where the lower limit of the mass integral, $M_{\text{min}}(z_c)$, prohibits low-mass halos from forming stars; its value is decided by different feedback processes. Here, we exclusively consider radiative feedback. For neutral regions, we assume that this quantity is determined by atomic cooling of gas within haloes (we neglect cooling via molecular hydrogen). Within ionised regions, photo-heating of the gas can result in a further suppression of star formation in low-mass haloes. We compute such (radiative) feedback self-consistently from the evolution of the thermal properties of the IGM, as discussed in Section 3.1.3.

We can then write the rate of emission of ionising photons per unit time per unit volume per unit frequency range, $\dot{n}_\nu(z)$, as

$$\dot{n}_\nu(z) = N_\gamma(\nu) f_{\text{esc}} \dot{\rho}_*(z), \quad (3.5)$$

where $N_\gamma(\nu)$ is the total number of ionising photons emitted per unit frequency range per unit stellar mass and f_{esc} is the escape fraction of photons from the halo. The quantity $N_\gamma(\nu)$ can be calculated using population synthesis, given the initial mass function and spectra of stars of different masses [Samui, Srianand & Subramanian, 2007]. Here, we have use the population synthesis code STARBURST99 [Leitherer et al., 1999; Vázquez & Leitherer, 2005] to calculate $N_\gamma(\nu)$ by evolving a stellar population of total mass $10^6 M_\odot$ with a 0.1 – 100.0 M_\odot Salpeter IMF and metallicity 0.001 (0.05 times

the solar metallicity, $Z_{\odot} = 0.02$). The total rate of emission of ionising photons per unit time per unit volume is obtained simply integrating by Equation (3.5) over suitable frequency range.

Given the above model, we obtain best-fit parameters by comparing with the redshift evolution of photoionisation rate obtained from the Ly α forest [Bolton & Haehnelt, 2007] and the electron scattering optical depth [Larson et al., 2011]. We should mention here that any model containing only a single population of atomic-cooled stellar sources with non-evolving $f_* f_{\text{esc}}$ cannot match both the Ly α forest and WMAP constraints [Choudhury, Ferrara & Gallerani, 2008; Bolton & Haehnelt, 2007]. Therefore, we choose the model which satisfies the Ly α constraints but under-predicts τ_e . In order to match both the constraints, one has to invoke either molecular cooling in minihaloes and/or metal-free stars and/or other unknown sources of reionization. This model is described by the parameter values $f_* = 0.2$ and $f_{\text{esc}} = 0.135$, and gives τ_e of 0.072. Figure 3.1 shows evolution of the filling factor of ionised regions, global star formation rate density, mass-weighted average temperature in ionised regions and average hydrogen photoionisation rate in this model (dashed curves in all panels). The filling factor of ionized regions is seen to rise monotonically from $z \approx 15$ and takes values close to unity at redshifts $z \approx 6$. Temperature of ionized regions also rises rapidly during reionization and flattens out to a few times 10^4 K at redshift $z \lesssim 4$ (not shown here). Lastly, the photoionisation rate also increases during reionization as the star formation rate builds up. However, the photoionisation rate increases rapidly with a sudden jump at $z \approx 6$ when the ionized regions overlap (filling factor becomes close to unity). This is because a given region in space starts receiving ionizing photons from multiple sources and as a result, the ionizing flux suddenly increases. This is our fiducial model, which satisfies observational constraints from Ly α forest, observations of star formation rate history, number density of Lyman-limit systems at high redshift and of the IGM temperature. In this model, reionization starts at $z \approx 15$ and is 90% complete by $z \approx 7$. Evolution of x_{HII} is consistent with constraints from Ly α emitters and the GP optical depths.

Having set up the reionization model, we then calculate the predicted luminosity function of galaxies in this model. Luminosity functions of objects are usually preferred for comparing theory with observations because of its directly observable nature. In this chapter, we closely follow the approach presented by Samui, Srianand & Subramanian [2007] to calculate the luminosity function. We obtain luminosity per unit mass, $l_{1500}(t)$, at 1500 Å as a function of time from population synthesis for an instantaneous burst. In our model, star formation does not happen in a burst, but is a continuous process spread out over a dynamical time-scale. Therefore, in order to determine the luminosity of a halo, $L_{1500}(t)$ with this kind of star formation, we convolve $l_{1500}(t)$ with the halo's star formation rate using

$$L_{1500}(M, T) = \int_T^0 d\tau \dot{M}_*(M, T - \tau, z_c) l_{1500}(\tau), \quad (3.6)$$

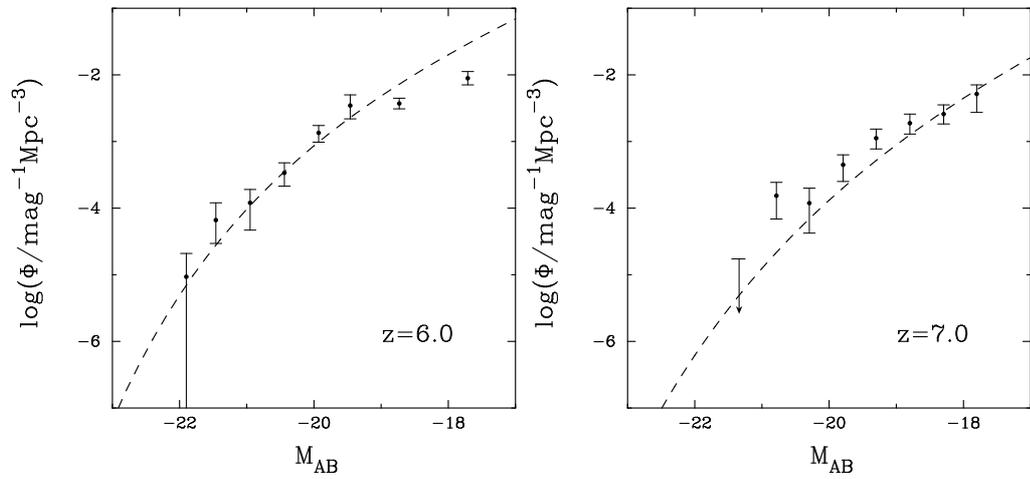


Figure 3.2: Luminosity function from our fiducial model at $z = 6$ and 7 compared with observations. This is the average case. Data points are from Bouwens & Illingworth [2006] ($z = 6$) and Bouwens et al. [2010b] ($z = 7$).

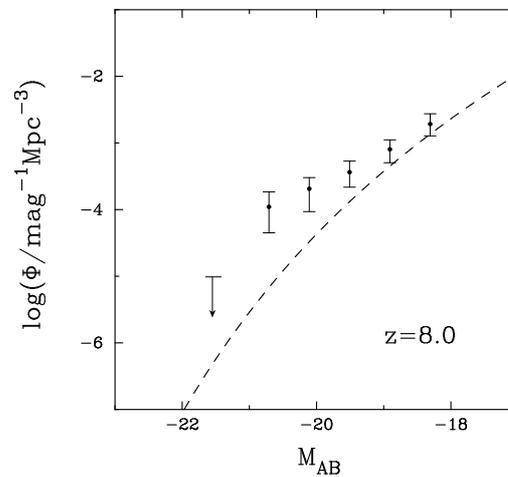


Figure 3.3: Luminosity function from our fiducial model at $z = 8$ compared with observations. This is the average case. Data points are from Bouwens et al. [2010b].

where T is the age of the halo, which has mass M and which collapsed at redshift z_c . This luminosity can be converted to absolute AB magnitude using

$$M_{AB} = -2.5 \log_{10}(L_{\nu 0}) + 51.60, \quad (3.7)$$

where the luminosity is in units of $\text{erg s}^{-1} \text{Hz}^{-1}$ [Oke & Gunn, 1983]. One can compute the luminosity evolution for any halo that collapses at redshift z_c and undergoes star formation according to Equation (3.3). The luminosity function at redshift z , $\Phi(M_{AB}, z)$, is now given by

$$\begin{aligned} & \Phi(M_{AB}, z) dM_{AB} \\ &= \int_z^\infty dz_c N(M, z, z_c) \frac{dM}{dL_{1500}} \frac{dL_{1500}}{dM_{AB}} dM_{AB}, \end{aligned} \quad (3.8)$$

where $N(M, z, z_c)$ is the number density at redshift z of halos of mass M collapsed at redshift z_c . We will use Equation (3.8) to study effect of overdensity on the luminosity function and to compare the luminosity function in our model with observations in the next section.

Figures 3.2 and 3.3 show the globally averaged luminosity function calculated using our model for different redshifts in comparison with observations presented by Bouwens & Illingworth [2006]. We find that our model reproduces the observed luminosity functions at high redshifts reasonably well. In particular, the match at $z = 6$ is remarkably good while the model predicts less number of galaxies than what is observed at $z = 7$ and 8. This could indicate that the star-forming efficiency f_* increases with z , and/or the time-scale of star formation is lower than t_{dyn} at higher redshifts. The match of the model with the data can be improved by tuning these parameters suitably, however we prefer not to introduce additional freedom in constraining the parameters; rather our focus is to estimate the effect of reionization and feedback on the luminosity function.

In our calculation of luminosities, we do not make any correction for dust. This is partly because of indications from observed very blue UV-continuum slopes [Bouwens et al., 2010d; Oesch et al., 2010; Finkelstein et al., 2010; Bunker et al., 2010] that dust extinction in $z \gtrsim 7$ is small¹. As discussed in the next section, we exclusively work with luminosity functions at these redshifts. Also, the effect of dust is degenerate with f_* to some extent. Therefore, the exclusion of dust extinction does not affect the general results of our calculation.

3.1.2 Biased regions

We have mentioned that galaxy luminosity functions provide valuable information regarding reionization. However, observations are carried out over relatively small fields

¹Although, work by Forero-Romero et al. [2010] that suggests that a clumpy dust attenuation model can reproduce the observed slopes.

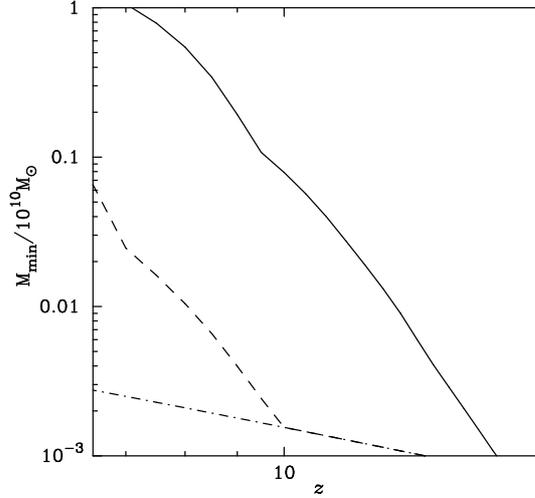


Figure 3.4: Evolution of the minimum mass $M_{\min}(z)$ of haloes that can host galaxies in ionized regions in the average case (dashed line) and the overdense case (solid line). The dot-dashed line shows the minimum mass in neutral regions, which is same for the average and overdense cases.

of view. The bright sources in these fields are typically hosted by high mass haloes. Hence, it is likely that these fields are biased tracers of luminosity function and reionization. In this section, we extend our model to study reionization within such biased regions and quantify the departure of various quantities from their globally averaged trends. Reionization in biased regions has been discussed in the literature. Wyithe & Loeb [2007] studied the correlation between high redshift galaxy distribution and the neutral Hydrogen 21 cm emission by considering reionization in the vicinity of these galaxies. Wyithe, Bolton & Haehnelt [2008] considered the ionization background near high redshift quasars. Geil & Wyithe [2009] studied effect of reionization around high redshift quasars on the power spectrum of 21 cm emission (see also Pritchard & Furlanetto 2007). The general conclusion of these studies is that overdense regions are ionised earlier. In this chapter, we will consider the behaviour of luminosity functions in such regions.

Overdense regions are characterised by their comoving Lagrangian size R , and their linearly extrapolated overdensity δ . At a given redshift, we can take a scale R corresponding to an observed field of view (e.g., WFC3/IR field in HST) and then determine δ by identifying the presence of a massive object, for example, a quasar or a bright galaxy. We follow a prescription discussed by Muñoz & Loeb [2008].

Note that if a galaxy with luminosity L_{1500} is observed at redshift z_g then we can assign a certain mass to the dark matter halo containing the galaxy, say M . The halo

mass M has to be obtained from galaxy luminosity L_{1500} by using some prescription or by fitting the galaxy's spectral energy distribution (SED; Vale & Ostriker 2004). In this chapter, however, since uncertainty in the value of overdensity δ is expected to be larger than any uncertainty in the mass of the galaxy's host halo, we choose to calculate this mass in an alternate, simpler manner. Note that we can invert equations (3.6) and (3.3) to obtain M given an observed value of L_{1500} if we have an estimate of the redshift at which the galaxy formed. In fact, we obtain the halo mass M such that the halo luminosity after a dynamical time from halo formation time, calculated according to our model, is equal to L_{1500} . In other words, we break the degeneracy between halo mass and formation redshift by assuming that the galaxy's age is equal to the dynamical time of the halo.

From this analysis, we conclude that a collapsed object with mass M exists at redshift z_g . Now suppose our field of observation corresponds to some linear scale R_o at this redshift. Then the linearly extrapolated overdensity at this scale can be obtained using the excursion set prescription [Muñoz & Loeb, 2008]. Recall that the probability distribution of the extrapolated Gaussian density field smoothed over scale R is also Gaussian

$$Q(\delta_R, \sigma^2(R))d\delta_R = \frac{1}{\sqrt{2\pi\sigma^2(R)}} \exp\left[-\frac{\delta_R^2}{2\sigma^2(R)}\right] d\delta_R. \quad (3.9)$$

The conditional probability distribution of overdensity δ_1 on a scale specified by the variance σ_1^2 , given a value of overdensity δ_2 on a larger scale specified by the variance $\sigma_2^2 < \sigma_1^2$ is given by

$$Q(\delta_1, \sigma_1^2 | \delta_2, \sigma_2^2) = Q(\delta_1 - \delta_2, \sigma_1^2 - \sigma_2^2). \quad (3.10)$$

Conversely, when the value of overdensity δ_1 on a smaller scale specified by variance σ_1^2 is given, the conditional probability distribution of δ_2 can be obtained using Bayes theorem as

$$Q(\delta_2, \sigma_2^2 | \delta_1, \sigma_1^2) \propto Q(\delta_1, \sigma_1^2 | \delta_2, \sigma_2^2) Q(\delta_2, \sigma_2^2) d\delta_2. \quad (3.11)$$

If we now set the smaller scale to be that of the observed collapsed halo, and the overdensity at that scale to be the critical overdensity for spherical collapse, Equation (3.11) will give the resulting overdensity at any larger scale due to the presence of this massive galaxy. In other words, we set $\delta_1 = \delta_c(z_g)$ and $\sigma_1^2 = \sigma^2(M)$ in Equation (3.11).

The larger scale corresponds to the field of observation. In order to calculate that, we first note that the excursion set principle functions entirely in Lagrangian coordinates. As a region evolves towards eventual collapse its Lagrangian size stays unchanged while its Eulerian size changes. For a spherical region the Eulerian evolution will follow the solution of the spherical collapse model. However, since the Eulerian and Lagrangian sizes of the region coincide at the initial instant, the spherical collapse solution is also a

relationship between these two sizes. Thus we have

$$R_E = \frac{3}{10} \frac{1 - \cos \theta}{\delta_L} \frac{D(z=0)}{D(z)} R_L, \quad (3.12)$$

where θ is a parameter given by

$$\frac{1}{1+z} = \frac{3 \times 6^{2/3} (\theta - \sin \theta)^{2/3}}{20 \delta_L}. \quad (3.13)$$

In our case, the Eulerian size of the region of interest is just the comoving distance corresponding to the angular field of view, which is just the angular diameter distance at the relevant redshift multiplied by the angular field of view. The WFC3/IR field is $136'' \times 123''$. For the best fit Λ CDM cosmology the diagonal size of this field corresponds to a comoving Eulerian distance $R_E = 1.365$ Mpc at $z = 8$. In a WFC3/IR field centred on the object UDFy-42886345 at redshift 8.0 and apparent magnitude $H_{160,AB} = 28.0$ we obtain a halo mass $M = 2.52 \times 10^{11} M_\odot$ and luminosity $L_{1500} = 2.21 \times 10^{29} \text{ erg s}^{-1} \text{ Hz}^{-1}$.

Notice, however, that since δ_L is unknown, Equation (3.12) implies that the relation between the Eulerian size R_E and Lagrangian size R_L is not one-to-one. Thus, for the probability distribution of linearly extrapolated overdensity δ given the halo mass M , we can only write

$$\frac{dP(\delta|M)}{d\delta} \propto Q[\delta, R_L(\delta, R_E)|\delta_c(z), R(M)], \quad (3.14)$$

where the constant of proportionality is calculated by using the normalization condition $\int [dP(\delta|M)/d\delta] d\delta = 1$.

In our calculations, we work with the value of δ for which $dP(\delta|M)/d\delta$ is maximum. For the WFC3/IR field at $z = 8$, this turns out to be $\delta = 8.86$ (linearly extrapolated to $z = 0$), which results in a Lagrangian size $R_L = 1.482$ Mpc for the region of interest. Notice that since $\delta > \delta_c$ the region must have collapsed at some redshift $z \lesssim 6.5$.

In order to incorporate this overdensity into our reionization model, note first that the number density of collapsed objects in such overdense regions is enhanced with respect to that in a region with average density. This enhancement can be calculated using the excursion set formalism [Bond et al., 1991]. It is then shown in the Appendix to this chapter that the comoving number density $N(M, z, z_c) dM dz_c$ at redshift z of collapsed halos having mass in the range M and $M + dM$ and redshift of collapse in the range z_c and $z_c + dz_c$ is given in this case by

$$N(M, z, z_c) dM dz_c = N(M, z_c) \left(\frac{\nu^2 \delta_c}{\delta_c/D(z_c) - \delta} \right) \frac{\dot{D}(z_c)}{D^2(z_c)} \times p_{\text{surv}}(z, z_c) \frac{dt}{dz_c} dz_c dM, \quad (3.15)$$

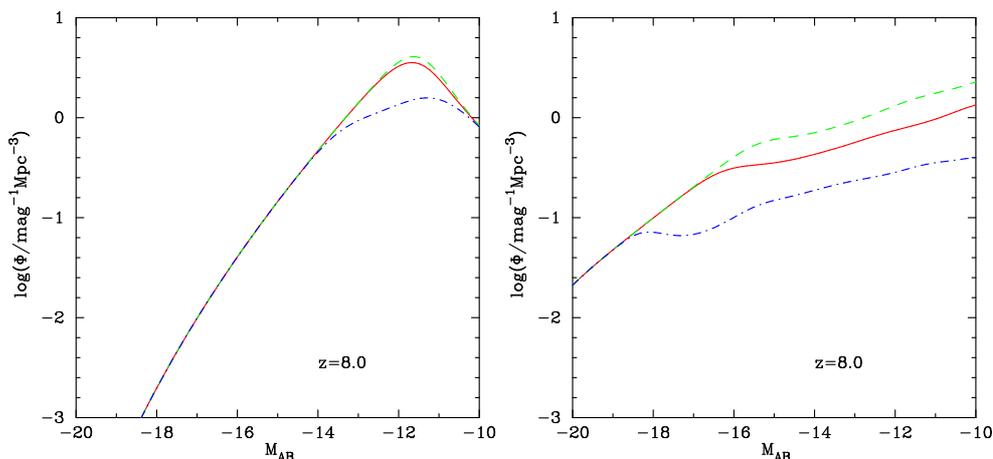


Figure 3.5: Effect of reionization history on luminosity function at $z=8.0$. Solid, dashed and dot-dashed lines have $\tau_e = 0.073$, 0.058 and 0.088 respectively. Left panel shows the average case. Right panel shows the overdense case.

where $N(M, z_c)$ is the PS mass function. The scale R enters via the definition of $\nu(M, z_c)$, which is now given by

$$\nu(M, z_c) = \frac{\delta_c/D(z_c) - \delta}{\sqrt{\sigma^2(M) - \sigma_R^2}}. \quad (3.16)$$

The survival probability $p_{\text{surv}}(z, z_c)$ is given by

$$p_{\text{surv}}(z, z_c) = \frac{\delta_c/D(z) - \delta}{\delta_c/D(z_c) - \delta}. \quad (3.17)$$

Another change when our reionization model is applied to overdense regions is that we now normalize the probability distribution of inhomogeneities in the IGM such that the average density in the region is $\Delta = \delta + 1$.

3.1.3 Radiative feedback

As we argue in the next section, the luminosity function of galaxies in an overdense region could carry an enhanced signature of feedback. We therefore highlight our feedback model in this subsection.

Radiation from stars in the first galaxies is expected to ionize and heat the surrounding medium. This increases the mass scale above which baryons can collapse in haloes. Also, as a result, the minimum mass of haloes that are able to cool is much higher in ionized regions than in the neutral ones. In our calculations, feedback appears through

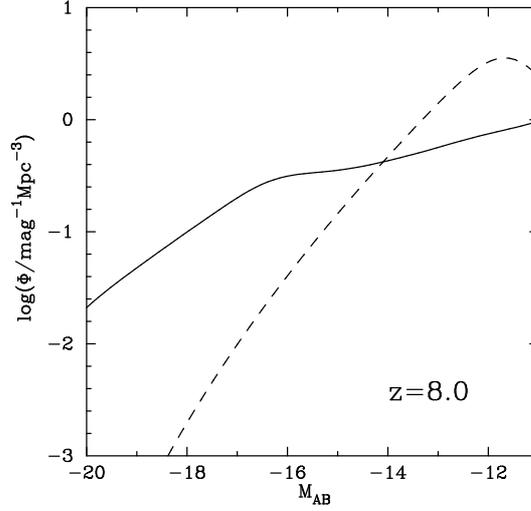


Figure 3.6: Effect of overdensity on the luminosity function via feedback at $z=8.0$. This is for $\delta = 8.8$ and $R_L = 1.482$ Mpc. Dashed line shows the average-region luminosity function, while the solid line shows overdense-region luminosity function.

the quantity $M_{\min}(z)$ in Equation 3.4. The temperature evolution of both regions is calculated self-consistently. In the ionized regions, we fix the cut-off mass to that corresponding to a virial temperature of 10^4 K or the local Jeans mass, whichever is higher. In the neutral regions, since the Jeans mass is always low, the cut-off mass always corresponds to the virial temperature of 10^4 K. The minimum mass corresponds to the circular velocity of

$$v_c^2 = \frac{2k_{\text{boltz}}T}{\mu m_p}, \quad (3.18)$$

where μ is the mean molecular weight. For a temperature of $\approx 10^4$ K, the minimum circular velocity is ≈ 25 km s $^{-1}$. Note that this value is comparable to values obtained in simulations [Gnedin, 2000] but is somewhat higher than that taken in the semi-analytic prescription of Samui, Srianand & Subramanian [2007].

We find that $M_{\min}(z)$ increases with time taking values of $\approx 10^7 M_\odot$ at $z \approx 10$ and $\approx 10^8 M_\odot$ at $z \approx 7$. In overdense regions the minimum mass is enhanced to about $10^{10} M_\odot$. Figure 3.4 shows the evolution of the minimum mass.

3.2 Results

The results for reionization and thermal histories within overdense regions are presented in this section.

3.2.1 Effect of overdensity on reionization history

We first consider the effect of overdensity on reionization history for our fiducial model. As is well known, reionization proceeds differently in overdense regions. The solid lines in Figure 3.1 show the evolution of the photoionisation rate, temperature in ionised regions, star formation rate density and the volume filling factor of ionised regions in an overdense region with size $R_L = 1.482$ Mpc and linearly extrapolated overdensity $\delta = 8.86$. This corresponds to the HUDF WFC3/IR field centred at the brightest source in Bouwens et al. [2010c]. (See Section 3.1.2.) Clearly while the average region is completely ionised at $z \approx 6$, the biased region is ionised much earlier, at $z \approx 7.5$. This result agrees with Wyithe & Loeb [2007], although note that unlike that work, here we calculate the clumping factor from a physical model for inhomogeneities. The reason for early reionization in overdense regions is the enhanced number of sources, which is clear from the plots of photoionisation rate and the star-formation rate, both of which are ~ 5 times higher than the corresponding globally averaged values. However, these overdense regions have more recombinations, which results in enhanced temperatures as is clear from the top right panel. This results in enhanced negative radiative feedback which will suppress star formation in low mass galaxies and hence affect the shape of the luminosity function. In fact, for the average case, haloes in ionised regions with masses below $10^8 M_\odot$ cannot form stars, whereas this cutoff mass rises to close to $10^{10} M_\odot$ in the overdense case. Clearly feedback is enhanced in overdense regions.

3.2.2 Effect of overdensity on luminosity function

We now discuss the effect of overdensity on luminosity function. Clearly, overdense regions tend to have enhanced number of sources, hence it is natural that the amplitude of the luminosity function for such regions should be higher than the globally averaged values. However, the overdense regions have enhanced radiative feedback too, which works towards diminishing the number of sources, particularly towards the fainter end.

Figure 3.6 shows the effect of overdensity on the luminosity function at $z = 8$ for our fiducial model. The ionised volume filling factor within the overdense region is $Q \approx 1.0$ for the overdense region under consideration at this redshift. The average region luminosity function (dashed line) is clearly very different from the luminosity function in the overdense region (solid line) at that redshift. Firstly, we can clearly see an enhancement in the source counts for brighter galaxies, which is as expected. In addition, there is a clear sign of a flattening for magnitudes $M_{AB} \gtrsim -17$, which is a signature of radiative feedback. In comparison, the effect of feedback for average regions occurs at much fainter magnitudes $M_{AB} \sim -12$. Note that there is no complete suppression of star formation for halo masses lower than the feedback threshold, rather the luminosity function for magnitudes below the knee continues to grow in the flattened region. This is simply due to the continued star formation in haloes with mass less than

the cutoff mass at $z = 8.0$, but which collapsed at higher redshifts when the feedback threshold mass was lower. Thus, for instance, if star formation is allowed to happen in a halo for only for a fraction of the dynamical time [see equation (3.3)], the luminosity function will rise less steeply at the fainter end. For small enough star formation time scale, the luminosity function will show an abrupt cutoff. Of course, an abrupt cutoff is always seen at low enough luminosities, which are not shown in the figure here.

It is important to understand here that the data points in Figures 3.2 and 3.3 do not represent luminosity function of the overdense region. Instead, those data points represent the globally averaged luminosity function derived using a maximum likelihood procedure from the observed luminosity distribution of sources. In this procedure, a likelihood function is defined, which describes the step-wise shape of the luminosity function that is most likely given the observed luminosity distribution in the search fields. Details of this procedure are described, for example, in Section 5.1 of Bouwens et al. [2010b] and references therein.

3.2.3 Luminosity function as a probe of reionization

Given the fact that the effect of radiative feedback shows up at brighter magnitudes for overdense regions, it is possible to use this feature for studying feedback using near-future observations. For this purpose, we consider two additional models (other than the fiducial one) of reionization. These models have parameter values $(f_*, f_{\text{esc}}) = (0.06, 0.3)$ and $(0.2, 0.07)$ and we obtain $\tau_e = 0.088$ and 0.058 respectively for these models. We fix f_* and only change the value of f_{esc} to ensure that any effect on the luminosity function is purely due to feedback. These two models predict photoionisation rates greater and lesser respectively than what are presented by Bolton & Haehnelt [2007].

The right panel of Figure 3.5 shows the luminosity function at $z = 8$ within the overdense region for three different reionization histories, which can be compared with the corresponding luminosity function in average region (shown in the left panel). In both cases a distinct “knee” is seen in the luminosity function as a signature of feedback. The luminosity function flattens at this luminosity, and is suppressed to very low values at much lower luminosities. As described in the previous section, this signature of feedback appears at brighter magnitudes for the overdense region. This is expected, because the cutoff mass depends directly on the temperature, which is enhanced in the overdense region. We also note that in the case of the first model the flattening occurs for $M_{\text{AB}} \gtrsim -19$ whereas for the second model at a fainter luminosity of $M_{\text{AB}} \simeq -16$. This is due to the fact that the photoionisation feedback is enhanced in the first model due to enhanced flux.

The evolution of the filling factor affects this result through the average temperature which sets the cutoff mass. Thus, early and late reionization models are distinguished by the difference in the nature of flattening in both cases. This also affects the evolution of the luminosity function.

We find that the reionization history has a strong effect on the luminosity function at the faint end. It is known that the bright end of the luminosity function is affected primarily by the star formation mode of a halo, and the overall bias, whereas its faint end is affected by the reionization history.

However, we also find, from Figure 3.5, that the effect of reionization history is much stronger in the case of overdense regions. This is because of the enhanced photoionisation feedback, which is more sensitive to changes in reionization history. This order of magnitude change in the overdense region luminosity function should be visible to the James Webb Space Telescope, which can observe up to $m_{\text{AB}} \approx 31.5$ ($M_{\text{AB}} \approx -16.0$ at redshifts of interest; Windhorst et al. 2006).

3.3 Discussion and Summary

In this chapter, we used a semi-analytic model, based on Choudhury & Ferrara [2005, 2006] to study reionization and thermal history of an overdense region. Studying such regions is important because observations of galaxy luminosity function at high redshifts typically focus fields of view of limited sizes preferentially containing bright sources; these regions possibly are overdense and hence biased with respect to the globally averaged regions. In particular, we study the effect of radiative feedback arising from reionization on the shape of galaxy luminosity function.

In summary, we find that

1. Reionization proceeds differently in overdense regions. Overdense regions are ionised earlier because of enhanced number of sources and star formation. In addition, these regions have higher temperatures because of enhanced recombinations and hence effect of radiative feedback is enhanced too.
2. In particular, the shape of the galaxy luminosity function for biased regions is very different from that for average regions. There is a significant enhancement in the number of high-mass galaxies because of bias, while there is a reduction in low-mass galaxies resulting from enhanced radiative feedback.
3. Luminosity function in overdense regions is more sensitive to reionization history compared to average regions. The effect of radiative feedback shows up at absolute AB magnitudes $M_{\text{AB}} \gtrsim -17$ in these regions, while it occurs at much fainter magnitudes $M_{\text{AB}} \sim -12$ for average regions. This order of magnitude change in the overdense region luminosity function should be visible to the James Webb Space Telescope in future.

Finally, we critically examine some of the simplifying assumptions made here and how they are likely to affect our conclusions. Firstly, we have seen that the presence

of a high mass galaxy within a region of size R does not uniquely specify the value of the overdensity δ . Rather we obtain a probability density (which is Gaussian in shape) and work with the value where this probability is maximum. In reality, however, the actual value of δ could be different and this may possibly affect the predicted luminosity function. Note that the luminosity function at the brighter end is almost independent of the details of reionization history, and this, in principle, can be used for constraining the value of δ . The effect of feedback can then be studied using the faint end of the luminosity function.

The radiative feedback prescription we use is based on a Jeans mass calculation [Choudhury & Ferrara, 2005]. However, alternate prescriptions for feedback exist in literature, e.g., Gnedin [2000] and hence the shape of the luminosity function at faint ends as predicted by our model may not be robust. Interestingly, the presence of a “knee” in the luminosity function can be used to estimate the value of the halo mass below which star formation can be suppressed (which in turn can indicate the temperature) while the shape of the function below this knee should indicate the nature of feedback. This study can also be complemented with proposed for studying feedback using other observations, e.g., 21 cm observation [Schneider et al., 2008] and CMBR [Burigana et al., 2008].

Appendix **A**

Formation rate and survival probability of haloes in overdense regions

As expressed in Equation (3.5), the number density of ionizing photons produced per unit time is related to the SFR density, which in turn depends on the SFR in each halo, given by Equation (3.4), and the number density of haloes of a certain age, given by Equation (3.1) for average regions, and by Equation (3.15) for overdense regions. We derive Equation (3.15) in this appendix.

We denote the number density at redshift z of haloes formed between redshifts z_c and $z_c + dz_c$, with mass between M and $M + dM$, by $N(M, z, z_c)dMdz_c$. This quantity is related to (1) the formation rate at redshift z_c of haloes with mass between M and $M + dM$, denoted by $\dot{N}_{\text{form}}(M, z_c)dM$, and (2) the probability of their survival at redshift z , denoted by $p_{\text{surv}}(z, z_c)$. We calculate these two quantities using a technique given by Sasaki [1994], applied to an overdense region with overdensity δ and size R .

Recall that in extended Press-Schechter theory [Bond et al., 1991], the mass function of dark matter haloes is defined as the comoving number density of haloes with mass between M and $M + dM$. At redshift z , this quantity is given by

$$N(M, z)dM = \sqrt{\frac{2}{\pi}} \frac{\bar{\rho}_m}{M} \exp\left(\frac{-\nu^2}{2}\right) \frac{d\nu}{dM} dM, \quad (\text{A.1})$$

where $\bar{\rho}_m$ is the average matter density, and, as before, $\nu(M, z) \equiv \delta_c/[D(z)\sigma(M)]$. The critical overdensity of collapse of a halo is denoted by δ_c , $D(z)$ is the growth function of density perturbations, and $\sigma(M)$ is the rms value of density perturbations at the comoving scale corresponding to mass M . In a region with overdensity δ and linear size R , the mass function is enhanced. This enhancement can be calculated using the excursion set formalism [Bond et al., 1991]. The resulting mass function is again given by Equation

(A.1), except that now the quantity $\nu(M, z)$ is defined as

$$\nu(M, z) \equiv \frac{\delta_c/D(z) - \delta}{\sqrt{\sigma^2(M) - \sigma_R^2}}, \quad (\text{A.2})$$

where σ_R is the rms value of density perturbations at comoving scale R . Closely following Sasaki [1994], we can write

$$\dot{N}(M, z) = \dot{N}_{\text{form}}(M, z) - \dot{N}_{\text{dest}}(M, z), \quad (\text{A.3})$$

where $\dot{N}_{\text{dest}}(M, z)dM$ is the destruction rate at redshift z of haloes of mass between M and dM . (The halo formation rate is defined as the number density of haloes formed per unit time from mergers of lower mass haloes. Similarly the halo destruction rate is defined as the number density of haloes destroyed per unit time due to mergers with other haloes.) Here, an overdot denotes the time derivative. We can write the destruction rate as

$$\dot{N}_{\text{dest}}(M, z) = \int_M^\infty N(M, z)\tilde{Q}(M, M', z)dM', \quad (\text{A.4})$$

$$\equiv \phi(M, z)N(M, z), \quad (\text{A.5})$$

and the formation rate as

$$\dot{N}_{\text{form}}(M, z) = \int_{M_{\text{min}}}^M N(M', z)Q(M', M, z)dM', \quad (\text{A.6})$$

where $\tilde{Q}(M, M', z)$ is the probability that a halo of mass M merges with another halo to result in a halo of mass M' per unit time, and $Q(M', M, z)$ that an halo of mass M forming at redshift z has a progenitor of mass M' . The threshold mass M_{min} is introduced at this stage to avoid divergence. This gives

$$\dot{N}_{\text{form}}(M, z) = \dot{N}(M, z) + \phi(M, z)N(M, z). \quad (\text{A.7})$$

We now assume that ϕ has no characteristic mass scale so that $\phi(M, z) = M^\alpha \tilde{\phi}(z)$. This gives

$$\tilde{\phi}(z) = \frac{-\dot{N}(M, z) + \dot{N}_{\text{form}}(M, z)}{N(M, z)M^\alpha}. \quad (\text{A.8})$$

But since the left hand side of Equation (A.8) is a function of time alone (through the redshift), the right hand side of this equation also has to be independent of mass. In particular, we can then set $M = M_{\text{min}}$ in this equation, giving us

$$\tilde{\phi}(z) = \frac{-\dot{N}(M_{\text{min}}, z)}{N(M_{\text{min}}, z)M_{\text{min}}^\alpha}. \quad (\text{A.9})$$

Now, in the case of the overdense region that we are considering here, we have

$$\dot{N}(M, z) = N(M, z) \frac{\dot{D}(z)}{D^2(z)} \frac{\delta_c}{\delta_c/D(z) - \delta} [\nu^2(m, z) - 1], \quad (\text{A.10})$$

which gives

$$\tilde{\phi} = \frac{\dot{D}}{D^2} \frac{\delta_c}{\delta_c/D(z) - \delta} [\nu^2(M_{\min}, z) - 1] M_{\min}^{-\alpha}. \quad (\text{A.11})$$

Since our choice of threshold mass M_{\min} is arbitrary, we now need to take the limit $M_{\min} \rightarrow 0$. However, since $\nu \rightarrow 0$ in this limit, $\tilde{\phi}$ becomes indeterminate, except when $\alpha = 0$. This implies that we must set $\alpha = 0$ for consistency. This gives $\phi(M, z) = \tilde{\phi}(z)$. Substituting the resultant expression in Equation (A.7), we get

$$\dot{N}_{\text{form}}(M, z) = N(M, z) \frac{\dot{D}}{D^2} \frac{\delta_c}{\delta_c/D(z) - \delta} \nu^2(M, z). \quad (\text{A.12})$$

This is the required formation rate of haloes in an overdense region.

Furthermore, from our definitions of probabilities in Equations (A.5) and (A.6), we can write the probability that a halo that has formed at redshift z_c continues to exist at redshift z as

$$p_{\text{surv}}(z, z_c) = \exp \left[- \int_{t(z_c)}^{t(z)} \phi(t') dt' \right], \quad (\text{A.13})$$

which in our case results in

$$p_{\text{surv}}(z, z_c) = \frac{\delta_c/D(z) - \delta}{\delta_c/D(z_c) - \delta}. \quad (\text{A.14})$$

From Equations (A.12) and (A.14), we can now write the comoving number density $N(M, z, z_c) dM dz_c$ at redshift z of collapsed halos having mass in the range M and $M + dM$ and redshift of collapse in the range z_c and $z_c + dz_c$ as

$$\begin{aligned} N(M, z, z_c) dM dz_c &= N(M, z_c) \left(\frac{\nu^2 \delta_c}{\delta_c/D(z_c) - \delta} \right) \frac{\dot{D}(z_c)}{D^2(z_c)} \\ &\times p_{\text{surv}}(z, z_c) \frac{dt}{dz_c} dz_c dM, \end{aligned} \quad (\text{A.15})$$

This is our Equation (3.15).

It is worth pointing out that Equations (A.12) and (A.14) reduce to the average forms for halo formation rate and survival probability in the limit $\delta \rightarrow 0$ and $R \rightarrow \infty$.

Formation rate of Dark Matter Haloes

A crucial ingredient of the analytical models of previous chapters was the global star formation rate history $\dot{\rho}_*(z)$. As we saw, the formation rate of dark matter halos was required to calculate this quantity. In fact, the hierarchical formation of dark matter halos is the key driver that leads to formation and evolution of galaxies and clusters of galaxies. Dark matter responds mainly to gravitational forces, and by virtue of a larger density than baryonic matter, assembly of matter into halos and large scale structure is primarily driven by gravitational instability of initial perturbations in dark matter. Galaxies are believed to form when gas in highly over-dense halos cools and collapses to form stars in significant numbers [Hoyle, 1953; Rees & Ostriker, 1977; Silk, 1977; Binney, 1977].

The halo mass function describes the comoving number density of dark matter halos as a function of mass and redshift in a given cosmology. It is possible to develop the theory of mass functions in a manner that makes no reference to the details of the cosmological model or the power spectrum of fluctuations. That is, we expect the mass function to take a universal form, when scaled appropriately. Simple theoretical arguments have been used to obtain this universal functional form of the mass function [Press & Schechter, 1974; Bond et al., 1991; Sheth, Mo & Tormen, 2001]. (Bond et al. [1991], and, Sheth, Mo & Tormen [2001] used the excursion set theory to derive the mass function.) Much work has also been done to determine the extent to which this form is consistent with results from N-body simulations [Jenkins et al., 2001; White, 2002; Reed et al., 2003; Warren et al., 2006; Reed et al., 2007; Lukić et al., 2007; Cohn & White, 2008; Tinker et al., 2008] with the conclusion that the agreement is fairly good. It is remarkable that a purely local approach provides a fairly accurate description of the manifestly non-linear and strongly coupled process of gravitational clustering. The success of the local description has been exploited in developing the semi-analytic theories of galaxy formation [White & Frenk, 1991; Kauffmann, White & Guiderdoni, 1993; Chiu & Ostriker, 2000; Madau, Ferrara & Rees, 2001; Samui, Srianand & Subrama-

nian, 2007].

The Press-Schechter mass function [Press & Schechter, 1974] that is commonly used in these semi-analytic models assumes spherical collapse of halos [Gunn & Gott, 1972]. The shape of this mass function agrees with numerical results qualitatively, but there are deviations at a quantitative level [Efstathiou et al., 1988; Jenkins et al., 2001]. Improvements to the Press-Schechter mass function have been made to overcome this limitation. In particular, the Sheth-Tormen mass function, which is based on the more realistic ellipsoidal collapse model [Sheth & Tormen, 1999; Sheth, Mo & Tormen, 2001] fits numerical results better. Many fitting functions with three or four fitting parameters have been proposed, these are based on results of simulations of the Lambda-Cold Dark Matter (Λ CDM) model [Jenkins et al., 2001; Reed et al., 2003; Warren et al., 2006; Fakhouri, Ma & Boylan-Kolchin, 2010].

In the application of the theory of mass functions to the semi-analytic models for galaxy formation, we often need to know comoving number density of halos of a certain age. Naturally, this quantity is related to the halo formation rates and the survival probability. While these details are known and well understood for the Press-Schechter mass function [Press & Schechter, 1974], the situation is not as clear for other models of the mass function. Furthermore, analytic estimates for the halo formation rate and survival probability are important in spite of the availability of accurate fitting functions for these quantities in the Λ CDM model. This is because analytic estimates can be used to study variation in these quantities with respect to, for instance, the underlying cosmology or the power spectrum of matter perturbations. Studying such variation with the help of simulations is often impractical. In this chapter, we focus on the computation of halo formation rates.

Several approaches to calculating halo formation rates have been suggested [Blain & Longair, 1993; Sasaki, 1994; Kitayama & Suto, 1996]. In particular, Sasaki [1994] suggested a very simple approximation for the formation rate as well as survival probability for halos. The approximation was suggested for the Press-Schechter mass function, though it does not use any specific aspect of the form of mass function. The series of arguments is as follows:

- Merger and accretion lead to an increase in mass of individual halos. Formation of halos of a given mass from lower mass halos leads to an increase in the number density, whereas destruction refers to halos moving to a higher mass range. The net change in number density of halos in a given interval in mass is given by the difference between the formation and destruction rate.
- Given the net rate of change, we can find the formation rate if we know the destruction rate.
- A simple but viable expression for the destruction rate is obtained by assuming

that the probability of destruction per unit mass (also known as the halo destruction efficiency) is independent of mass.

- This approximate expression for the destruction rate is then used to derive the formation rate as well as the survival probability.

The resulting formulae have been applied freely to various cosmologies and power spectra, including the CDM class of power spectra. The Sasaki approach has been used in many semi-analytic models of galaxy formation [Chiu & Ostriker, 2000; Choudhury & Ferrara, 2005; Samui, Srianand & Subramanian, 2007] mainly due to its simplicity. Attempts have also been made to generalize the approximation to models of mass function other than the Press-Schechter mass function [Samui, Subramanian & Srianand, 2009], though it has been found that a simple extension of the approximation sometimes leads to unphysical results. In particular, when applied to the Sheth-Tormen mass function, the Sasaki approach yields negative halo formation rates.

In this chapter, we investigate the application of the Sasaki approach to the Sheth-Tormen mass function. We test the Sasaki approach by explicitly computing the halo formation and destruction rates for the Press-Schechter mass function using the excursion set formalism. We then generalize this same method to compute the halo formation rates for the Sheth-Tormen mass function. We find that halo formation rates computed in this manner are always positive. Finally, we compare our analytical results with N-body simulations.

A reason for choosing the approach presented in this chapter, as compared to other competing approaches based on the excursion set formalism, is that we wish to be able to differentiate between major and minor mergers. This is an essential requirement in semi-analytical models of galaxy formation and is not addressed by other approaches for halo formation rate [Percival & Miller, 1999; Percival, Miller & Peacock, 2000; Percival, 2001; Giocoli et al., 2007; Moreno, Giocoli & Sheth, 2008, 2009].

Many attempts at calculating halo formation and merger rates using analytical or numerical techniques have been made in the literature. Neistein & Dekel [2008] and Neistein, Macciò & Dekel [2010] provided analytical expressions for merger rates of dark matter haloes within the framework of extended Press-Schechter formalism. They showed that these merger rates can fit to the results of N-body simulation better than that from the estimates of Lacey & Cole [1994] and hence this can be a useful tool for studying the galaxy formation. Miller et al. [2006] investigated the cosmological growth and merger rates of dark matter haloes for studying the accretion history of supermassive black holes. Similarly, Li et al. [2007b] studied the growth history and formation rates of dark matter haloes numerically using the Lagrangian perturbation code PINOCCHIO and showed that the mass assembly histories of haloes obtained by PINOCCHIO are in good agreement with those obtained using N-body simulations.

Using high-resolution N-body simulations, Cohn & White [2008] investigated the abundance, clustering and mass assembly histories of high-mass haloes at high red-

shift. They also compared the mergers and mass accretion histories to the extended Press-Schechter formalism. Benson, Kamionkowski & Hassani [2005] described a new approach, which is different from the extended Press-Schechter theory, to find more physically reasonable estimates of dark matter halo merger rates. They showed that while the extended PS theory contains an intrinsic inconsistency in its predictions for halo merger rate, their approach can produce always a self-consistent merger rate.

In this chapter, we first discuss the Sasaki and the excursion set formalisms in Section 4.1. We then compare these methods with N-body simulations by first describing our simulations in Section 4.2 and then discussing our findings in Section 4.3.

4.1 Rate of Halo Formation

The total change in number density of collapsed halos at time t with mass between M and $M + dM$ per unit time is denoted by $\dot{N}(M, t)dM$ and is due to halos gaining mass through accretion or mergers. Lower-mass halos gain mass so that their mass is now between M and $M + dM$, and some of the halos with mass originally between M and $M + dM$ gain mass so that their mass now becomes higher than this range. We call the former process halo formation and the latter as halo destruction, even though the underlying physical process is the same in both cases; the different labels of formation or destruction arise due to our perspective from a particular range of mass. We denote the rate of halo formation by $\dot{N}_{\text{form}}(M, t)dM$ and the rate of halo destruction by $\dot{N}_{\text{dest}}(M, t)dM$. We immediately have

$$\dot{N}(M, t) = \dot{N}_{\text{form}}(M, t) - \dot{N}_{\text{dest}}(M, t). \quad (4.1)$$

Following Sasaki [1994], in general we can formulate each term in the above expression as follows. The rate of halo destruction can be written as

$$\dot{N}_{\text{dest}}(M, t) = \int_M^{\infty} N(M, t) \tilde{Q}(M, M'; t) dM' \quad (4.2)$$

$$\equiv \phi(M, t) N(M, t), \quad (4.3)$$

where, $\tilde{Q}(M, M'; t)$ represents the probability of a halo of mass M merging with another halo to form a new halo of mass M' per unit time. The fraction of halos that are destroyed per unit time is denoted by $\phi(M, t)$. This quantity is also referred to as the efficiency of halo destruction. The rate of halo formation can be written as

$$\dot{N}_{\text{form}}(M, t) = \int_0^M N(M', t) Q(M', M; t) dM' \quad (4.4)$$

where $Q(M', M; t)$ represents the probability of a halo of mass M' evolving into another halo of mass M per unit time. We can now write, from Equation (4.1) and from our definitions in Equations (4.3) and (4.4),

$$\dot{N}_{\text{form}}(M, t) = \dot{N}(M, t) + \phi(M, t)N(M, t). \quad (4.5)$$

This reduces the calculation of rate of halo formation to a computation of $\phi(M, t)$.

Sasaki [1994] proposed a simple ansatz to compute $\phi(M, t)$: If we assume that the efficiency of halo destruction has no characteristic mass scale and we require that the destruction rate remains finite at all masses then it can be shown that ϕ does not depend on mass at all. (We have proved this claim in Appendix A of Chapter 3.)

4.1.1 Sasaki prescription: Press-Schechter mass function

To understand the Press-Schechter formalism [Press & Schechter, 1974; Bond et al., 1991], which gives the co-moving number density of collapsed halos at a time t with mass between M and $M + dM$, let us consider a dark matter inhomogeneity centered around some point in the universe. The smoothed density contrast within a smoothing scale of radius R around this point is defined as $\delta(R) = [\rho(R) - \bar{\rho}]/\bar{\rho}$, where $\rho(R)$ is the density of dark matter within R and $\bar{\rho}$ is the mean background density of the universe. If this density contrast $\delta(R)$ is greater than the threshold density contrast for collapse δ_c obtained from spherical collapse model [Gunn & Gott, 1972], the matter enclosed within the volume collapses to form a bound object. In hierarchical models, density fluctuations are larger at small scales so with decreasing R , $\delta(R)$ will eventually reach δ_c . The problem then is to compute the probability that the first up-crossing of the barrier at δ_c occurs on a scale R . This problem can be addressed by excursion set approach.

The excursion set approach consists the following principles: Consider a trajectory $\delta(R)$ as a function of the filtering radius R at any given point and then determine the largest R at which $\delta(R)$ up-crosses the density threshold $\delta_c(t)$ corresponding to the formation time t . The solution of the problem can be enormously simplified for Brownian trajectories [Chandrasekhar, 1943b], that is for sharp k -space filtered density fields, as in this case contribution of each wave mode is independent of all others. In such a case we have to solve the Fokker-Planck equation for the probability density $\Pi(\delta, S)d\delta$, where $S \equiv \sigma^2(R)$ and $\sigma(R)$ is the standard deviation of fluctuations in the initial density field, smoothed at a scale R ,

$$\frac{\partial \Pi(\delta, S)}{\partial S} = \frac{1}{2} \frac{\partial^2 \Pi(\delta, S)}{\partial \delta^2} \quad (4.6)$$

The solution [Porciani et al., 1998; Zentner, 2007] can be obtained using the absorbing boundary condition $\Pi(\delta_c(t), S) = 0$ and the initial condition $\Pi(\delta, S = 0) = \delta_D(\delta)$,

where $\delta_D(\delta)$ is the Dirac delta function

$$\begin{aligned} \Pi(\delta, S; \delta_c) d\delta &= \frac{1}{\sqrt{2\pi S}} \\ &\times \left[\exp\left(-\frac{\delta^2}{2S}\right) - \exp\left(-\frac{(\delta - 2\delta_c(t))^2}{2S}\right) \right] d\delta. \end{aligned} \quad (4.7)$$

Now define $F(S, \delta_c(t)) = \int_{-\infty}^{\delta_c(t)} d\delta \Pi(\delta, S; \delta_c(t))$ as the survival probability of trajectories and obtain the differential probability for a first barrier crossing:

$$f(S) = -\frac{\partial F(S, \delta_c(t))}{\partial S} = \frac{\delta_c(t)}{\sqrt{2\pi S^3}} \exp\left(-\frac{\delta_c(t)^2}{2S}\right) \quad (4.8)$$

From this, one can obtain the comoving number density of collapsed halos at time t with mass between M and $M + dM$

$$\begin{aligned} N_{\text{PS}}(M, t) dM &= \frac{\rho_{\text{nr}}}{M} f(S) \left| \frac{dS}{dM} \right| dM \\ &= \sqrt{\frac{2}{\pi}} \frac{\rho_{\text{nr}}}{M} (\nu)^{\frac{1}{2}} \left| \frac{d \ln \sigma}{dM} \right| \exp\left[-\frac{\nu}{2}\right] dM. \end{aligned} \quad (4.9)$$

here ρ_{nr} is the comoving density of non-relativistic matter and $\nu \equiv [\delta_c(t)/\sigma(M)]^2 \equiv [\delta_c/(D(t)\sigma(M))]^2$, where δ_c is the threshold density contrast for collapse, $D(t)$ is the linear rate of growth for density perturbations and $\sigma(M) (\equiv S^{1/2})$ is the standard deviation of fluctuations in the initial density field, which is smoothed over a scale that encloses mass M .

In the following discussion, we will denote the mass function by $N(M, t)$ if the statement is independent of the specific form of the mass function. We will use a subscript PS when the statements apply only to the Press-Schechter form of the mass function.

With Sasaki's ansatz, the destruction rate efficiency ϕ can be written in this case as

$$\phi(t) = \frac{1}{D(t)} \frac{dD(t)}{dt}. \quad (4.10)$$

With this, we can write down the rate of halo formation for the Press-Schechter mass function from Equation (4.5) as:

$$\begin{aligned} \dot{N}_{\text{form}}(M, t) &= \dot{N}_{\text{PS}}(M, t) + \frac{1}{D(t)} \frac{dD(t)}{dt} N_{\text{PS}}(M, t) \\ &= \frac{1}{D(t)} \frac{dD(t)}{dt} N_{\text{PS}}(M, t) \frac{\delta_c^2}{\sigma^2(M) D^2(t)}. \end{aligned} \quad (4.11)$$

Note that for halos with large mass, that is in the limit $\delta_c \gg \sigma(M)D(t)$, \dot{N}_{form} approaches \dot{N}_{PS} . In other words, the total change in the number of halos is determined by formation of new halos. For halos with low mass, where σ is much larger than unity, although \dot{N}_{form} remains positive, the total change is dominated by destruction and \dot{N}_{PS} becomes negative.

We can also define two related, useful quantities now. Firstly, the probability $p(t_1, t_2)$ that a halo which exists at t_1 continues to exist at t_2 without merging is given by

$$p(t_1, t_2) = \exp \left[- \int_{t_1}^{t_2} \phi(t') dt' \right] = \frac{D(t_1)}{D(t_2)} \quad (\text{where } t_2 > t_1) \quad (4.12)$$

This is usually known as the survival probability of halos, and is independent of halo mass in the Sasaki prescription. Secondly, in this picture the distribution of epochs t_f of formation of halos with mass M at time t can be defined as

$$F(M; t_f, t) dM dt_f = \dot{N}_{\text{form}}(M, t_f) p(t_f, t) dM dt_f. \quad (4.13)$$

4.1.2 Sasaki prescription: Sheth-Tormen mass function

The Press-Schechter mass function does not provide a very good fit to halo mass function obtained in N-body simulations. In particular, it under-predicts the number density of large mass halos, and over-predicts that of small mass halos. Hence it is important to generalize the calculation of formation rates to other models for mass function that are known to fit simulations better. The Sheth-Tormen form of mass function [Sheth & Tormen, 1999] is known to fit simulations much better than the Press-Schechter form.¹ (For a comparison of both of these forms of halo mass function with simulations, see Figure 3 of Jenkins et al. 2001.) The Sheth-Tormen mass function is given by

$$N_{\text{ST}}(M, t) dM = A \sqrt{\frac{2}{\pi}} \frac{\rho_{\text{nr}}}{M} (a\nu)^{1/2} \left| \frac{d \ln \sigma}{dM} \right| \times [1 + (a\nu)^{-p}] \exp \left[-\frac{a\nu}{2} \right] dM, \quad (4.14)$$

where the parameters a , p , and A have best fit values of $a = 0.707$, $p = 0.3$ and $A = 0.322$ [Sheth & Tormen, 1999], and the quantity ν is as defined before. This form

¹Even this form of halo mass function has poor accuracy in some cases, namely, for conditional mass functions with large mass ratios and for mass function in overdense regions [Sheth & Tormen, 2002]. In applications involving these regimes it is perhaps advisable to use more accurate fitting functions to simulation data. However, the Sheth-Tormen form still has the property of being considerably better than the Press-Schechter form while having a physical interpretation. It is thus preferable in many semi-analytic models where the Press-Schechter form is used.

of mass function has the added advantage of being similar to the mass function derived using a variable barrier motivated by ellipsoidal collapse of overdense regions [Sheth, Mo & Tormen, 2001; Sheth & Tormen, 2002]. Note that if we choose $A = 0.5$, $p = 0$ and $a = 1$ then we recover the Press-Schechter mass function derived using spherical collapse. Recently, it has been shown that the best fit values of these parameters depend on the slope of the power spectrum [Bagla, Khandai & Kulkarni, 2009].

We can now apply the Sasaki prescription to this form of mass function and calculate the rates of halo formation and destruction [Ripamonti, 2007]. We get for the destruction rate efficiency

$$\phi(t) = \frac{1}{D} \frac{dD}{dt} [1 - 2p]. \quad (4.15)$$

Note that the destruction rate efficiency is independent of mass. The rate of halo formation is then given by

$$\dot{N}_{\text{form}}^{\text{ST}}(M, t) = -\frac{1}{D} \frac{dD}{dt} \left[\frac{2p}{1 + (a\nu)^{-p}} - a\nu \right] N_{\text{ST}}(M, t). \quad (4.16)$$

Note that in this case, because of the extra term, the halo formation rate can be negative for some values of halo mass. Since negative values of rate of halo formation are unphysical, this indicates that the generalization of Sasaki approximation to the Sheth-Tormen mass function is incorrect. The same problem is encountered if we use other models of the halo mass function [Samui, Subramanian & Srianand, 2009].

However, since the basic framework outlined in the beginning of this section is clearly correct, there should not be any problems in generalizing it to other mass functions. It is therefore likely that the simplifying assumptions of the Sasaki method that led to the estimate of the halo destruction rate efficiency of Equation (4.15) are responsible for negative halo formation rate.

4.1.3 Excursion set approach to halo formation rates: Press-Schechter mass function

To check this assertion we perform an explicit calculation of the rate of halo formation using the excursion set formalism. Recall that from Equations (4.2) and (4.3), we can write for the halo destruction rate efficiency as

$$\phi(M_1, t) = \int_{M_1}^{\infty} \tilde{Q}(M_1, M_2; t) dM_2, \quad (4.17)$$

where $\tilde{Q}(M_1, M_2; t)$ represents the probability that an object of mass M_1 grows into an object of mass M_2 per unit time through merger or accretion at time t . This quantity is also known as the transition rate.

In the excursion set formalism, the conditional probability for a halo of mass M_1 present at time t_1 to merge with another halo to form a larger halo of mass between M_2 and $M_2 + dM_2$ at time $t_2 > t_1$ [Lacey & Cole, 1993, 1994] can be written for the extended Press-Schechter mass function as

$$f(M_2, \delta_2 | M_1, \delta_1) dM_2 = \sqrt{\frac{2}{\pi}} \frac{\delta_2(\delta_1 - \delta_2)}{\delta_1} \sigma_2^2 \left[\frac{\sigma_1^2}{\sigma_2^2(\sigma_1^2 - \sigma_2^2)} \right]^{\frac{3}{2}} \times \exp \left[-\frac{(\delta_2\sigma_1^2 - \delta_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2 - \sigma_2^2)} \right] \left| \frac{d\sigma_2}{dM_2} \right| dM_2. \quad (4.18)$$

Here, σ_1 and σ_2 are values of the standard deviation of the density perturbations when smoothed over scales that contain masses M_1 and M_2 respectively, and δ_1 and δ_2 are the values of the threshold density contrast for spherical collapse at time t_1 and t_2 respectively. Taking the limit t_2 tends to t_1 , i. e. δ_2 tends to δ_1 , we can determine the mean transition rate at time $t = t_1$:

$$\tilde{Q}(M_1, M_2; t) dM_2 = \sqrt{\frac{2}{\pi}} \sigma_2^2 \left[\frac{\sigma_1^2}{\sigma_2^2(\sigma_1^2 - \sigma_2^2)} \right]^{\frac{3}{2}} \left| \frac{d\delta}{dt} \right| \times \exp \left[-\frac{\delta^2(\sigma_1^2 - \sigma_2^2)}{2\sigma_1^2\sigma_2^2} \right] \left| \frac{d\sigma_2}{dM_2} \right| dM_2. \quad (4.19)$$

This represents the probability that a halo of mass M_1 will accrete or merge to form another halo of mass M_2 at time t . We can use this with Equation (4.17) to explicitly compute the destruction rate, and hence the halo formation rate.

However, in the excursion set method, an arbitrarily small change in the halo mass is treated as creation of a new halo. As a result, the integral in Equation (4.17) diverges unless we specify a ‘‘tolerance’’ parameter. We assume that a halo is assumed to have *survived* unless its mass increases such that $M_1 \rightarrow M_2 \geq M_1(1 + \epsilon)$ due to either accretion or merging, where ϵ is a small number. This assumption allows us to introduce a lower cutoff in the integral in Equation (4.17) and the lower limit changes to $M_1(1 + \epsilon)$, leading to a convergent integral. This is also physically pertinent for our application as infinitesimal changes do not lead to variations in dynamical structure of halos, and hence we do not expect any changes in galaxies hosted in halos that do not undergo a major merger. This is similar in spirit to the assumption made elsewhere in the literature that a halo is assumed to survive until its mass increases by a factor of two [Lacey & Cole, 1994; Kitayama & Suto, 1996]. Note that N-Body simulations have a natural cutoff due to the discrete nature of N-Body particles. With the introduction of this new parameter, the modified formula for the halo destruction rate efficiency is given by

$$\phi(M_1, t) = \int_{M_1(1+\epsilon)}^{\infty} \tilde{Q}(M_1, M_2; t) dM_2 \quad (4.20)$$

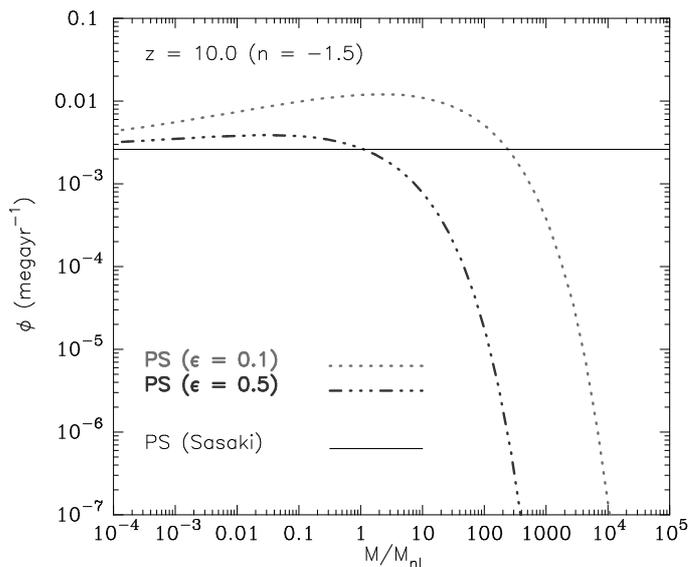


Figure 4.1: Destruction rate $\phi(M, t)$ at $z = 10$ for the Press-Schechter mass function for a power law model with index -1.5 . Curves have been plotted for $\epsilon = 0.1$ and $\epsilon = 0.5$.

This can then be used to calculate the rate of halo formation using Equation (4.5).

Figure 4.1 shows the destruction rate efficiency $\phi(M, t)$ computed in this manner for the Press-Schechter mass function for an Einstein-de Sitter cosmology with power law spectrum of density perturbations with index -1.5 . Curves have been plotted for $\epsilon = 0.1$ and $\epsilon = 0.5$. We have also shown the Sasaki approximation in the same panel. The excursion set result has three features:

1. At small M , the excursion set value approaches the destruction rate computed using the Sasaki approximation.
2. The destruction rate has a peak, more pronounced for smaller ϵ , near the scale of non-linearity.
3. At larger scales the destruction rate falls rapidly; this is the region where deviations from the Sasaki result are the largest. Thus the halo destruction rate efficiency vanishes at large masses.

A similar trend is seen for other power spectra. We postpone a detailed discussion of these issues to the end of this section.

4.1.4 Excursion set approach to halo formation rates: Sheth-Tormen mass function

As discussed in Subsection 4.1.2, the Sheth-Tormen mass function is known to be a much better fit to N-Body simulations than the Press-Schechter mass function. Several other forms of halo mass function have also been fitted to results of high resolution N-Body simulations [Jenkins et al., 2001; Reed et al., 2003; Warren et al., 2006]. In this chapter we focus on the Sheth-Tormen mass function. Recall that the Sasaki prescription gives unphysical results when applied to this form of the mass function. Therefore, we now derive the halo destruction rate efficiency, and the halo formation rates for the Sheth-Tormen mass function. This requires obtaining analogs of Equations (4.18) and (4.19).

Sheth, Mo & Tormen [2001] showed that once the barrier shape is known, all the predictions of the excursion set approach, like the conditional mass function, associated with that barrier can be computed easily.² Further, they showed that the barrier shape for ellipsoidal collapse is

$$B(\sigma, t) \equiv \delta_{\text{ec}}(\sigma, t) = \sqrt{a}\delta_c(t) [1 + \beta(a\nu)^{-\gamma}], \quad (4.21)$$

where $a = 0.75$, $\beta = 0.485$, $\gamma = 0.615$, and, $\delta_c(t)$ is the threshold value of overdensity required for spherical collapse (also see Sheth & Tormen 2002). They also found that, for various barrier shapes $B(S)$, the first-crossing distribution of the excursion set theory is well approximated by

$$f(S)dS = \frac{|T(S)|}{\sqrt{2\pi S^{3/2}}} \exp\left[-\frac{B(S)^2}{2S}\right] dS, \quad (4.22)$$

where $T(S)$ denotes the sum of the first few terms in the Taylor series expansion of $B(S)$

$$T(S) = \sum_{n=0}^{\infty} \frac{(-S)^n}{n!} \frac{\partial^n B(S)}{\partial S^n}. \quad (4.23)$$

(Here, for conformity with the literature, we use the symbol $S \equiv \sigma^2$.) This expression gives the exact answer in the case of constant and linear barriers. For the ellipsoidal barrier, we can get convergence of the numerical result if we retain terms in the Taylor expansion up to $n = 5$.

For Press-Schechter mass function, the conditional mass function $f(S_1, \delta_1 | S_2, \delta_2)$ can be obtained from the first crossing distribution $f(S)$ by just changing the variables $\delta \rightarrow \delta_1 - \delta_2$ and $S \rightarrow S_1 - S_2$. This can be done because, despite the shift in the origin, the second barrier is still one of constant height. This is no longer true for

²These can be also calculated for non-Gaussian initial conditions, see, e.g., De Simone, Maggiore & Riotto [2011]

ellipsoidal collapse and hence we cannot simply rescale the function in Equation 4.22 to get the conditional mass function. Instead, this can be done by making the replacements $B(S) \rightarrow B(S_1) - B(S_2)$ and $S \rightarrow S_1 - S_2$ to get

$$f(S_1|S_2)dS_1 = \frac{|T(S_1|S_2)|}{\sqrt{2\pi}(S_1 - S_2)^{3/2}} \exp \left[-\frac{(B(S_1) - B(S_2))^2}{2(S_1 - S_2)} \right] dS_1, \quad (4.24)$$

where we now have

$$T(S_1|S_2) = \sum_{n=0}^5 \frac{-(S_1 - S_2)^n}{n!} \frac{\partial^n (B(S_1) - B(S_2))}{\partial S_1^n}. \quad (4.25)$$

Using Bayes' theorem, we now have

$$\begin{aligned} f(S_2|S_1) dS_2 &= \frac{|T(S_1|S_2)||T(S_2)|}{|T(S_1)|} \frac{1}{\sqrt{2\pi}} \left[\frac{S_1}{S_2(S_1 - S_2)} \right] \\ &\times \exp \left[-\frac{[B(S_1) - B(S_2)]^2}{2(S_1 - S_2)} - \frac{B^2(S_2)}{2S_2} + \frac{B^2(S_1)}{2S_1} \right] dS_2. \end{aligned} \quad (4.26)$$

A change of variables from S to M now gives us an analog of Equation (4.18) for the Sheth-Tormen mass function. In other words, we get the conditional probability $f_{\text{ST}}(M_2|M_1)d \ln M_2$ that a halo of mass M_1 present at time t_1 will merge to form a halo of mass between M_2 and $M_2 + dM_2$ at time $t_2 > t_1$. Further, taking the limit as t_2 tends to $t_1 (= t)$, we obtain $\tilde{Q}(M_1, M_2; t)$. As before, we can then use it to calculate the halo destruction rate efficiency $\phi(M, t)$ and the rate of halo formation $\dot{N}_{\text{form}}^{\text{ST}}(M_1, z)$ using equations (4.5) and (4.20). We perform this part of the calculation numerically. It is also possible to use this formalism to calculate formation rates for the square-root barrier [Moreno, Giocoli & Sheth, 2009, 2008; Giocoli et al., 2007], which is a good approximation for the ellipsoidal collapse model. We do not attempt this calculation here.

Figure 4.2 is the analog of Figure 4.1 for the Sheth-Tormen mass function. It shows the destruction rate efficiency $\phi(M, t)$ computed using the excursion set method for an Einstein-de Sitter cosmology with power law spectrum of density perturbations with index -1.5 at $z = 10.0$. Curves have been plotted for $\epsilon = 0.1$ and $\epsilon = 0.5$. We have also shown the Sasaki approximation for ST mass function in the same panel for comparison. This result for the Sheth-Tormen mass function has the same features as the result for the Press-Schechter mass function. We also see that the destruction rate efficiency is far from constant at small M/M_{nl} . Thus the central assumption of the Sasaki prescription is invalid in the case of Sheth-Tormen mass function as well.

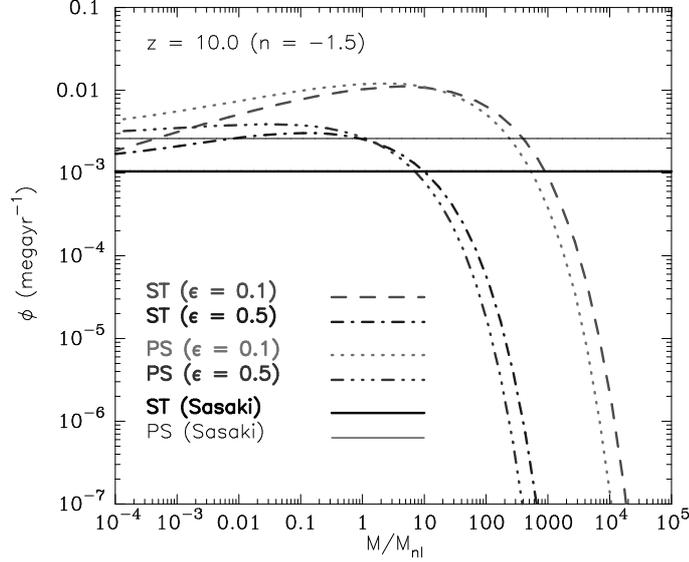


Figure 4.2: Same as Figure 4.1 but for the Sheth-Tormen (as well as Press-Schechter) mass function.

n	N_{box}	N_{part}	r_{nl}^i	r_{nl}^f	$r_{\text{nl}}^{\text{max}}$	z_i
-1.5	400^3	400^3	2.5	12.0	10.0	103.38
-0.5	256^3	256^3	2.5	12.0	18.2	291.53

Table 4.1: Simulations used in this chapter. Power law power spectrum: here n is the index of the power spectrum, N_{box} is the size of the simulation box, N_{part} represents the number of particles, r_{nl}^i is the scale of non-linearity at the earliest epoch, r_{nl}^f is the actual scale of non-linearity for the last epoch, $r_{\text{nl}}^{\text{max}}$ represents the maximum scale of non-linearity and z_i is the starting redshift of the simulations for every model.

L_{box}	N_{part}	m_{part}	ϵ	z_{f}	z_{out}
23.04	512^3	6.7×10^6	1.35	5.0	5.04
51.20	512^3	7×10^7	3.00	3.0	3.34
76.80	512^3	2.3×10^8	4.50	1.0	1.33

Table 4.2: Simulations used in this chapter. Λ CDM power spectrum: columns 1 and 2 list the size of the box (in $h^{-1}\text{Mpc}$) and the number of particles used in the simulations. Columns 3 and 4 give the mass (in $h^{-1}M_{\odot}$) and force resolution (in $h^{-1}\text{kpc}$; not to be confused with the ϵ used in the text) of the simulations, while columns 5 and 6 tell us the redshift at which the simulations were terminated and the redshift for which the analyses were done.

4.2 N-Body simulations

From the excursion set calculation described in the previous section, we thus find that the halo destruction rate efficiency is not independent of mass as is assumed in the Sasaki prescription. Clearly, this is the reason why Sasaki prescription yields unphysical values for the rate of halo formation. In this section and the next, we now compare the results of our excursion set calculation with results of N-body simulations.

We used the TreePM code [Khandai & Bagla, 2009] for these simulations. The TreePM [Bagla, 2002; Bagla & Ray, 2003] is a hybrid N-Body method which improves the accuracy and performance of the Barnes-Hut (BH) Tree method [Barnes & Hut, 1986] by combining it with the PM method [Miller, 1983; Klypin & Shandarin, 1983; Bouchet, Adam & Pellat, 1985; Bouchet & Kandrup, 1985; Hockney & Eastwood, 1988; Bagla & Padmanabhan, 1997; Merz, Pen & Trac, 2005]. The TreePM method explicitly breaks the potential into a short-range and a long-range component at a scale r_s —the PM method is used to calculate the long-range force and the short-range force is computed using the BH Tree method. Use of the BH Tree for short-range force calculation enhances the force resolution as compared to the PM method.

The mean inter-particle separation between particles in the simulations used here is $l_{\text{mean}} = 1.0$ in units of the grid-size used for the PM part of the force calculation. In our notation this is also cube root of the ratio of simulation volume N_{box}^3 to the total number of particles N_{part} .

Power law models do not have any intrinsic scale apart from the scale of non-linearity introduced by gravity. We can therefore identify an epoch in terms of the scale of non-linearity r_{nl} . This is defined as the scale for which the linearly extrapolated value of the mass variance at a given epoch $\sigma_L(a, r_{\text{nl}})$ is unity. All power law simulations are normalized such that $\sigma^2(a = 1.0, r_{\text{nl}} = 8.0) = 1.0$. The softening length in grid units is 0.03 in all runs.

The Λ CDM simulations were run with the set of cosmological parameters favored by the WMAP 5-yr data (Komatsu et al. 2009) as the best fit for the Λ CDM class of models: $\Omega_{nr} = 0.2565$, $\Omega_{\Lambda} = 0.7435$, $n_s = 0.963$, $\sigma_8 = 0.796$, $h = 0.719$ and $\Omega_b h^2 = 0.02273$. The simulations were done with 512^3 particles in a comoving cube of three different values of the physical volume as given in Table 4.1.4.

Simulations introduce an inner and an outer scale in the problem and in most cases we work with simulation results where $L_{\text{box}} \gg r_{\text{nl}} \geq L_{\text{grid}}$, where L_{grid} , the size of a grid cell, is the inner scale in the problem. L_{box} is the size of the simulation and represents the outer scale. In Table (4.1.4) we list the power law models simulated for the present study. We list the index of the power spectrum n (column 1), size of the simulation box N_{box} (column 2), number of particles N_{part} (column 3), the scale of non-linearity at the earliest epoch used in this study (column 4), and, the maximum scale of non-linearity, $r_{\text{nl}}^{\text{max}}$ (column 6) given our tolerance level of 3% error in the mass variance at this scale. For some models with very negative indices we have run the simulations beyond this epoch. This can be seen in column 5 where we list the actual scale of non-linearity for the last epoch. The counts of halos in low mass bins are relatively unaffected by finite box considerations. We therefore limit errors in the mass function by running the simulation up to $r_{\text{nl}}^{\text{max}}$. Column 7 lists the starting redshift of the simulations for every model. Similarly, in Table (4.1.4), we mention the details of the LCDM simulations used in this work. We list the size of the simulation box L_{box} in $h^{-1}\text{Mpc}$ (column 1), number of particles used in the simulations N_{part} (column 2), mass of the particles m_{part} in $h^{-1}\text{M}_{\odot}$ (column 3), force resolution ϵ (not to be confused with the ϵ used in the text) of the simulations in $h^{-1}\text{kpc}$ (column 4), the redshift z_f at which the simulations were terminated (column 5) and the redshift z_{out} for which the analyses were done (column 6).

In order to follow the merger history of dark matter halos in each of these simulations, we store the particle position and velocities at different redshifts. A friend-of-friend group finding algorithm is used to locate the virialised halos in each of these slices. We adopt a linking length that is 0.2 times the mean inter-particle separation, corresponding to the density of virialised halos. Only groups containing at least 20 particles are included in our halo catalogs. A merger tree is then constructed out of the halo catalogs by tracking the evolution of each particle through various slices. This lets us identify a halo as it evolves with time through mergers with other halos. We then describe the formation and destruction of halos in terms of change in number of particles between consecutive snapshots of the simulation. When a halo of mass M at redshift z turns into a halo of mass M' at $z' (< z)$, then we say that a halo of mass M was destroyed at redshift z and a halo of mass M' has formed at z' if $M' \geq M(1 + \epsilon)$. We identify the resolution parameter ϵ with that used in our excursion set calculation and experiment with different values as described in the next section.

The presence of ϵ in our analytical model helps in comparison with the N-body

results. It is clear that this parameter is not needed if we care about $\dot{N}(M, t)$ alone. But the formation rate that we are interested in will depend on this parameter, as we will show in the next section. This kind of parameter is also needed for identifying major and minor mergers [Li et al., 2007b].

4.3 Results and Discussion

In this section we present the results of a comparison of our calculations presented in Section 4.1 with N-body simulations. We present comparison of the destruction rate efficiency and the rate of halo formation and then discuss our results at the end of this section. We also consider two related quantities, the halo survival probability and the distribution of halo formation times, which were defined in Section 4.1.

4.3.1 Halo destruction rate efficiency

Figures 4.3 and 4.5 show the halo destruction rate efficiency $\phi(M, t)$ for Sheth-Tormen and Press-Schechter mass functions in an Einstein-de Sitter universe with a power law power spectrum of density fluctuations with indices $n = -0.5$ and $n = -1.5$ respectively. The top row of both figures shows the halo destruction rate efficiency at $z = 0.8$ and the second row shows the same at $z = 0.0$. In each case, we compute the halo destruction rate efficiency using the Sasaki method as well as our excursion set method. We then derive $\phi(M, t)$ from our N-body simulations for a comparison. See Bagla, Khandai & Kulkarni [2009] for details of the simulations and best fit parameters for the ST mass function. These results are superimposed on the plots. For the excursion set calculation and for the comparison with simulations, we use $\epsilon = 0.5$ (left column) and $\epsilon = 0.1$ (right column). For the two power spectra, the two redshifts that we consider correspond to $r_{\text{nl}} = 5$ and $r_{\text{nl}} = 8$ grid lengths, and $r_{\text{nl}} = 4$ and $r_{\text{nl}} = 8$ grid lengths respectively.

As we saw in Figures 4.1 and 4.2, we find that Sasaki's assumption is not valid for ST or PS mass functions, that is $\phi(M, t)$ depends on the halo mass. We also see that the value of $\phi(M, t)$ derived from simulations matches well with that calculated by our method. On the other hand, the predictions of Sasaki's approximation do not match the simulations. This difference is more pronounced for the smaller value of ϵ . Note that the points from N-Body simulations have large error-bars at higher mass as the number of halos decreases at these scales. The most notable feature of the destruction rate efficiency in the excursion set picture is that it cuts off very sharply for large masses. Another aspect is that for small ϵ , there is a pronounced peak in ϕ and it drops off towards smaller masses.

We have also calculated the destruction rate efficiency for the Λ CDM cosmological model for both Press-Schechter and Sheth-Tormen mass functions and compared it with

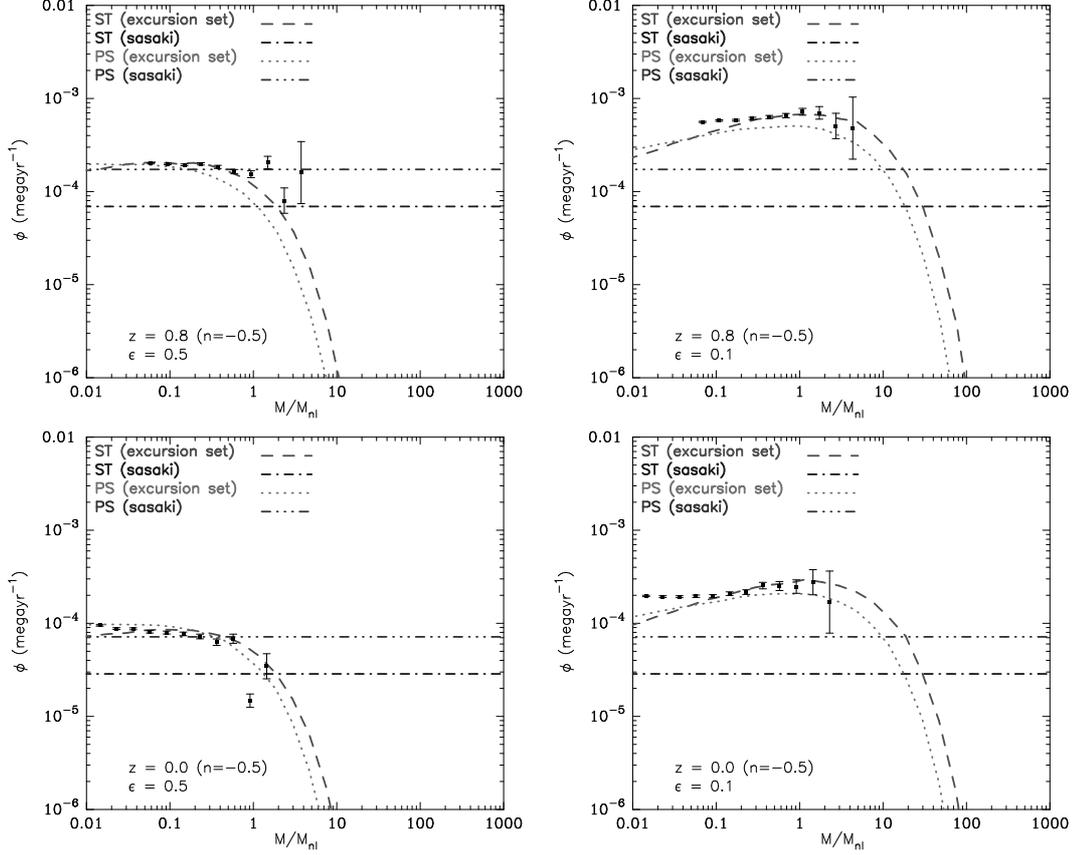


Figure 4.3: Comparison of the destruction rate efficiencies computed using our method and Sasaki formalism for both ST and PS mass function at $r_{nl} = 5$ grid lengths (top row) and $r_{nl} = 8$ grid lengths (second row). All curves are plotted for power-law model with index $n = -0.5$. Curves for $\epsilon = 0.5$ are shown in the left panel and $\epsilon = 0.1$ in the right panel. Points with error bars represent the corresponding results obtained from N-body simulations.

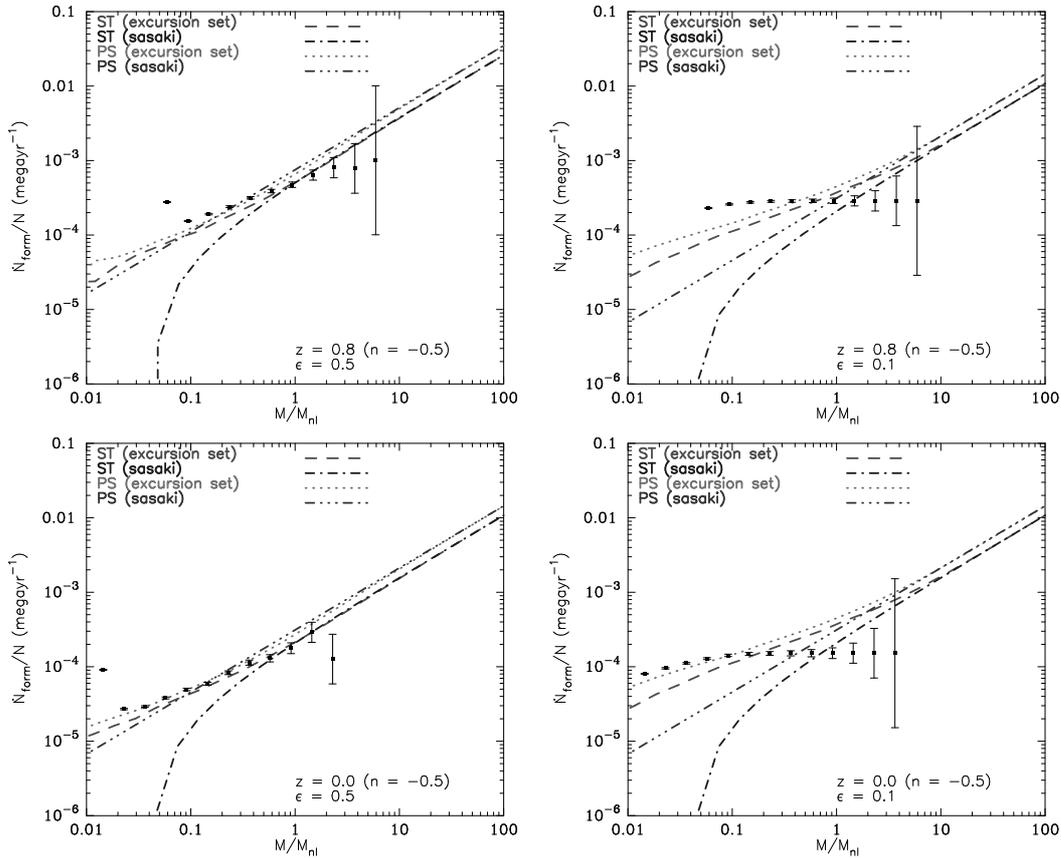


Figure 4.4: Comparison of the formation rates computed using our method and Sasaki formalism for both ST and PS mass function for $r_{nl} = 5$ (third row) and $r_{nl} = 8$ (lowest row). All curves are plotted for power-law model with index $n = -0.5$. Curves for $\epsilon = 0.5$ are shown in the left panel and $\epsilon = 0.1$ in the right panel. Points with error bars represent the corresponding results obtained from N-body simulations.

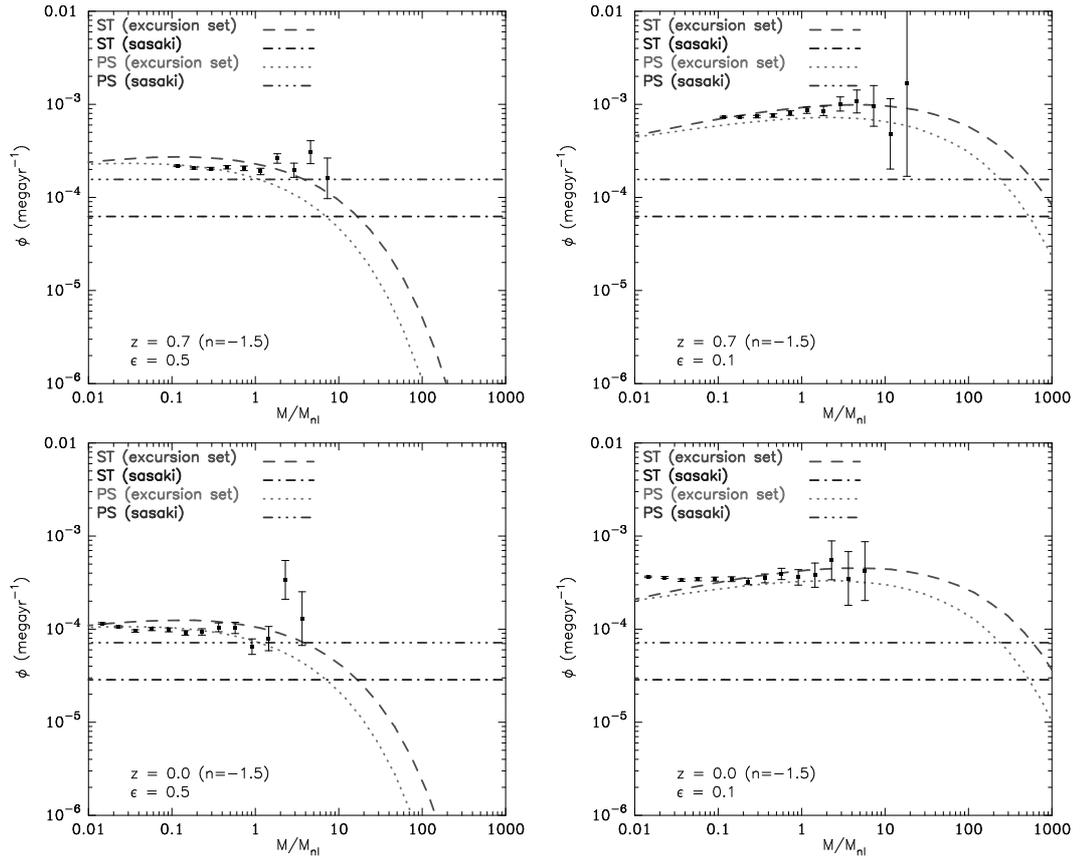


Figure 4.5: Same as Figure 4.3 but now for $n = -1.5$. The two epochs correspond to $r_{nl} = 4$ and $r_{nl} = 8$ grid lengths respectively.

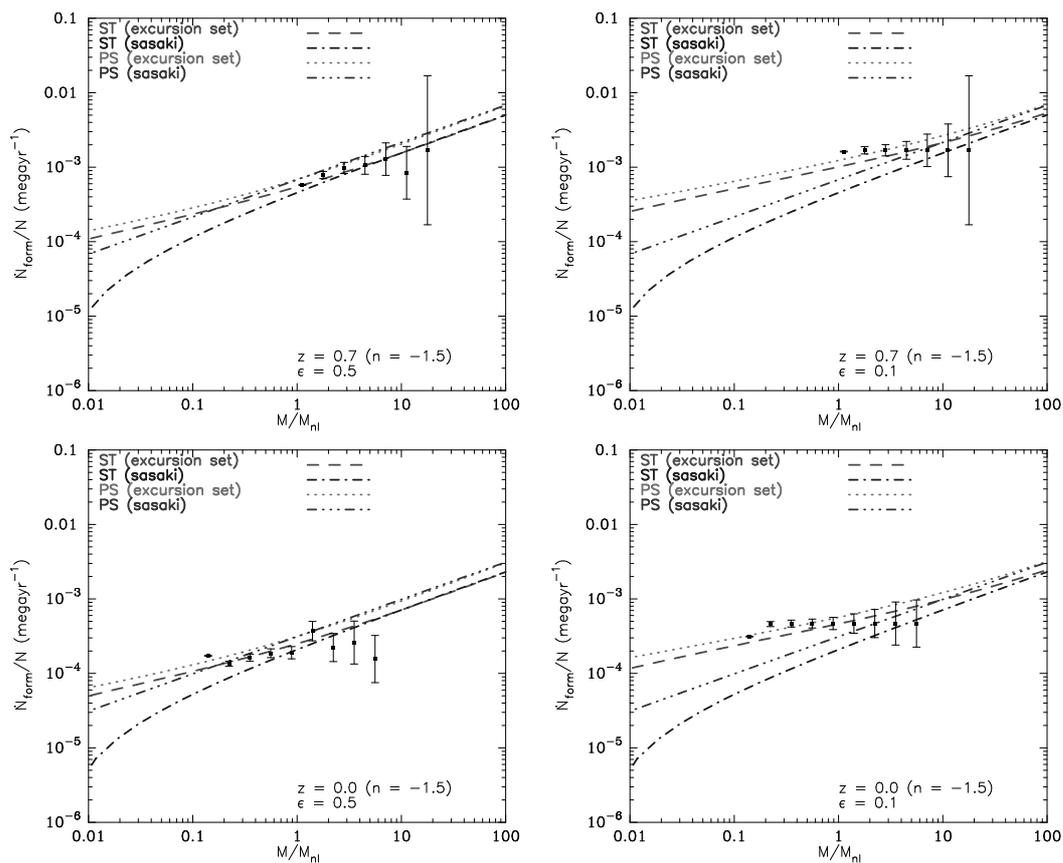


Figure 4.6: Same as Figure 4.4 but now for $n = -1.5$. The two epochs correspond to $r_{nl} = 4$ and $r_{nl} = 8$ grid lengths respectively.

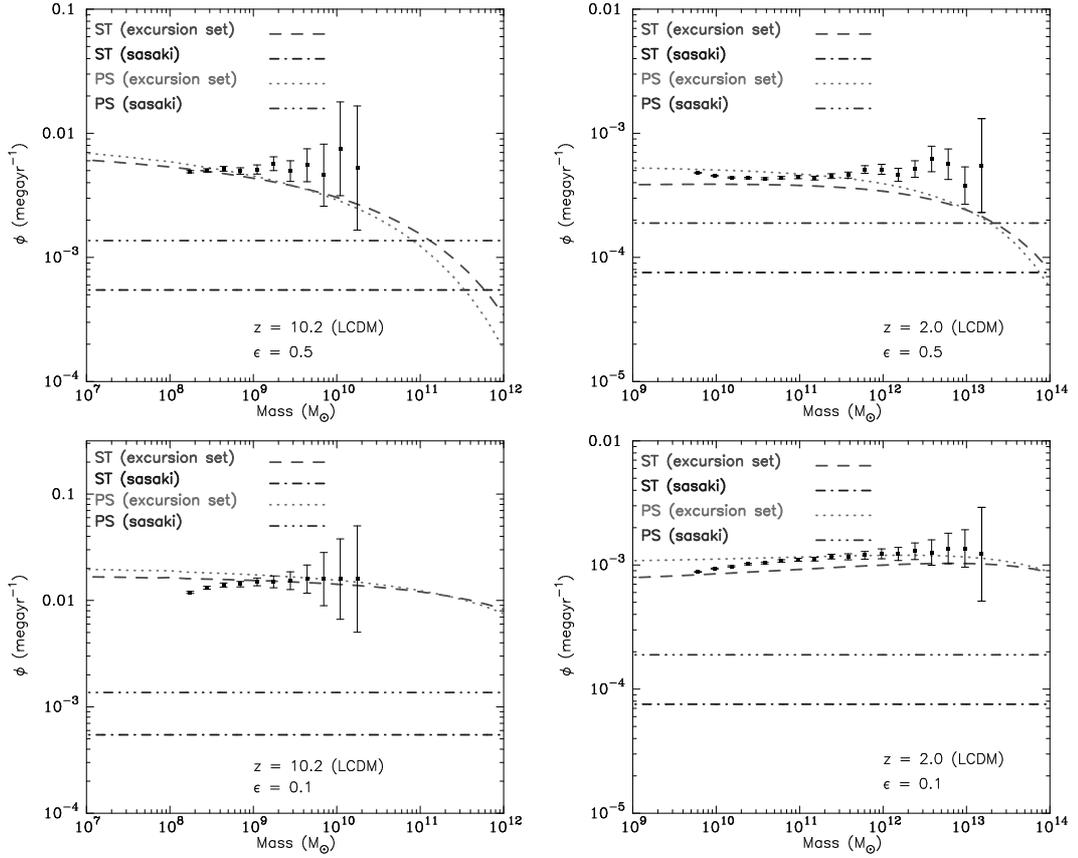


Figure 4.7: Destruction rates for ΛCDM model for both PS and ST mass functions using different thresholds ($\epsilon = 0.5$ for first; $\epsilon = 0.1$ for second) and different redshifts ($z = 10.2$ for left panel, $z = 2.0$ for right panel). Again, points with error bars represent the corresponding results obtained from N-body simulations.

derived values from simulations. The results are shown in Figure 4.7 for three redshifts (2.0, and 10.2) and two values of ϵ (0.5 and 0.1). We can see that results calculated by our technique fit numerical results better.

4.3.2 Halo formation rate

Having calculated the destruction rate efficiency, we can now calculate the halo formation rate using the formalism described in Section 4.1 and compare it with the derived halo formation rates from our simulations. The results are shown in Figures 4.4 and 4.6 for an Einstein-de Sitter Universe with a power law power spectrum of density fluctuations with indices $n = -0.5$ and $n = -1.5$ respectively. The third row of both figures shows the formation rate at redshift $z = 0.8$ and the fourth row shows the same at red-

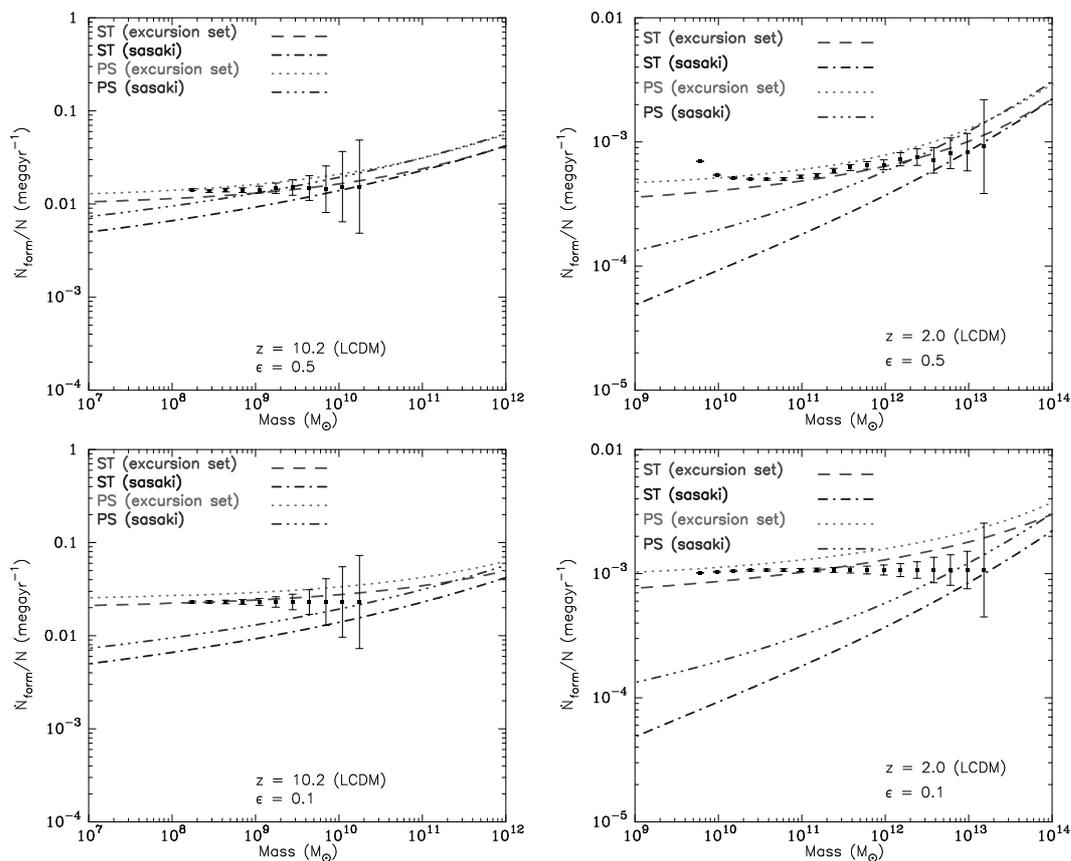


Figure 4.8: Formation rates for ΛCDM model for both PS and ST mass functions using different thresholds ($\epsilon = 0.5$ for first; $\epsilon = 0.1$ for second row) and different redshifts ($z = 10.2$ for left panel, $z = 2.0$ for right panel). As usual, points with error bars represent the corresponding results obtained from N-body simulations.

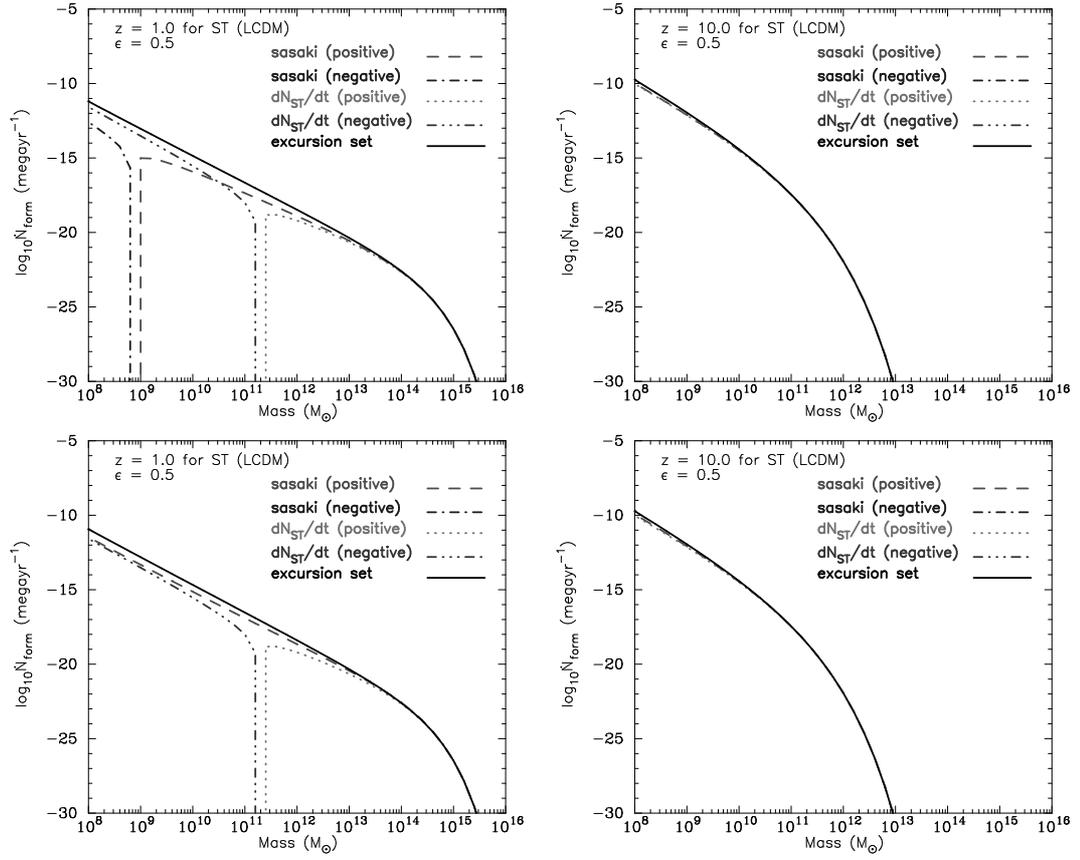


Figure 4.9: Upper panels show formation rates for ST mass function. Lower panels show the same where we used ϕ computed from excursion set approach in the PS mass function and used that to compute the formation rate in the ST mass function.

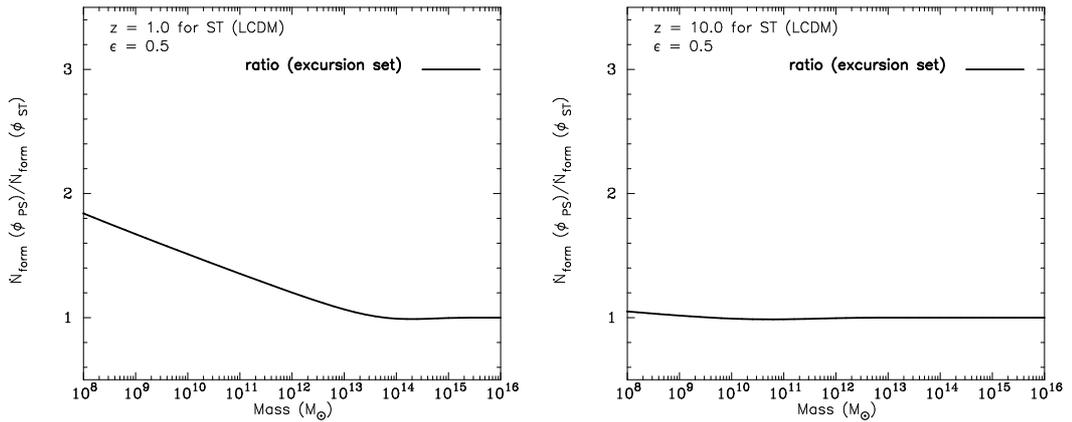
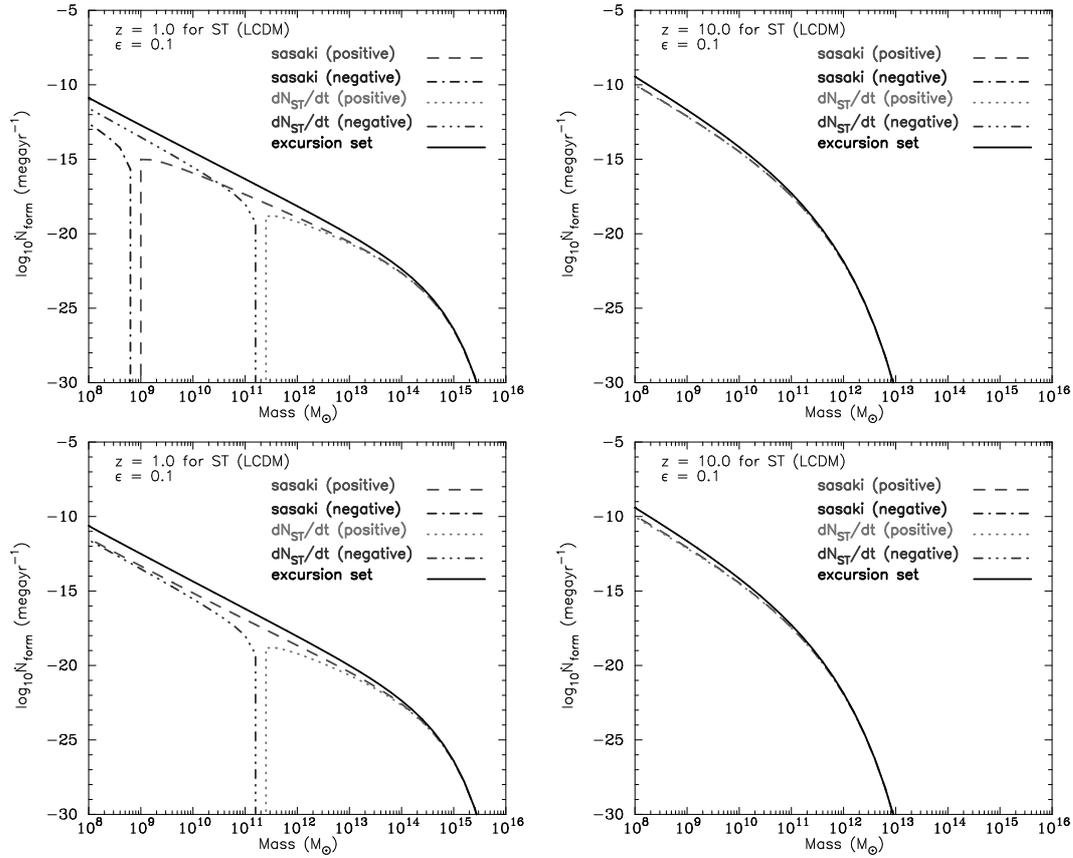
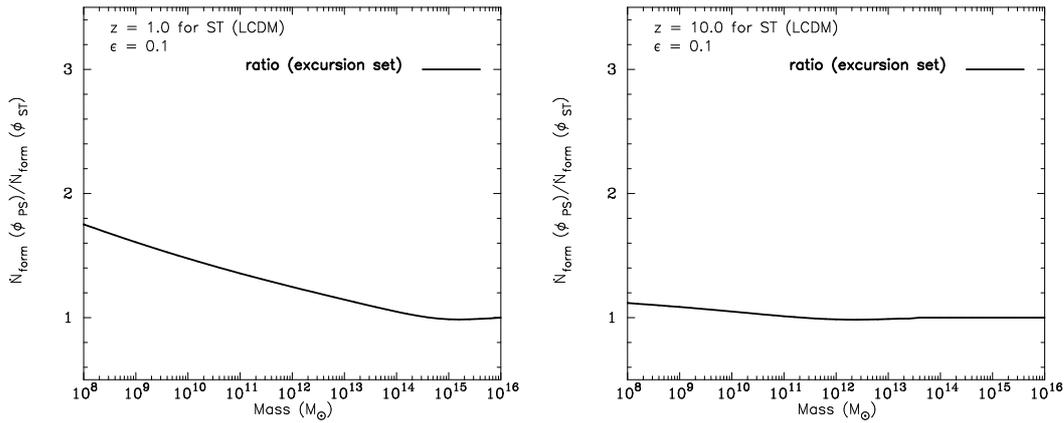


Figure 4.10: Ratio of formation rates estimated in the two approaches shown in Figure 4.9.

Figure 4.11: Same as Figure 4.9 but for $\epsilon = 0.1$ Figure 4.12: Same as Figure 4.10 but for $\epsilon = 0.1$

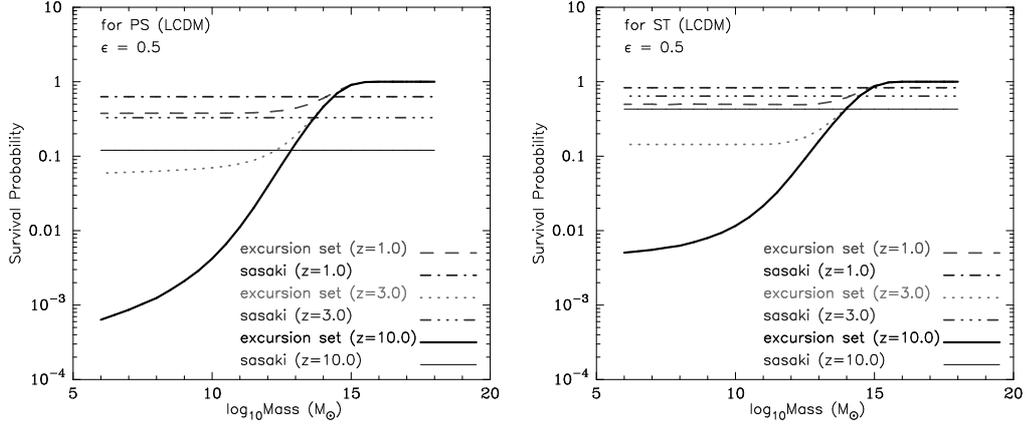


Figure 4.13: Comparison of the survival probabilities computed using our method and Sasaki formalism for both PS (left panel) and ST (right panel) mass functions with different redshifts ($z = 1, 3$ and 10) for $\epsilon = 0.5$. Curves have been plotted for the Λ CDM model. These curves show the probability that the halo survives from that redshift up to the present epoch.

shift $z = 0.0$. Note the quantity plotted here is the ratio $\dot{N}_{\text{form}}(M, t)/N(M, t)$. We have shown the results from the Sasaki prescription and the excursion set calculations and have superimposed formation rates derived from N-Body simulations. As before, for the excursion set calculation and for the comparison with simulations, we use $\epsilon = 0.5$ (left column) and $\epsilon = 0.1$ (right column). For the two power spectra, the two redshifts that we consider correspond to $r_{\text{nl}} = 5$ and $r_{\text{nl}} = 8$ grid lengths, and $r_{\text{nl}} = 4$ and $r_{\text{nl}} = 8$ grid lengths respectively.

Again, we see that the excursion set results fit simulation data much better as compared to the results from Sasaki prescription. The Sasaki method underestimates the formation rates by a large factor for low mass halos. Results from the two methods tend to converge in the large mass limit, although a systematic difference remains between the Sheth-Tormen and Press-Schechter estimates, with the former always being larger than the later. The difference in the Sasaki estimate and the excursion set estimate for the destruction rate efficiency and the formation rate is as high as an order of magnitude at some scales so the close proximity of simulation points to the excursion set calculations is a clear vindication of our approach. It is worth noting that there is a clear deviation of simulation points from the theoretical curves at small mass scales and this deviation is more pronounced at small mass scales for $\epsilon = 0.5$. It may be that some of the deviations arise due to a series representation of the barrier shape, and the number of terms taken into account may not suffice for the estimate. We have found that truncation of the series can affect results at small masses, though in most cases results converge with the five terms that we have taken into account for the range of masses considered here.

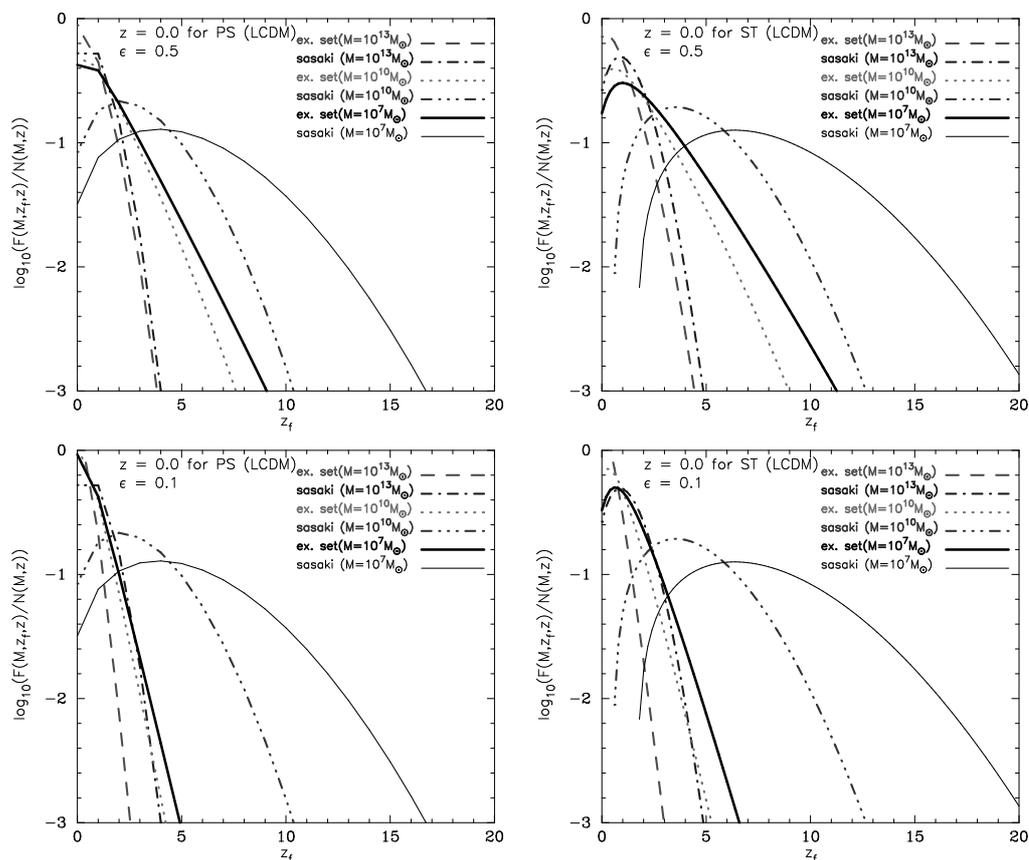


Figure 4.14: Plots for formation epoch distribution of halos. Left column is for the PS and the right column is for the ST mass function. Curves have been plotted for the LCDM model. The formation epoch distribution as computed using the Sasaki formalism and the excursion set approach described in this work is shown in the top panel.

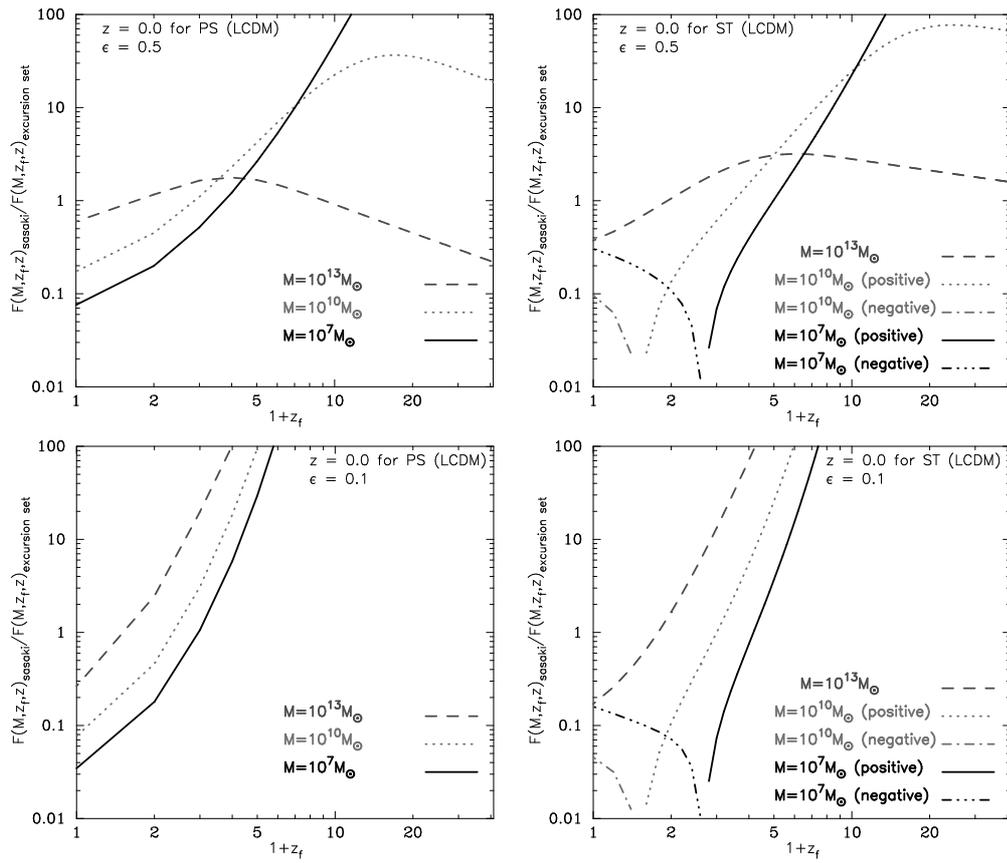


Figure 4.15: Ratio of the two different approaches used in Figure 4.14 to highlight less obvious differences.

4.3.3 Halo survival probability

An important auxiliary quantity in the ongoing discussion is the halo survival probability, defined in Section 4.1. From our calculation of the halo destruction rate efficiency, we calculated the survival probability of dark matter halos using both the excursion set formalism and the Sasaki prescription and compared results. These results are shown in Figure 4.13, which shows the survival probabilities in the Λ CDM cosmological model for the Press-Schechter (left panel) and Sheth-Tormen (right panel) mass functions using the two approaches at three different redshifts ($z = 1, 3$ and 10). In this case, we have used $\epsilon = 0.5$ for the excursion set calculation.

In Sasaki approximation, the destruction rate is independent of mass and hence the survival probability is also independent of mass. Our calculations show that this approximation is not true, and hence the survival probability of halos must also depend on mass. We note that the survival probability is high for large mass halos: if a very large mass halo forms at a high redshift then it is likely to survive without a significant addition to its mass. Smaller halos are highly likely to merge or accrete enough mass and hence do not survive for long periods. Survival probability drops very rapidly as we go to smaller masses. While this is expected on physical grounds, it is an aspect not captured by the Sasaki approximation where equal survival probability is assigned to halos of all masses. The mass dependence of survival probability is qualitatively similar to that obtained by Kitayama & Suto [1996]. There is no significant qualitative difference between the curves for the Press-Schechter and the Sheth-Tormen mass functions.

4.3.4 Formation time distribution

Finally, another interesting quantity is the distribution $F(M; t_f, t)$ of formation epochs t_f of halos with mass M at t , defined in Section 4.1. This distribution can be obtained once the survival probability and formation rate of halos is known. We calculate the formation time distribution using the excursion set formalism and the Sasaki prescription. The results are shown in Figure 4.14. We plot $F(M; z_f, z = 0)/N(M, z = 0)$ versus the formation redshift z_f for three different masses (10^{13} , 10^{10} and $10^7 M_\odot$) in the standard Λ CDM model for both Press-Schechter (left column) and Sheth-Tormen (right column) mass functions with $\epsilon = 0.5$ (first row) and $\epsilon = 0.1$ (second row). A common feature is that F as a function of z_f increases up to a certain redshift and then starts to decline. The epochs at which F drops by an order of magnitude from its peak can be interpreted as typical range of redshifts for the formation of bound systems of respective masses which exist at $z = 0$.

The differences between the formation redshift distribution for $\epsilon = 0.5$ and $\epsilon = 0.1$ are along expected lines: the formation redshifts are smaller for the lower value of ϵ as a smaller change in mass is required to declare that a new halo has formed and hence typical halos do not survive for a very long time. We see that the excursion set calcula-

tion suggests that halos formed more recently as compared to the Sasaki approximation based estimate. This can be understood in terms of the equal survival probability assigned by the Sasaki approximation to halos of all masses. For a clearer comparison, the ratio of the estimate based on Sasaki approximation and the excursion set calculation is shown in Figure 4.15. We note that for very low mass halos these two estimates differ by more than an order of magnitude. The main qualitative difference between the plots for the Press-Schechter and the Sheth-Tormen mass functions is caused by the negative formation rates in the Sasaki approximation.

4.3.5 Discussion

The results described above show conclusively that the excursion set approach predicts halo formation and destruction rates that match with simulations much better than the Sasaki approximation.

Another noteworthy aspect is that the destruction and formation rates depend on the value of ϵ in simulations as well as the excursion set calculation thereby allowing us to differentiate between major and minor mergers. In comparison, there is no natural way to bring in this dependence in the Sasaki approximation. While the match between simulations and the excursion set approach for the two values of ϵ is satisfying, it raises the question of the appropriate value of this parameter. In our view the appropriate value of the parameter should depend on the application in hand. In semi-analytic galaxy formation models, we should use a value of ϵ that corresponds to the smallest ratio of masses of the in-falling galaxy and the host galaxy where we expect a significant dynamical influence on star formation rate. For instance, Kauffmann et al. [1999] use $\epsilon = 0.3$ in their semi-analytic galaxy formation model while considering formation of bulges in merger remnants. In case of galaxy clusters we may base this on the smallest ratio of masses where the intra-cluster medium is likely to be disturbed in a manner accessible to observations in X-ray emission or the Sunyaev-Zel'dovich effect [Sunyaev & Zeldovich, 1972; Navarro, Frenk & White, 1995; Kay, 2004].

While the close match between simulations and the excursion set calculation is useful, it also implies that we should not use the simpler Sasaki approximation. The excursion set calculation of the halo destruction rate is fairly simple for the Press-Schechter mass function, but the corresponding calculation for the Sheth-Tormen mass function is much more complicated. Plots of the destruction rate efficiency $\phi(M)$ for all the models suggest that its variation with mass and ϵ is very similar for the PS and ST mass function. This suggests an approximation where we use $\phi(M, z; \epsilon)$ computed using the Press-Schechter mass function and use that to compute the halo formation rate in the Sheth-Tormen mass function. Figures 4.9 and 4.11 show the halo formation rate for the Λ CDM model at three redshifts and compares the excursion set calculation, the Sasaki approximation and the intermediate approximation suggested above. We find that the intermediate approximation is not plagued by negative halo formation rates and that it

is an excellent approximation at all mass scales at higher redshifts. At lower redshifts, the approximation is still good at high masses but not so at smaller masses.

4.4 Conclusions

Key points presented in this chapter can be summarized as follows:

- We revisited the Sasaki approximation for computing the halo formation rate and computed the destruction rate explicitly using the excursion set approach.
- We introduced a parameter ϵ , the smallest fractional change in mass of a halo before we consider it as destruction of the old halo and formation of a new halo.
- We showed that the halo destruction rate is not independent of mass even for power law models and hence the basis for the Sasaki ansatz does not hold. Two prominent features of the halo destruction rate are the rapid fall at large masses, and a pronounced peak close to the scale of non-linearity. The peak is more pronounced for smaller values of ϵ .
- Using the excursion set approach for the Sheth-Tormen mass function leads to positive halo formation rate, unlike the generalization of the Sasaki ansatz where formation rate at some mass scales is negative.
- We compared the destruction rate and the halo formation rate computed using the excursion set approach with N-Body simulations. We find that our approach matches well with simulations for all models, at all redshifts and also for different values of ϵ .
- In some cases there are deviations between the simulations and the theoretical estimate. However, these deviations are much smaller for the excursion set based method as compared to the Sasaki estimate. It may be that some of the deviations arise due to a series representation of the barrier shape, and the number of terms taken into account may not suffice for the estimate. We have found that truncation of the series can affect results at small masses, though in most cases results converge with the five terms that we have taken into account for the range of masses considered here.
- We showed that we can use the halo destruction rate computed for the Press-Schechter mass function to make an approximate estimate of the halo formation rate in Sheth-Tormen mass function using Equation (4.5). This approximate estimate is fairly accurate at all mass scales in the Λ CDM model at high redshifts.

- The halo survival probability is a strong function of mass of halos, unlike the mass independent survival probability obtained in the Sasaki approximation.
- The halo formation redshift distribution for halos of different masses is also very different from that obtained using the Sasaki approximation. This is especially true for the Sheth-Tormen mass function where the Sasaki approximation gives negative halo formation rates in some range of mass scales and redshifts.

The formalism used here for calculation of halo formation rate and other related quantities can be generalized to any description of the mass function if the relevant probabilities can be calculated. Within the framework of the universal approach to mass functions, it can also be used to study formation rates of halos in different cosmological models [Linder & Jenkins, 2003; Macciò et al., 2004]. This allows for an easy comparison of theory with observations for quantities like the major merger rate for galaxy clusters [Cohn, Bagla & White, 2001].

In case of semi-analytic models of galaxy formation, our approach allows for a nuanced treatment where every merger need not be treated as a major merger and we may only consider instances where mass ratios are larger than a critical value for any affect on star formation in the central galaxy.

Formation of galactic nuclei with multiple supermassive black holes at high redshifts

In the previous chapters, we looked at analytic models of evolution of galaxies and the IGM. These models calculated various properties of galaxies and the IGM—like the star formation history and average IGM temperature—in their global average. This chapter onwards, we go beyond the global average to look into some specific aspects of galaxy formation.

Most local galaxies host supermassive black holes (SMBHs) at their centres [Richstone et al., 1998; Ferrarese & Ford, 2005]. The SMBH mass M_{bh} is correlated with properties of the spheroidal nucleus of the host galaxy, such as velocity dispersion [Ferrarese & Merritt, 2000; Gebhardt et al., 2000; Ferrarese, 2002; Gültekin et al., 2009] and luminosity [Magorrian et al., 1998; McLure & Dunlop, 2002; Marconi & Hunt, 2003; Gültekin et al., 2009]. Detection of bright quasars at redshifts $z \gtrsim 6$ [Fan et al., 2001; Mortlock et al., 2011] suggests that SMBHs with masses as high as $\sim 2 \times 10^9 M_{\odot}$ already existed at $z \sim 7$. In the standard Λ CDM cosmological model, growth of galaxies is hierarchical and galaxy mergers are expected to be particularly frequent at redshifts $z \sim 6$ –20. As galaxies merge, their central SMBHs can grow through coalescence and accretion of gas. It is commonly postulated that SMBHs at lower redshifts grew out of seed black holes (BHs) in the first galaxies [Loeb & Rasio, 1994; Eisenstein & Loeb, 1995; Kauffmann & Haehnelt, 2000; Menou, Haiman & Narayanan, 2001; Bromm & Loeb, 2003; Volonteri, Haardt & Madau, 2003; Hopkins et al., 2006; Tanaka & Haiman, 2009].

Existing merger tree models are based on the assumption that any binary black hole system, which inevitably forms in a galaxy’s merger history, coalesces on a short time-scale. However, the evolution of SMBH binaries is a complex open problem and it is unclear if a binary can merge within a Hubble time [Merritt & Milosavljević, 2005]. One expects that during a merger event of two galaxies, the dynamics of their constituent

SMBHs would proceed in three stages [Begelman, Blandford & Rees, 1980]. In the first stage, the SMBHs sink to the centre of the gravitational potential of the merger remnant by dynamical friction and form a gravitationally bound binary. The newly-formed binary continues to lose energy and angular momentum through its global gravitational interaction with many stars until the separation between the SMBHs reduces to a value at which the dominant mechanism of energy loss is the 3-body interaction between the binary and individual stars. This is the second stage of the binary’s evolution, and is known as the ‘hard stage.’ The precise definition of a hard SMBH binary varies in the literature, but it is commonly assumed that the binary becomes hard when its semi-major axis a reaches a value given by [Yu, 2002]

$$a \approx a_h \equiv \frac{Gm}{4\sigma^2} = 2.8 \left(\frac{m}{10^8 M_\odot} \right) \left(\frac{200 \text{ km s}^{-1}}{\sigma} \right)^2 \text{ pc}, \quad (5.1)$$

where stars in the galactic nucleus are assumed to have a one-dimensional velocity dispersion σ , and m denotes the mass of the lighter SMBH. Finally, once the binary separation decreases to a small-enough value, gravitational wave emission becomes the dominant mode of energy loss and the SMBHs coalesce rapidly. This is the third stage of the SMBH binary evolution. The value of semi-major axis a at which the coalescence time scale due to gravitational wave emission alone is t is given by [Peters, 1964; Loeb, 2010]

$$a(t) \equiv a_{\text{gw}}(t) = 4.3 \times 10^{-3} \left(\frac{t}{10^5 \text{ yr}} \right)^{1/4} \left(\frac{M}{2 \times 10^8 M_\odot} \right)^{3/4} \text{ pc}, \quad (5.2)$$

where M is the total mass of the binary, and we have considered two SMBHs with mass $10^8 M_\odot$ each on a circular orbit (with shorter time scale at increasing eccentricity). Gravitational wave emission takes over as the dominant mode of energy loss when $a = a_{\text{gw}}(t_h)$, where t_h is the hardening time scale.

Among these three stages of evolution of an SMBH binary, the largest uncertainty in the binary’s lifetime originates from the hard stage, which can be the slowest stage since the binary quickly ejects all low angular momentum stars in its vicinity, thus cutting off its supply of stars. This is known as the “final parsec problem” [Milosavljević & Merritt, 2003b]. For example, Yu [2002] studied coalescence of SMBH binaries in a sample of galaxies observed by Faber et al. [1997] and found that spherical, axisymmetric or weakly triaxial galaxies can all have long-lived binary SMBHs that fail to coalesce. Similarly, Merritt & Milosavljević [2005] found that the time spent by a binary is less than 10^{10} yr only for binaries with very low mass ratios ($\lesssim 10^{-3}$).¹ Furthermore, Merritt & Milosavljević [2005] showed that a binary may not be able to interact with all the stars in its loss cone, thereby increasing the time spent in the hard stage even further;

¹However, for such low mass ratios the time taken by the lighter black hole to reach the galactic nucleus due to dynamical friction is itself expected to exceed the Hubble time.

they found that in a nucleus with a singular isothermal sphere stellar density profile, an equal-mass binary will stall at a separation of $a \approx a_h/2.5$, where we have defined a_h in Equation (5.1). The final separation is expected to be even higher for galaxies with shallower density profiles.

Several ways have been discussed in the literature to efficiently extract energy and angular momentum from a hard SMBH binary and overcome the final parsec problem. An example is the work by Armitage & Natarajan [2002], who suggested that gas can catalyse the coalescence of a hard SMBH binary by serving as an effective sink for the binary's angular momentum. In particular, they found that a binary with a separation of 0.1 pc embedded in a gaseous accretion disk would merge in 10^7 years without significant enhancement in the gas accretion rate. Similarly, Escala et al. [2004, 2005] found that in SPH simulations, clouds of hot gas ($T_{\text{gas}} \approx T_{\text{virial}}$) can induce decay of orbits of embedded binary point masses due to gravitational drag. A caveat to these studies is that feedback from gas accretion onto the SMBHs can remove the rest of the gas from the merger remnant before the binary coalesces. However, stellar dynamical processes could also accelerate binary coalescence, without gas. For example, Merritt & Poon [2004] considered the effect of chaotic orbits in steep triaxial potentials. They found that stars are supplied to the central black hole at a rate proportional to the fifth power of the stellar velocity dispersion and that the decay rate of a central black hole binary would be enhanced even if only a few percent of the stars are on chaotic orbits, thus solving the final parsec problem. As another example, it was suggested that a third SMBH closely interacting with a hard SMBH binary can reduce the binary separation to a small value either due to the eccentricity oscillations induced in the binary via the Kozai-Lidov mechanism [Blaes, Lee & Socrates, 2002] or due to repopulation of the binary's loss cone due to the perturbation in the large-scale potential caused by the third black hole [Hoffman & Loeb, 2007]. Blaes, Lee & Socrates [2002] found that the merger time scale of an inner circular binary can be shortened by as much as an order of magnitude, and that general relativistic precession does not destroy the Kozai-Lidov effect for hierarchical triples that are compact enough.

In summary, there is substantial uncertainty in the current understanding of the evolution of binary SMBHs. Clearly, if the SMBH binary coalescence time is longer than the typical time between successive major mergers of the galaxy, then more than two SMBHs may exist in the nucleus of a merger remnant. We study this possibility in this chapter. We calculate the relative likelihood of binary, triple, and quadruple SMBH systems, by considering the timescales for relevant processes and combining galaxy merger trees with direct-summation N-body simulations for the dynamics of stars and SMBHs in galactic nuclei. An obvious question regarding galactic nuclei with multiple SMBHs is whether such systems can be long-lived. We consider this question here. Finally, systems with multiple SMBHs are likely to be interesting because of observational effects involving their effect on the properties of the host bulge, the enhancement in the rate of

tidal disruption of stars, their associated gravitational wave and electromagnetic signals, and the slingshot ejection of SMBHs at high speeds. We study some of these effects.

5.1 Previous work

Galactic nuclei with multiple SMBHs were first studied by Saslaw, Valtonen & Aarseth [1974], who computed orbits of three and four SMBH systems by sampling the parameter space of the problem. They showed that if an infalling SMBH is lighter than the components of the pre-existing binary, then the most probable outcome is a slingshot ejection in which the infalling SMBH escapes at a velocity that is about a third of the orbital velocity of the binary. Valtonen [1976] further showed that the ejection velocity can be significantly enhanced if drag forces due to gravitational radiation are accounted for in the three-body dynamics. The formation of systems with three or four SMBHs in a hierarchical merger of smooth galactic potentials was first studied by Mikkola & Valtonen [1990] and Valtonen et al. [1994] with the objective of understanding the structure of extragalactic radio sources. This line of work was extended to binary-binary scattering of SMBHs by Heinämäki [2001], and by Hoffman & Loeb [2007], who studied repeated triple interactions in galactic nuclei. Both of these studies used cosmologically consistent initial conditions based on the extended Press-Schechter theory. Systems with a larger number of black holes were studied by Hut & Rees [1992] and Xu & Ostriker [1994] using simple analytical models and numerical calculations of massive particles in smooth galactic potentials. Xu & Ostriker [1994] concluded that the most-likely outcome in these cases is one in which most black holes are ejected and the galactic center is left with zero, or one, or two black holes. Finally, full N-body simulations of galactic nuclei with constituent SMBHs were performed for the case of two successive mergers by Makino & Ebisuzaki [1996], Makino [1997], and Iwasawa, Funato & Makino [2006]. Much of this work on SMBHs was based on earlier studies of stellar-mass black holes in globular clusters. Sigurdsson & Hernquist [1993] and Kulkarni, Hut & McMillan [1993] considered the evolution of ~ 100 stellar mass black holes in globular clusters. They concluded that after mass segregation, most of these black holes are ejected out on a short time scale, and the globular cluster is left with none or a few black holes. Mass segregation and associated effects of stellar-mass black holes in a galactic nucleus with a central SMBH was also considered [Miralda-Escudé & Gould, 2000; Freitag, Amaro-Seoane & Kalogera, 2006].

The possible formation of systems with multiple SMBHs due to successive galactic mergers arises naturally in any model describing the hierarchical assembly of galaxies. One approach to modeling SMBH growth involves constructing semi-analytic prescriptions of various characteristic processes, like mergers of galaxies, formation of spheroids, star formation, and gas thermodynamics, coupled with merger trees of dark matter haloes. This approach has been adopted, for example, by Kauffmann & Haehnelt

[2000], who also extended it to study possible formation of multiple SMBH systems and implications for the $M_{\text{bh}}-\sigma$ relation and density profiles observed in luminous elliptical galaxies [Haehnelt & Kauffmann, 2002]. Another study by Volonteri, Haardt & Madau [2003] followed merger trees of dark matter haloes and their component SMBHs using Monte Carlo realizations of hierarchical structure formation in the Λ CDM cosmology. They modeled dark matter haloes as singular isothermal spheres and calculated the in-spiral of less massive halos in more massive ones by using the Chandrasekhar formula for dynamical friction. Gas accretion to the SMBHs was modeled so as to reproduce the empirical $M_{\text{bh}}-\sigma$ relation and the SMBH dynamics was described with analytic prescriptions. In particular, the coalescence time of hard SMBH binaries was calculated from a set of coupled differential equations based on scattering experiments involving the ejection of stellar mass from the loss cone due to the hard SMBH binary and the resultant change in the hardening rate [Quinlan, 1996; Merritt, 2000]. For galaxies that underwent another major merger before their constituent binary SMBH coalesced, a three-body interaction was implemented between the binary and the intruder SMBH. They found that the smallest SMBH was kicked out of the galaxy in 99% of cases, while the binary escapes the galaxy in 8 % of cases. Thus, a significant fraction of galactic nuclei could end up with no SMBHs or offset SMBHs with mass lower than that expected from the $M_{\text{bh}}-\sigma$ relation. These results were later extended to incorporate recoil in the SMBH merger remnant due to asymmetric emission of gravitational waves, which mainly affected the $M_{\text{bh}}-\sigma$ relation for low-mass haloes by increasing the scatter [Volonteri & Rees, 2006; Volonteri, 2007; Blecha et al., 2011]. Similar semi-analytic models were studied by several other authors to understand the assembly of $z \sim 6$ quasars. However, most of these models ignored the dynamics of multiple SMBHs and assumed prompt coalescence [Haiman & Loeb, 1999b; Wyithe & Loeb, 2003b; Yoo & Miralda-Escudé, 2004; Tanaka & Haiman, 2009; Shen, 2009]. As a result, these models did not treat systems with multiple SMBHs.

Lastly, SMBH assembly has also been studied using smooth particle hydrodynamic simulations that attempted to calculate effects of both the gas physics as well as the gravitational dynamics of the large-scale structure within and around galaxies [Hopkins et al., 2006; Sijacki et al., 2007; Li et al., 2007a; Hopkins et al., 2007]. However, due to poor mass resolution and particle smoothing, these simulations cannot accurately calculate the detailed dynamics of a multiple SMBH systems. Indeed, in most of these studies, black hole coalescence occurs on scales smaller than the smoothing length, which is usually much larger than the expected separation of a hard SMBH binary. As a result, SMBH coalescence is implemented via a subgrid model. Prompt BH coalescence is assumed in these models, without calculating either the effect of the environment on BH evolution, or the back-reaction of the BHs on the environment. Since these models have been used to explain the existence of the $z \gtrsim 6$ SDSS quasars, it is crucial to confirm the assumptions made in their sub-grid prescriptions. Here, we explore for the

first time numerical simulations that incorporate the cosmological process of galaxy mergers in the cosmological context along with an accurate treatment of black hole dynamics.

5.2 Formation of multiple-SMBH systems

Unless they coalesce rapidly, or get kicked out of the host galactic nucleus, we expect multi-SMBH systems to form in galactic nuclei at high redshift due to mergers of galaxies if the typical black hole coalescence timescale is longer than the feeding timescale of new incoming black holes. In this section, we establish a simple theoretical framework for this formation path using analytical estimates of its relevant timescales: *(i)* the major merger time scale of galaxies; *(ii)* the time scale on which a satellite galaxy sinks to the center of a host galaxy so that a close interaction between SMBHs can occur; and *(iii)* the time scale of SMBH coalescence.

5.2.1 Time scale of incoming SMBHs

Fakhouri, Ma & Boylan-Kolchin [2010] have quantified the average merger rate of dark matter haloes per halo per unit redshift per unit mass ratio for a wide range of halo mass, progenitor mass ratios and redshift. The result is given by a fitting formula derived from the Millennium [Springel et al., 2005] and Millennium-II [Boylan-Kolchin et al., 2009] simulations:

$$\frac{dN}{d\xi dz}(M, \xi, z) = A \left(\frac{M}{10^{12} M_{\odot}} \right)^{\alpha} \xi^{\beta} \exp \left[\left(\frac{\xi}{\tilde{\xi}} \right)^{\gamma} \right] (1+z)^{\eta}. \quad (5.3)$$

Here, M is the halo mass at redshift z , and ξ is the mass ratio of progenitors. Mergers with $\xi > 0.3$ are considered major mergers. The best fit values of various parameters are $\alpha = 0.133$, $\beta = -1.995$, $\gamma = 0.263$, $\eta = 0.0993$, $A = 0.0104$ and $\tilde{\xi} = 9.72 \times 10^{-3}$. The average major merger rate per unit time is then given by

$$\frac{dN_m}{dt}(M, z) = \int_{0.3}^1 d\xi \frac{dN}{d\xi dz}(M, \xi, z) \frac{dz}{dt}. \quad (5.4)$$

Fakhouri, Ma & Boylan-Kolchin [2010] also provide a fitting formula for average mass growth rate of halos that can be used to calculate the halo mass at redshift z for use in equation (5.3),

$$\dot{M}(z) = 46.1 \frac{M_{\odot}}{\text{yr}} (1 + 1.1z) \sqrt{\Omega_m (1+z)^3 + \Omega_{\Lambda}} \left(\frac{M}{10^{12} M_{\odot}} \right)^{1.1}. \quad (5.5)$$

Using equation (5.4) we can now define the time scale of major mergers for a halo as

$$t_{\text{mrg}} = \left[\frac{dN_m}{dt} \right]^{-1}. \quad (5.6)$$

The behavior of this quantity is shown in Figure 5.1 for three halo masses that discussed here: a Milky Way-like halo that has a mass $M_0 = 10^{12} M_\odot$ at $z = 0$, the typical halo today that has mass $M_0 = 10^{14} M_\odot$ at $z = 0$, and rare haloes with mass $M_0 = 10^{16} M_\odot$ at $z = 0$.² This is the time scale at which we expect new (satellite) haloes to enter the halo. As expected, halo mergers are more frequent at higher redshift. At redshift $z \lesssim 1$ the major merger time scale for a Milky Way-like halo is greater than the Hubble time.

After two dark matter haloes have merged, the smaller halo becomes a satellite halo within the virial radius of the host halo. It then takes this satellite a dynamical friction time to sink to the center of the host halo, so that the constituent galaxies can merge. As a result, the timescale for major mergers of galaxies is expected to be different that the time scale for major mergers of dark matter haloes calculated in Equation (5.6).

The dynamical friction time scale is often estimated using Chandrasekhar's formula [Chandrasekhar, 1943a; Lacey & Cole, 1993; Binney & Tremaine, 2008]:

$$t_{\text{df}} = \frac{f_{\text{df}} \Theta_{\text{orb}}}{\ln \Lambda} \frac{M_{\text{host}}}{M_{\text{sat}}} t_{\text{dyn}}, \quad (5.7)$$

where M_{host} and M_{sat} are the masses for the host and satellite haloes respectively, $\ln \Lambda$ is the coulomb logarithm, Θ_{orb} is a function of the orbital energy and angular momentum of the satellite, f_{df} is an adjustable parameter of order unity and t_{dyn} is the halo dynamical time scale calculated at the virial radius. Equation (5.7) is valid only in the limit of small satellite mass in an infinite, isotropic and homogeneous collisionless medium. Still, it has been used in the literature even for large satellite masses by modifying the Coulomb logarithm. In recent years, deviations from predictions by equation (5.7) have been reported in both the $M_{\text{sat}} \ll M_{\text{host}}$ and $M_{\text{sat}} \lesssim M_{\text{host}}$ regimes [Taffoni et al., 2003; Monaco, Fontanot & Taffoni, 2007; Boylan-Kolchin, Ma & Quataert, 2008; Jiang et al., 2008; Wetzel, Cohn & White, 2009].

To correct the problems associated with Chandrasekhar's formula, several groups have developed full dynamical models of evolution of merging haloes [Taylor & Babul, 2001; Gnedin, 2003; Taffoni et al., 2003; Zentner et al., 2005]. For example, one of the approaches to overcome the limits of Chandrasekhar's formula is the theory of linear response (TLR; Colpi, Mayer & Governato 1999). TLR captures the backreaction of the

²In this chapter, M_0 always denotes the halo mass at redshift $z = 0$. We also refer to the average mass of such haloes at other redshifts, by e. g. $M_{z=4}$ and $M_{z=6}$. A halo with $M_0 = 10^{12} M_\odot$ will have $M_{z=6} = 2 \times 10^{10} M_\odot$. A halo with $M_0 = 10^{14} M_\odot$ will have $M_{z=6} = 5 \times 10^{11} M_\odot$. A cluster-size halo, with mass $M_0 = 10^{15} M_\odot$ will have $M_{z=6} = 5 \times 10^{12} M_\odot$ and is expected to hold a single galaxy at that redshift.

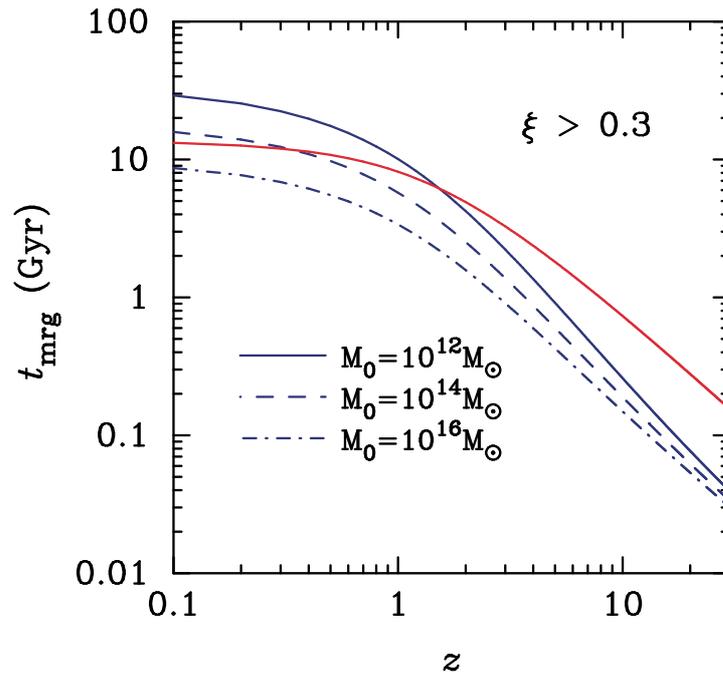


Figure 5.1: Halo major merger time scale (mass ratio > 0.3), according to equation (5.6), for haloes with mass $M_0 = 10^{12} M_\odot$ (blue solid line), $10^{14} M_\odot$ (blue dashed line) and $10^{16} M_\odot$ (blue dot-dashed line). The Hubble time is shown by the solid red curve. Major mergers are more frequent at higher redshifts. On average, Milky Way-sized haloes are not expected to undergo a major merger for $z \lesssim 1$. Galaxy major merger time scale is always longer than the subsequent dynamical friction time scale.

stellar distribution to the intruding satellite by correlating the instantaneous drag force on it with the drag force at an earlier time via the fluctuation-dissipation theorem. Tidal stripping of a satellite halo is an important ingredient in this formulation. In a singular isothermal sphere with 1D velocity dispersion σ and density profile $\rho(r) = \sigma^2/[2\pi Gr^2]$, TLR predicts a sinking time

$$t_{\text{df}} = 1.17 \frac{r_{\text{cir}}^2 V_{\text{cir}}}{GM_{\text{sat}} \ln \Lambda} \epsilon^\alpha, \quad (5.8)$$

where ϵ is the circularity (defined as the ratio between the angular momentum of the current orbit relative to that of a circular orbit of equal energy), r_{cir} and V_{cir} are the initial radius and velocity of the circular orbit with the same energy of the actual orbit, and M_{S} is the mass of the incoming satellite halo. Numerical simulations suggest a value of $0.4 - 0.5$ for the exponent α [van den Bosch et al., 1999; Colpi, Mayer & Governato, 1999; Volonteri, Haardt & Madau, 2003].

Given the limitations of analytical treatments, we turn to results of numerical simulations to understand the dynamical friction time scale. Using N-body simulations, Boylan-Kolchin, Ma & Quataert [2008] give a fitting formula that accurately predicts the time-scale for an extended satellite to sink from the virial radius of a host halo down to the halo's centre for a wide range of mass ratios and orbits (including a central bulge in each galaxy changes the merging time scale by $\lesssim 10\%$). Their fitting formula is given by

$$\frac{t_{\text{df}}}{t_{\text{dyn}}} = A \frac{\xi^{-b}}{\ln(1 + 1/\xi)} \exp \left[c \frac{j}{j_{\text{cir}}(E)} \right] \left[\frac{r_{\text{cir}}(E)}{r_{\text{vir}}} \right]^d, \quad (5.9)$$

where $A = 0.216$, $b = 1.3$, $c = 1.9$ and $d = 1.0$. Here ξ is the mass ratio $M_{\text{sat}}/M_{\text{host}}$, j is the specific angular momentum of the satellite halo, and j_{cir} is the specific angular momentum of a circular orbit with the same energy E . This formula is expected to be valid for $0.025 \leq \xi \leq 1.0$, and for circularities $0.3 \leq \eta \equiv j/j_{\text{cir}}(E) \leq 1.0$. Most likely value of circularity in dark matter simulations is $\eta \approx 0.5$ [Benson, 2005; Zentner et al., 2005; Khochfar & Burkert, 2006]. Lastly, it is valid for range of orbital energy $-0.65 \leq r_{\text{cir}}(E)/r_{\text{vir}} \leq 1.0$. This covers the peak value of distribution seen in cosmological N-body simulations. We fix $r_{\text{cir}}(E)/r_{\text{vir}} = 1.0$ and $\eta = 0.5$, which are the typical values found in simulations.

We can now obtain the instantaneous merger rate of galaxies by combining the halo merger rate and dynamical friction time scale. We closely follow the method of Shen [2009] and write

$$B_{\text{gal}}(M, \xi, z) = B[M, \xi, z_e(z, \xi)] \frac{dz_e}{dz}, \quad (5.10)$$

where $B(M, \xi, z)$ (per unit volume per unit mass per unit redshift per unit mass ratio) is the instantaneous merger rate of halos with mass M , progenitors with mass ratio ξ at

redshift z , B_{gal} is the same quantity for galaxies. The redshift $z_e(z, \xi)$ is a function of z and ξ , and is given implicitly by

$$t(z) - t(z_e) = t_{\text{mrg}}(\xi, z_e), \quad (5.11)$$

where $t(z)$ is the cosmic time at redshift z . Shen [2009] finds that dz_e/dz is almost constant at all redshifts for $\xi = 0.1 - 1$ and can be approximated by

$$\frac{dz_e}{dz} \approx 1 + 0.09[\xi^{1.3} \ln(1 + 1/\xi)]^{-1}, \quad (5.12)$$

for the fitting formula in equation (5.9). We assume this form in our calculations. Once we have calculated $B_{\text{gal}}(M, \xi, z)$, we normalize it by $n(M, z)$, the abundance of haloes of mass M at redshift z . We use the Sheth-Tormen mass function [Sheth & Tormen, 1999] to calculate $n(M, z)$. This gives us the galaxy merger rate *per halo* per unit ξ per unit redshift, which is the galaxy's counterpart of equation (5.3), and which we denote by dN_{gal}/dz . The rate of mergers of galaxies is the rate at which new black holes are added to the host halo's nucleus. Thus, the time scale of incoming black holes is

$$t_{\text{in}} = \left[\frac{dN_{\text{gal}}}{dz} \frac{dz}{dt} \right]^{-1}. \quad (5.13)$$

The result is shown by the solid black line in Figure 5.2 for a mass ratio of $\xi = 0.4$ and a halo that has mass of $10^{12} M_{\odot}$ at $z = 0$.

5.2.2 Binary SMBH coalescence time scale

In order to find whether there is a generic possibility of formation of systems with multiple SMBHs, we compare the time scale on which new black holes are added to the galactic nucleus at a certain redshift with the coalescence time scale of a binary SMBH at that redshift.

As described above, the formation and coalescence of a black hole binary is expected to take place in three stages. We define the coalescence time as the time that the binary spends in the second of these stages, that is the time from when the binary separation is $a = a_h$, defined in equation (5.1), up to when the separation is $a = a_{\text{gr}}$ at which point the binary enters the third stage of evolution, and gravitational waves become the dominant mechanism of energy loss. For a hard binary, the dominant channel through which energy is lost is three-body interactions in which stars passing in close proximity to the binary are ejected at a much higher velocity $v_{\text{ej}} = [GM_{\text{tot}}/a]^{1/2}$, where M_{tot} is the total mass of the binary. The hardening time scale was quantified for a fixed stellar distribution by Quinlan [1996], who found a time scale of

$$t_h(a) \equiv \left| \frac{a}{\dot{a}} \right| = \frac{\sigma}{G\rho a H}, \quad (5.14)$$

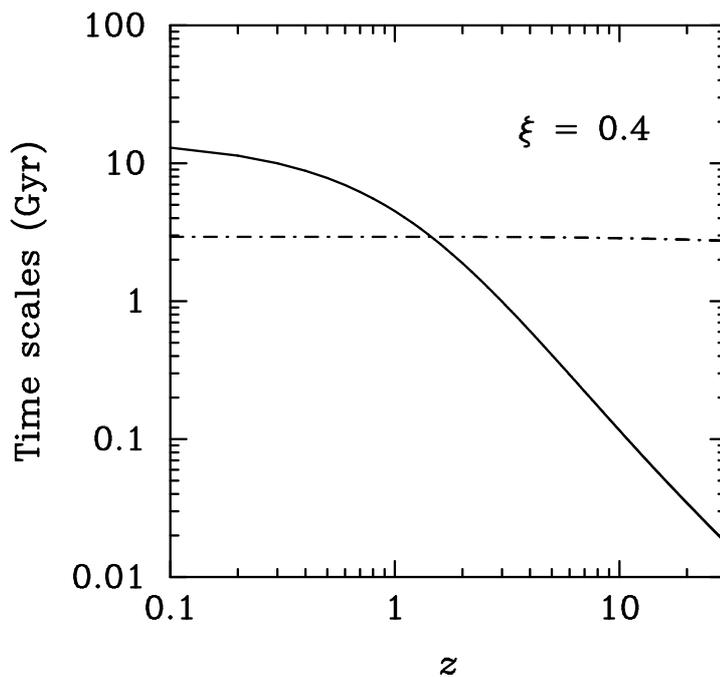


Figure 5.2: A comparison between the feeding time scale of incoming black holes t_{in} (black solid line; Eq. 5.13) and the time scale of black hole coalescence t_{coal} (black dot-dashed line; Eq. 5.23), for a halo mass $M_0 = 10^{12} M_{\odot}$ and considering only mergers with a mass ratio $\xi = 0.4$. The coalescence time t_{coal} has only a weak dependence on redshift because its dependence on M_{bh} and σ cancel out due to the $M_{\text{bh}}-\sigma$ relation. This figure shows that at high redshift new black holes would arrive to the center of a galaxy faster than they could merge via dynamical processes.

where a is the binary separation, ρ and σ are the density and one-dimensional velocity dispersion of the stellar background, and H is a dimensional parameter whose value was found from scattering experiments to be 16 for a hard, equal-mass binary. In practice, however, the above expression for t_h is valid only during the initial stages of the binary's evolution. As the binary shrinks further, it ejects stellar mass from the central regions and modifies the stellar density ρ that appears in equation (5.14). This feedback can be quantified using a simple analytical model given by Merritt [2000], in which the binary evolution is described by two coupled equations, the first describing the binary's hardening due to the presence of stars,

$$\frac{d}{dt} \left(\frac{1}{a} \right) = H \frac{G\rho}{\sigma}, \quad (5.15)$$

and the second describing the change in stellar density due to ejection of mass by the hard SMBH binary,

$$\frac{dM_{\text{ej}}}{d \ln(1/a)} = JM_{\text{tot}}, \quad (5.16)$$

where M_{ej} is the ejected mass, and J is another dimensionless parameter that was measured by Quinlan [1996] to be close to unity and nearly independent of a .

By assuming a singular isothermal sphere profile for the stellar density and assuming that the ejected stellar mass causes a constant-density core to form at the center of this profile, Merritt [2000] finds that evolution of the binary separation can be described as

$$\frac{t - t_{\text{init}}}{t_0} = \frac{a_h}{a} \left[\ln^2 \left(\frac{a_h}{a} \right) - 2 \ln \left(\frac{a_h}{a} \right) + 2 \left(1 - \frac{a}{a_h} \right) \right], \quad (5.17)$$

where a_h is as defined in Equation (5.1), $a(t_{\text{init}}) = a_h$, and t_0 is given by

$$t_0 = \frac{9\pi J^2}{H} \left(\frac{M_{\text{tot}}}{2m_2} \right) \left(\frac{GM_{\text{tot}}}{\sigma^3} \right). \quad (5.18)$$

This result is found to closely match with the evolution observed in N-body simulations.

On the other hand, the timescale for emission of gravitational waves is given by

$$t_{\text{gr}} = \frac{5}{256} \frac{c^5 a^4}{G^3 m_1 m_2 M_{\text{tot}}}. \quad (5.19)$$

As a result, the binary will continue to harden only up to the time when hardening time $t_h = t_{\text{gr}}$, after which it will coalesce rapidly due to gravitational wave emission. Using equation (5.17), it can be shown that this occurs when $a = a_{\text{gr}}$ where [Merritt, 2000],

$$\frac{a_{\text{gr}}}{a_h} \approx A |\ln A|^{0.4}, \quad (5.20)$$

and

$$A = 9.85 \left(\frac{m_1}{m_2} \right)^{0.2} \left(\frac{M_{\text{tot}}}{2m_2} \right)^{0.4} \left(\frac{\sigma}{c} \right). \quad (5.21)$$

Here m_1 and m_2 are masses of the components of the SMBH binary. Finally, we can again use equation (5.17) to calculate the time it takes for the binary to shrink from $a = a_h$ to $a = a_{\text{gr}}$ [Merritt, 2000]:

$$t_{\text{coal}} \approx 8t_0 A^{-1} |\ln A|^{8/5}, \quad (5.22)$$

which can be simplified as

$$t_{\text{coal}} \approx 1.4 \times 10^{10} \text{yT} \left(\frac{m_2}{m_1} \right)^{0.2} \left(\frac{M_{\text{tot}}}{2m_2} \right)^{0.6} \left(\frac{M_{\text{tot}}}{10^9 M_\odot} \right) \left(\frac{\sigma}{200 \text{km/s}} \right)^{-4}. \quad (5.23)$$

Clearly, there is a possibility for the formation of multiple-SMBH system if $t_{\text{in}} < t_{\text{coal}}$. These two time scales are compared in Figure 5.2 for a halo that has a mass of $M_0 = 10^{12} M_\odot$ at $z = 0$. For simplicity, we have fixed the mass ratio of merging haloes to be $\xi = 0.4$. At each redshift, we calculate t_{in} from equation (5.13). In order to estimate t_{coal} at a given redshift using equation (5.23), we first infer the mass of the halo at that redshift from the fitting function for the halo's assembly history from equation (5.5). We then assume that a galaxy belonging to a satellite halo with mass ratio ξ has merged with this host halo at this redshift.

In order to estimate the mass of black holes in the nuclei of these galaxies, we follow the approach of Hoffman & Loeb [2007] in employing the $M_{\text{bh}}-\sigma$ relation. The virial velocity (defined as the circular velocity at virial radius) for a halo of mass M at redshift z is given by

$$v_{\text{vir}} = 23.4 \left(\frac{M}{10^8 h^{-1} M_\odot} \right)^{1/3} \left[\frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2} \right]^{1/6} \left(\frac{1+z}{10} \right)^{1/2} \text{km/s}, \quad (5.24)$$

where

$$\Omega_m^z = \frac{\Omega_m (1+z)^3}{\Omega_m (1+z)^3 + \Omega_\Lambda + \Omega_k (1+z)^2}, \quad (5.25)$$

and Δ_c is the overdensity of the halo relative to the critical density, given for the Λ CDM cosmology by

$$\Delta_c = 18\pi^2 + 82d - 39d^2, \quad (5.26)$$

where $d = \Omega_m^z - 1$ [Barkana & Loeb, 2001]. Further, we equate the halo virial velocity with the circular velocity v_c of its constituent spheroid and obtain the velocity dispersion of the spheroid using the relation [Ferrarese, 2002]

$$v_c \approx 314 \left[\frac{\sigma}{208 \text{km/s}} \right]^{0.84} \text{km/s}. \quad (5.27)$$

This combined with the $M_{\text{bh}}-\sigma$ relation [Tremaine et al., 2002]

$$\frac{\sigma}{208\text{km/s}} \approx \frac{M_{\text{bh}}}{1.56 \times 10^8 M_{\odot}}^{1/4.02}, \quad (5.28)$$

gives

$$\left(\frac{M_{\text{halo}}}{10^{12} M_{\odot}}\right) = 8.28 \left(\frac{M_{\text{bh}}}{10^8 M_{\odot}}\right) \left[\frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2}\right]^{-1/2} (1+z)^{-3/2}. \quad (5.29)$$

We obtain the black hole masses in the host and the satellite haloes using equation (5.29) and use the spheroid velocity dispersion from equation (5.27) to estimate the coalescence time from equation (5.23). The result is shown by the dashed line in Figure 5.2.

At high redshift, early on in the assembly history of a halo, the galaxy merger rate is higher than the SMBH binary coalescence rate and systems with multiple SMBHs can form. Note that the time scale t_{coal} obtained above will change if effect of loss-cone replenishment and gas are taken into account. However, Yu [2002] finds that in realistic spheroidal galaxies, even loss-cone replenishment is insufficient to cause early coalescence.

5.3 Evolution of multiple SMBHs

We have described the literature on systems with more than two SMBHs in §5.1. If the infalling SMBH is less massive than either of the components of a pre-existing binary then we expect the ultimate outcome to be ejection of the smaller SMBH and recoil of the binary. Hoffman & Loeb [2007] studied the statistics of close triple SMBH encounters in galactic nuclei by computing a series of three-body orbits with physically motivated initial conditions appropriate for giant elliptical galaxies. Their simulations included a smooth background potential consisting of a stellar bulge and a dark matter halo, and also accounted for the effect of dynamical friction due to stars and dark matter. They found that in most cases the intruder helped the binary SMBH to coalesce via the Kozai-Lidov mechanism and by scattering stars into the binary's loss cone. In this case, the intruder itself was left wandering in the galactic halo, or even kicked out of the galaxy altogether. It was also found that escape of all three black holes is exceedingly rare.

Dynamical evolution of multiple massive black holes in globular clusters has received much attention [Kulkarni, Hut & McMillan, 1993; Sigurdsson & Hernquist, 1993]. From these studies, it is expected that systems with more than two SMBHs will last for about a crossing time.

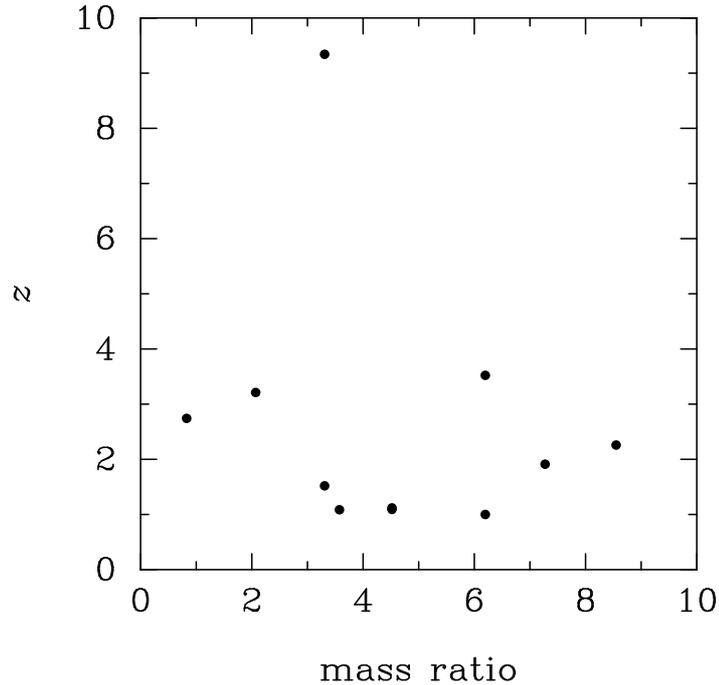


Figure 5.3: An example merger tree from the Millennium simulation of a halo that has a mass of $M_0 \sim 10^{12} M_\odot$. This plot shows major mergers (mass ratio > 0.1) *in all branches* of the halo’s merger tree.

5.4 Simulations

In order to accurately calculate the formation and evolution of galactic nuclei with multiple black holes, we perform direct-summation N-body simulations of galactic nuclei merging in a cosmological context. This essentially involves generating physically consistent initial conditions for galactic nuclei with SMBHs at high redshift and evolving them while taking into account the mergers of such nuclei and the resultant close interaction of their SMBHs.

We obtain merger histories of galactic nuclei by extracting merger trees of gravitationally bound subhaloes from the Millennium Simulation Database³, which stores results of the Millennium Simulation [Springel et al., 2005]. The Millennium Simulation is a pure dark matter simulation with a Λ CDM model with 2160^3 particles in a periodic cube $500 h^{-1}\text{Mpc}$ on a side. This corresponds to a particle mass of $8.6 \times 10^8 h^{-1} M_\odot$. The output of this simulation is stored in 64 snapshots between $z = 127$ and $z = 0$. Particles in each snapshot are grouped into friends-of-friends (FOF) clusters that

³<http://www.mpa-garching.mpg.de/millennium/>

Simulation	Mass of halo at $z = 0$ (M_\odot)	Max. BH no.	SMBH Coalescences	SMBH Escapes
L1	1.21×10^{14}	4	7	2
L2	1.31×10^{14}	2	1	1
L3	1.31×10^{14}	2	3	2
L4	1.24×10^{14}	2	5	5
L5	1.28×10^{14}	5	8	4
L6	1.31×10^{14}	6	6	0
L7	1.23×10^{14}	3	2	0
L8	1.31×10^{14}	2	3	1

Table 5.1: Summary of simulations and results for haloes that have a mass of $M_0 \sim 10^{14} M_\odot$. The maximum BH number denotes the number of black holes in the biggest BH group found in a simulation. The last two columns show number of BH coalescences and escapes in the simulation. A halo with $M_0 = 10^{14} M_\odot$ has average mass $M_{z=6} = 5 \times 10^{11} M_\odot$.

Simulation	Mass of halo at $z = 0$ (M_\odot)	Max. BH no.	SMBH Coalescences	SMBH Escapes
H1	1.25×10^{15}	6	4	3
1.65×10^{15}	2	1	1	
1.81×10^{15}	3	2	0	
1.24×10^{15}	5	6	3	
1.37×10^{15}	3	7	1	
1.40×10^{15}	4	3	0	
1.41×10^{15}	6	9	1	
1.45×10^{15}	3	4	1	
1.46×10^{15}	2	2	0	
1.48×10^{15}	4	7	1	
1.54×10^{15}	2	1	1	
1.59×10^{15}	5	10	1	
1.66×10^{15}	8	15	4	
1.71×10^{15}	4	3	0	
1.81×10^{15}	4	20	7	
1.86×10^{15}	3	7	4	
4.04×10^{15}	8	11	2	

Table 5.2: Summary of simulation runs with haloes that have mass $M_0 \gtrsim 10^{15} M_\odot$ at $z = 0$. Various columns are same as Table 5.1. A halo with $M_0 = 10^{15} M_\odot$ has average mass $M_{z=6} = 5 \times 10^{12} M_\odot$.

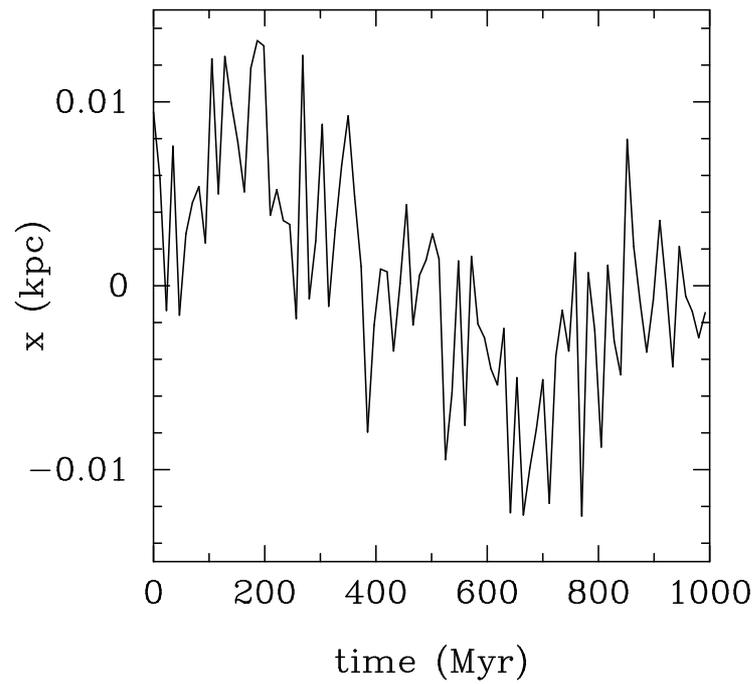


Figure 5.4: Evolution of a single SMBH in our simulations. The left hand panel shows evolution of the x -component of the position of a $9.95 \times 10^5 M_{\odot}$ back hole near the centre of a Hernquist bulge of mass $5.41 \times 10^7 M_{\odot}$ and scale length of 0.2 kpc. The particle mass is $5.411 \times 10^3 M_{\odot}$. The secular motion is due to that of the cusp.

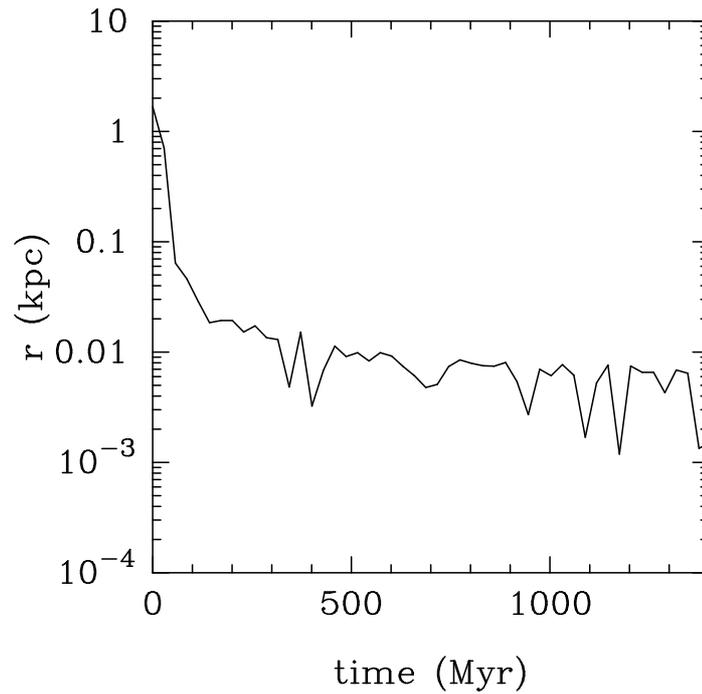


Figure 5.5: Evolution of a binary SMBH in our simulations. The right hand panel shows evolution of the separation between SMBHs in a binary with initial separation 2 kpc and eccentricity 0.5. The black hole masses were $8.65 \times 10^4 M_{\odot}$ and the binary evolved near the center of a Hernquist halo with mass $5.41 \times 10^7 M_{\odot}$ and scale length of 10.0 kpc. The particle mass is $5.411 \times 10^3 M_{\odot}$.

are expected to correspond to virialised structures. Each FOF halo contains substructure of gravitationally bound subhaloes that can be related to each other across snapshots as progenitors and descendants. Because a halo can contain multiple galaxies, we expect the subhalo merger tree to reflect the merger history of the galaxies within a halo. Since our goal in this chapter is to understand formation and evolution of systems of multiple black holes due to the hierarchical merger history of a galaxy, we extract subhalo merger trees from the Millennium Simulation Database. Each such merger tree typically shows growth of a subhalo via accretion of dark matter particles and via mergers. We process these merger trees to keep only major mergers, which we define to be mergers having mass ratio larger than 0.1. To identify the mass ratio of two subhaloes, we use the masses of the distinct FOF haloes that these subhaloes were a part of before the FOF haloes merged. This is to account for the mass loss of the satellite subhalo due to tidal stripping after it enters the FOF group of the host subhalo, but before the eventual merger of the two subhaloes. (See discussion in §5 of Bundy, Treu & Ellis 2007.) Figure 5.4 shows the resultant merger history of a Milky Way sized halo. The main reason behind removing minor mergers from our calculation is that for such mergers the dynamical friction time taken by the satellite halo to reach the center of the host halo is longer than the Hubble time. As a result, in such mergers, we do not expect the constituent galactic nuclei of these haloes to interact closely. Since, as we describe below, we model only the spheroidal galactic nuclei in our simulations, we only need to account for mergers in which such nuclei will closely interact. This approach is very similar to that used by Li et al. [2007a], with the main difference being our use of direct-summation N-body simulations instead of SPH simulations.

Once we have a galaxy merger tree, we set up the initial conditions of our simulation in the “leaves” of the tree, that is, in haloes that do not have a progenitor, and follow the evolution using an N-body calculation. The initial conditions of our simulation consist of a stellar spheroid with a Hernquist density profile,

$$\rho(r) = \frac{M}{2\pi} \frac{a}{r(r+a)^3}, \quad (5.30)$$

where M is the total mass of the spheroid and the scale length a is related to the half mass radius $r_{1/2}$ of the spheroid by $a = 0.414r_{1/2}$. Values for the parameters M and a were obtained from the halo mass as follows [Hoffman & Loeb, 2007]. We first obtain the black hole mass M_{bh} from the halo mass M_{halo} using Equation (5.29). We then use the empirical relation between the SMBH mass and the spheroid’s virial mass [Magorrian et al., 1998; Marconi & Hunt, 2003; Peng et al., 2006] to obtain the latter as

$$M_{\text{sph}} = 4.06 \times 10^{10} M_{\odot} \left[\frac{M_{\text{bh}}}{10^8 M_{\odot}} \right]^{1.04}. \quad (5.31)$$

The virial mass of the spheroid is related to its velocity dispersion σ_e and half light

radius R_e by

$$M_{\text{sph}} = \frac{kR_e\sigma_e^2}{G}. \quad (5.32)$$

We follow Marconi & Hunt [2003] and set $k = 3$ to get an average ratio of unity between this mass estimate and the dynamically measured masses of galaxies. The velocity dispersion in the above equation is usually measured over either a circular aperture of radius $R_e/8$ or a linear aperture of length R_e . These two methods are in essential agreement, as argued by Tremaine et al. [2002]. Assuming a constant mass-to-light ratio for the Hernquist profile, we have $R_e = 1.815a$ and the velocity dispersion at radius $R_e/8$ is given by

$$\sigma_e^2 = \frac{0.104GM}{a}. \quad (5.33)$$

This lets us obtain the value of the parameter M of the Hernquist profile as $M = 1.765M_{\text{sph}}$. The scale length a is readily obtained as

$$a = \frac{GM_{\text{sph}}}{3\kappa_1\sigma_{\text{bh}}^2}, \quad (5.34)$$

where σ_{bh} is obtained using the $M - \sigma$ relation of equation (5.28). Having obtained a density profile for the bulge, we place a black hole at its center and set the black hole mass to be ten times that obtained from equation (5.29). This factor of ten is introduced to keep the ratio between the black hole mass and the particle mass high enough [Milosavljević & Merritt, 2001; Makino & Ebisuzaki, 1996]. We confirm that the radius of influence $r_{\text{inf}} = Gm_{\text{bh}}/\sigma^2$ of this black hole is still much smaller than a . Velocities of the stars in the spheroid are then generated from the unique, isotropic velocity distribution that corresponds to the gravitational potential of the density profile in Equation (5.30) and the SMBH [Tremaine et al., 2002]. These initial conditions are then scaled to standard N-body units of $G = 1$, $M = 1$ and $E = -0.25$, where M is the total mass of the system and E is its total energy [Heggie & Mathieu, 1986; Aarseth, 2003]. In these units, in virial equilibrium, the mean square velocity $\langle v^2 \rangle = 1/2$ and the system's crossing time is $t_{\text{cr}} = 2\sqrt{2}$, independent of the number of particles. The conversion factors from physical units to these N-body units can be easily obtained via dimensional analysis.

Note that we ignore presence of gas in this set-up. Simulations of binary BHs in gaseous environment have not reached sufficient resolution to establish the role played by gas in evolution of SMBHs in galactic nuclei [Merritt & Milosavljević, 2005; Colpi & Dotti, 2009]. Moreover, we expect that at high redshifts, quasar activity triggered by galaxy mergers could efficiently drive gas away from the shallow potential wells of the galaxies.

To investigate the actual dynamical evolution of this system, we use the direct-summation code NBODY6 written by Sverre Aarseth [Aarseth, 1999, 2003]. This code

has been well-tested for various applications since around 1992. Its purpose is to perform an exact integration, without particle softening, of a large number of particles. It integrates equations of motion of individual particles using a fourth-order Hermite method with block time steps [Makino & Aarseth, 1992]. This integrator is coupled with the Ahmed-Cohen neighbour scheme [Ahmad & Cohen, 1973], which selects a subset of neighbours of a particle whose forces on it are calculated at a higher time resolution than other, more distant, particles. This scheme reduces the computational cost from $\mathcal{O}(N^2)$ to about $\mathcal{O}(N^{1.6})$. Close two-body encounters are treated using the Kustaanheimo-Stiefel (KS) regularization method that eliminates the $r = 0$ singularity in Newtonian gravity by using a coordinate transformation. Triples, quadruples and compact subsystems of up to six particles (called “chains”) are treated using the chain regularization method [Mikkola & Aarseth, 1990]. Details of the various algorithms in this code and their implementation are given by Aarseth [2003]. In all simulations reported in this chapter, the time-step parameter for irregular force polynomial, η_I , and the time-step parameter for regular force polynomial, η_R are set to 0.02. The energy tolerance is set to $Q_E = 4 \times 10^{-5}$ and the regularized time-step parameter is set to $\eta_U = 0.2$.

We check the stability of our initial conditions by evolving standalone realizations of the Hernquist bulge with a central BH and then traverse the merger tree of a given halo using NBODY6, starting from the initial conditions as described above. We scale the physical time between two successive nodes of the tree to N-body units and run NBODY6 for that duration. If a merger happens at a certain node, we place the two galactic nuclei at a distance of 2 kpc apart and evolve in an head-on approach. Although such head-on mergers would be unlikely, we choose it to reduce the computational time while still retaining some realism. When two galaxies, that are in equilibrium separately, merge we expect some transient response in the resulting dynamics. However, as discussed by Milosavljević & Merritt [2001], any such effects in the dynamics of the central regions of the merger remnant of these galaxies are essentially negligible.

Under these conditions, the component black holes approach after a merger event and the remnant galactic nucleus is left with two black holes, which gradually harden due to dynamical friction and three-body interactions with stars in their vicinity. Black hole coalescence is implemented in our simulation by monitoring the separation of hard black hole binaries. Once members of a SMBH binary get closer than a fixed distance d_{crit} , we replace them with a single black hole with mass equal to the sum of the masses of component black holes. In all the runs reported in this chapter, we set $d_{\text{crit}} = 0.1$ pc. Note that this is the only mechanism in which black holes grow in our simulations. Thus, the initial SMBH masses are set according to the $M - \sigma$ relation, but the later growth of these SMBHs occurs only via coalescence.

Recoil due to anisotropic emission of gravitational waves is a natural consequence of asymmetric merger of black holes, either due to unequal masses or due to unequal

spins [Peres, 1962; Bekenstein, 1973]. Until recently, it was unclear whether this recoil is large enough to be astrophysically relevant. However, recent results from numerical relativity have revealed the resultant kick velocities in a variety of merger configurations [Pretorius, 2005; Baker et al., 2006a]. When the black hole spins are aligned with each other and with the orbital spin, these simulations find recoil velocity of $v_{\text{recoil}} \lesssim 200 \text{ km s}^{-1}$ [Baker et al., 2006b; González et al., 2007; Herrmann et al., 2007; Lousto & Zlochower, 2009]. In the absence of spins, this recoil velocity is only a function of the ratio of black hole masses. For random orientations of spins, recoil velocities as high as 2000 km s^{-1} have been obtained [Campanelli et al., 2007a,b]. Bogdanović, Reynolds & Miller [2007] argue that a circumbinary gas disk can align the binary spins with the orbital axis thereby reducing v_{recoil} to about 200 km s^{-1} . In our simulations we assume a constant kick velocity of 200 km s^{-1} , which we impart to the remnant of every unequal-mass binary SMBH coalescence.

We follow the approach of Makino & Aarseth [1992] and keep the particle number fixed at $N = 10^4$ throughout the simulation. Thus, at every merger, we combine particles in each merging galactic nucleus and double the particle mass. This lets us keep the particle number high throughout the merger tree of the halo. The ratio of black hole mass to the stellar mass is typically a few hundred, which is also roughly the ratio of the spheroid's total mass to the black hole's mass. These values are comparable to other simulations of this kind [Makino & Ebisuzaki, 1996; Milosavljević & Merritt, 2001].

In summary, the unique features of our simulations are: *(i)* kinematically consistent initial conditions with black holes; *(ii)* calculation of mergers of galactic nuclei in a cosmological setting using merger trees extracted from cosmological N-body simulations; *(iii)* calculation of merger of galactic nuclei resulting in a formation of SMBH binaries starting from the results of each nucleus having evolved in isolation; and *(iv)* accurate calculation of SMBH-star and SMBH-SMBH dynamics throughout the assembly history of a galactic nucleus and its constituent SMBH with the effect of gravitational wave recoil taken into account.

5.5 Results

We perform some basic checks on our code, such as ensuring energy conservation and stable evolution of equilibrium systems. In all of our runs, the relative error in the total energy is maintained at $|\Delta E/E| < 4 \times 10^{-5}$. The treatment of BH-BH and BH-star interaction is handled by the original NBODY6 code, and is expected to be accurate. One caveat here is that the neighbour criterion in NBODY6 for regularization of close particles is based on inter-particle distance. As a result, while evolving a set of particles in the vicinity of a massive BH, the code either selects a large number of particles for chain regularization, or selects every close pair of particles for two-body regularization. This usually slows down the code. Indeed, in three of our runs the code run time ex-

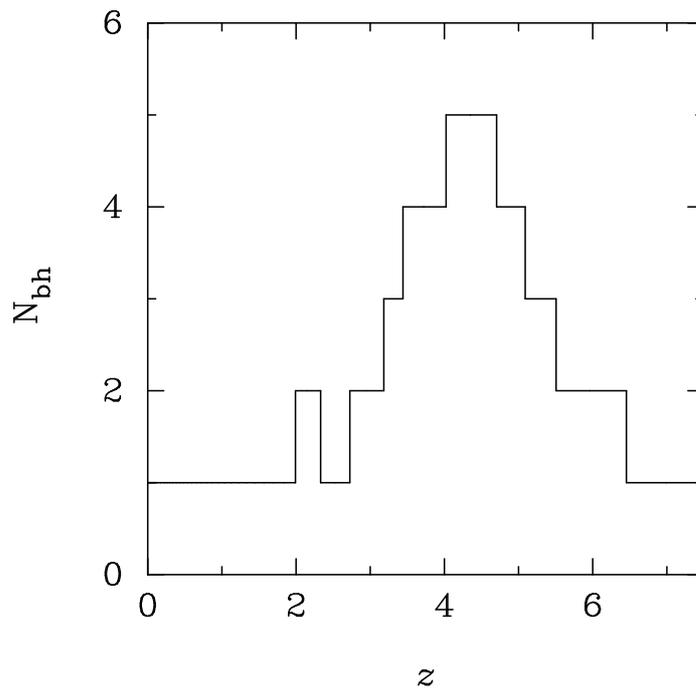


Figure 5.6: Number of black holes as a function of redshift in a simulation with $M_0 = 1.29 \times 10^{14} M_\odot$.

ceeded practical constraints because of this effect. These three runs are excluded from the results presented below.

5.5.1 Dynamics of single and binary SMBHs

In a stellar environment, a single SMBH exhibits a random fluctuating motion arising due to discrete interactions with individual stars. As a result, the effect of the stellar environment on the SMBH can be decomposed into two distinct components: (1) a smooth component arising due to the large scale distribution of the whole system, and (2) a stochastic fluctuating part coming from the interaction with individual stars [Chatterjee, Hernquist & Loeb, 2002]. This random motion is illustrated in the left hand panel of Figure 5.4, which shows evolution of the x -component of the position of a $9.95 \times 10^5 M_\odot$ black hole near the centre of a Hernquist bulge of mass $5.41 \times 10^7 M_\odot$ and scale length of 0.2 kpc. The particle mass is $5.411 \times 10^3 M_\odot$. As expected, the SMBH wanders around due to stochastic interactions with the stars in its vicinity. The mean square amplitude of these fluctuations is expected to be [Chatterjee, Hernquist & Loeb, 2002;

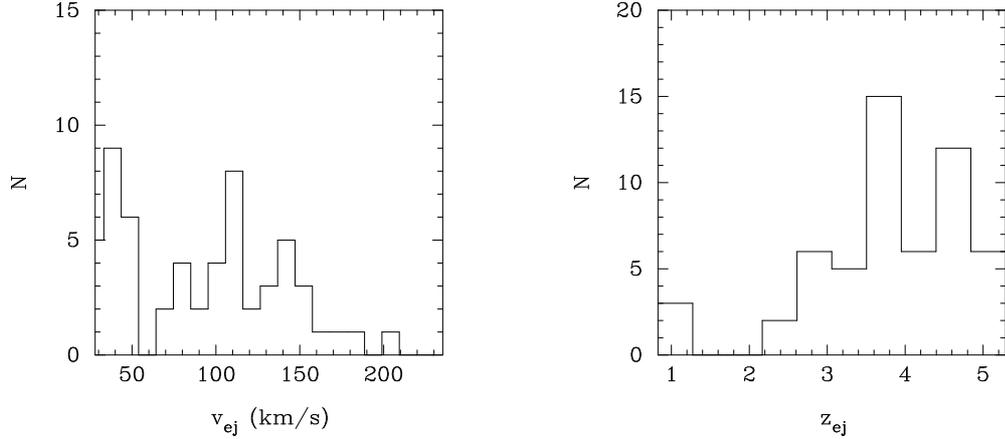


Figure 5.7: Histograms of ejection velocities of BHs. Left: Velocities of ejected black holes in all of our high mass runs. Note that this does not include ejected black holes with the highest velocities ($> 2000 \text{ km s}^{-1}$). Right: number of ejections as a function of redshift in our high mass runs.

Milosavljević & Merritt, 2003a]

$$\langle x^2 \rangle \approx \frac{m_*}{m_{\text{BH}}} r_{\text{core}}^2, \quad (5.35)$$

where r_{core} is the radius within which the stellar distribution flattens out. The Hernquist distribution that we have used here does not have a well-defined core, since the density keeps rising as r^{-1} near the origin. Milosavljević & Merritt [2003a] argue that the effective core radius for such distribution can be taken as the radius of influence of the black hole. The resultant mean square value of fluctuations is somewhat smaller than that for Figure 5.4 by roughly a factor of 2 as is known to happen in N-body simulations [Quinlan & Hernquist, 1997; Milosavljević & Merritt, 2003a].

As described above, the evolution of a binary black hole in a gas-poor galaxy takes place in three stages. Right hand panel of Figure 5.4 shows evolution of the separation between SMBHs in a binary with initial separation 2 kpc and eccentricity 0.5 in our code. The black hole masses were $8.65 \times 10^4 M_{\odot}$ and the binary evolved near the center of a Hernquist halo with mass $5.41 \times 10^7 M_{\odot}$ and scale length of 10.0 kpc. The particle mass is $5.411 \times 10^3 M_{\odot}$. In the first stage of evolution, the SMBHs sink to the centre of the galactic nucleus by losing energy via dynamical friction and become bound to each other. This stage ends when the separation between the SMBHs is equal to the radius of influence of the binary [Merritt & Milosavljević, 2005]. In the second evolutionary stage, the binary loses energy predominantly ejection of nearby stars via three-body interaction. The binary loses energy rapidly in this stage, which continues until $t \approx 200$ Myr for the case depicted in Figure 5.4. The final stage of the SMBH binary evolution

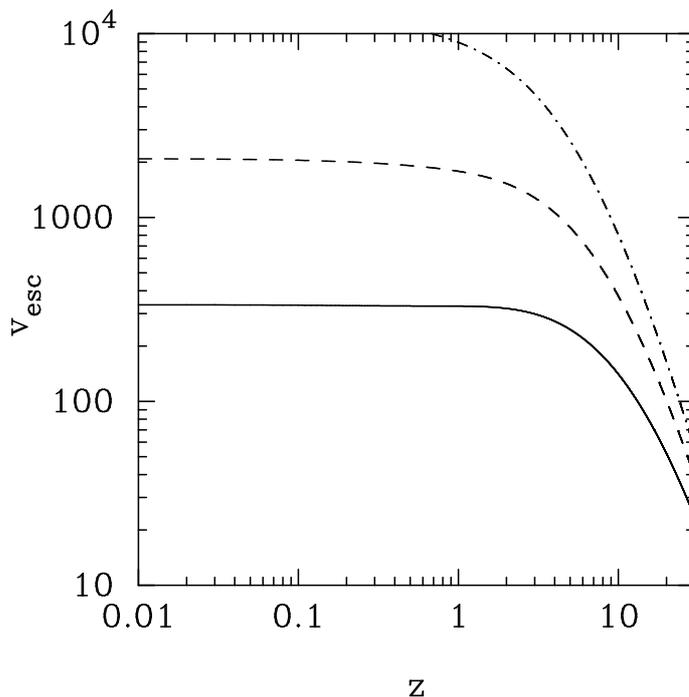


Figure 5.8: Escape velocities from the bulges of haloes in our three categories of present-day masses of haloes. Solid line: $M_0 \approx 10^{12} M_\odot$, Dashed line: $M_0 \approx 10^{14} M_\odot$, Dot-dashed line: $M_0 \gtrsim 10^{15} M_\odot$. Note that these are average values computed from the fitting functions to the Millennium simulation. Therefore, case by case comparison with our runs is not straightforward.

begins when the rapid hardening of the second stage stops. This happens when the binary semi-major axis takes the value given by Equation (5.1). The binary semi-major axis is related to the separation r by

$$\frac{1}{a} = \frac{2}{r} - \frac{v^2}{\mu}, \quad (5.36)$$

where v is the relative velocity of the BHs and μ is the reduced mass [Makino & Funato, 2004; Berczik et al., 2006; Merritt, Mikkola & Szell, 2007; Khan, Just & Merritt, 2011]. In N -body simulations, the last stage is known to have a dependence on the number of particles N such that the hardening rate decreases with increasing N [Makino & Funato, 2004]. For real spherical galaxies, the binary separation would stop evolving after this point because the loss cone is empty.

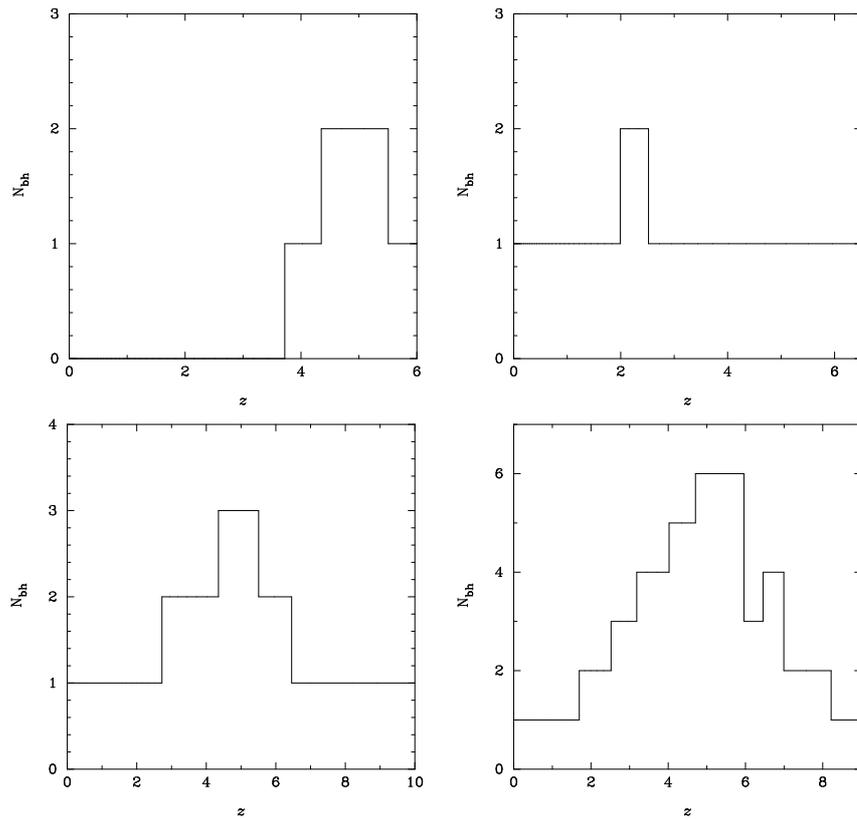


Figure 5.9: Number of black holes as a function of redshift in a few of our simulation runs.

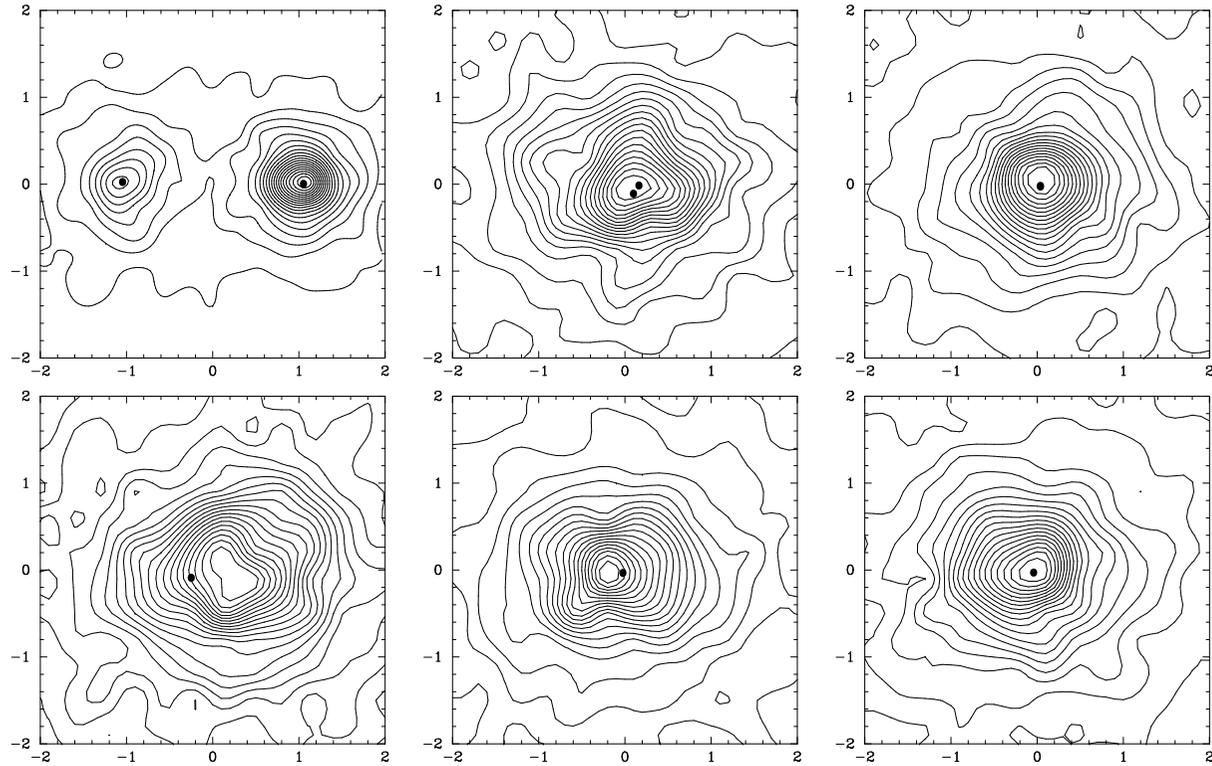


Figure 5.10: Projected stellar density contours in the presence of a binary in the simulation H5. Each panel is 400 pc on a side. Clockwise from top left to bottom right, the redshifts are $z = 10.073, 8.54, 7.27, 6.19, 5.28,$ and 4.52 . The total time span is about 800 Myr. Core-SMBH oscillations are clearly visible.

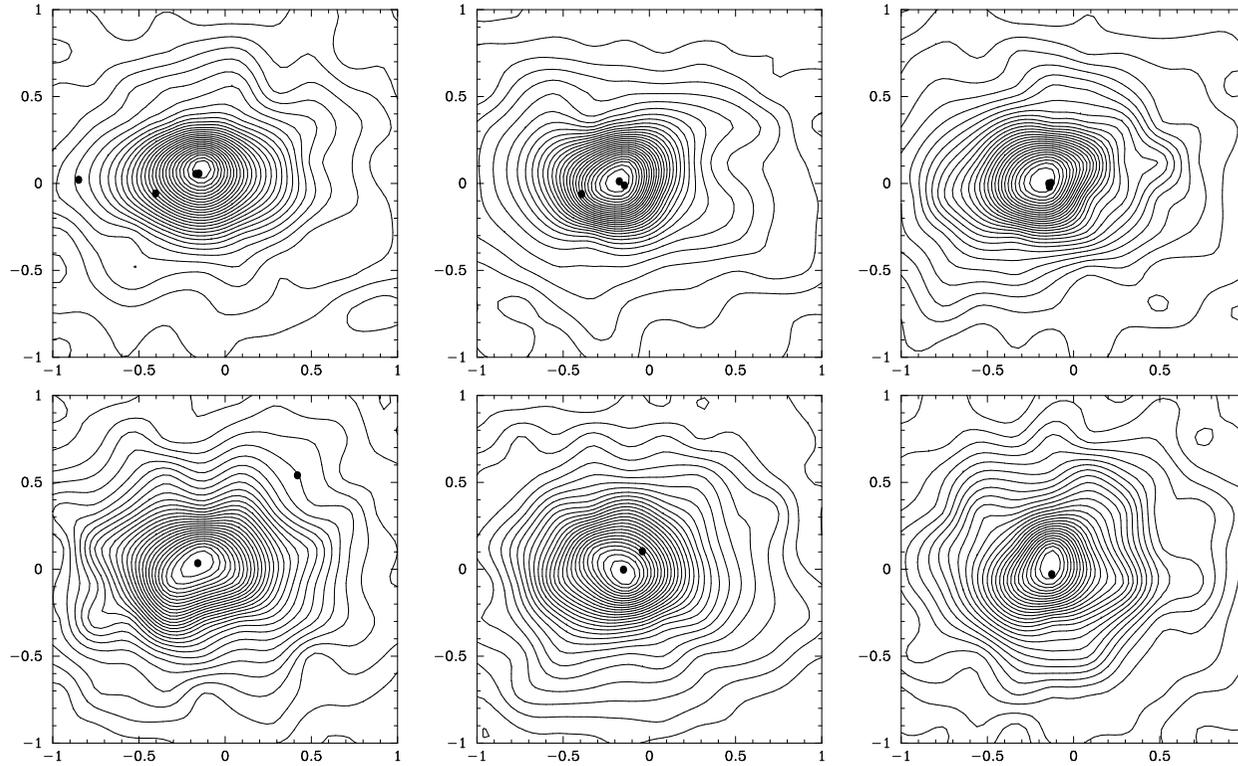


Figure 5.11: Projected stellar density contours in the presence of multiple BHs in the simulation H4. Each panel is 100 pc on a side. The total time span, clockwise from top left to bottom right, is about 1 Gyr. Most BHs are stripped of their cusps in nuclei with multiple BHs.

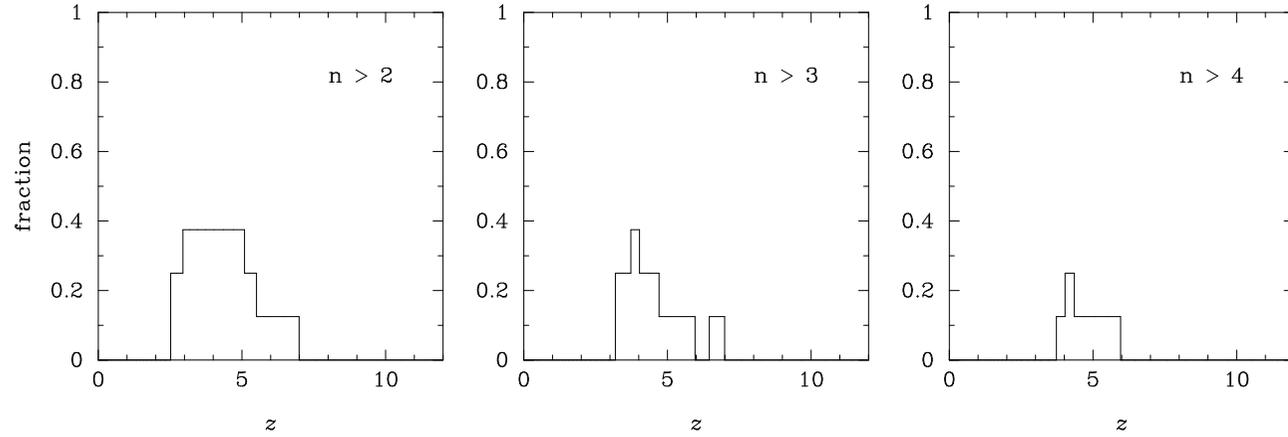


Figure 5.12: The fraction of runs with multiple SMBHs at different redshift bins for haloes with a mass $M_0 \sim 10^{14}M_\odot$ at $z = 0$. The results of these runs are summarised in Table 5.1. The three panels from left to right describe the occurrence of systems with more than 2, 3 and 4 black holes respectively. At each redshift, this number can be interpreted as the likelihood of finding such systems in haloes of mass $M_0 \sim 10^{14}M_\odot$ at $z = 0$. It is seen that systems with multiple SMBHs are rare at redshift $z \lesssim 2$. Note that a halo with $M_0 = 10^{14} M_\odot$ will have $M_{z=6} = 5 \times 10^{11} M_\odot$.

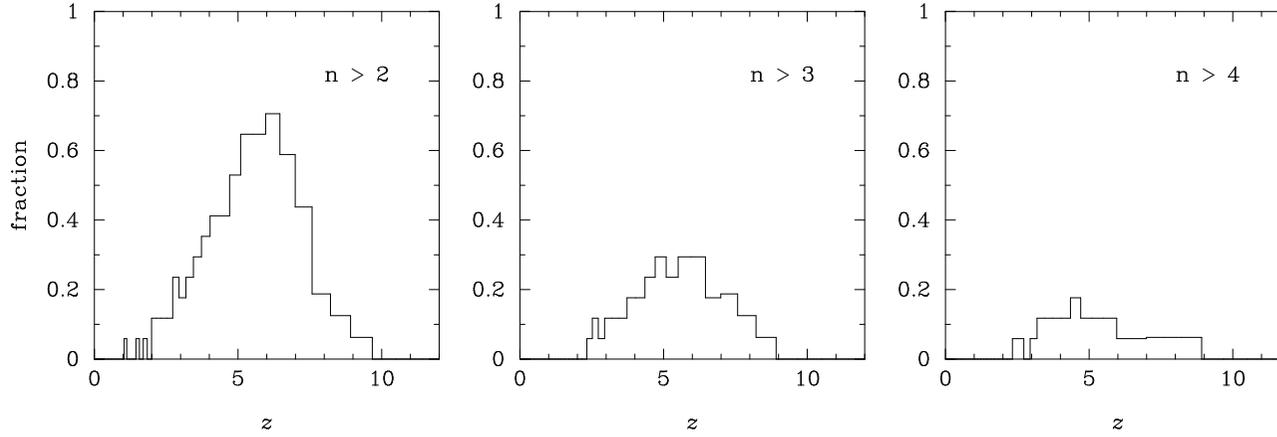


Figure 5.13: The fraction of runs with multiple SMBHs at different redshift bins for halo masses $M_0 \gtrsim 10^{15} M_\odot$ at $z = 0$. The results of these runs are summarised in Table 5.2. The three panels from left to right describe the occurrence of systems with more than 2, 3 and 4 black holes respectively. At each redshift, this number can be interpreted as the likelihood of finding such systems in haloes of mass $M_0 \gtrsim 10^{15} M_\odot$ at $z = 0$. It is seen that systems with multiple SMBHs are rare at redshift $z \lesssim 2$. These results can be compared with those in figure 5.12. Nuclei with multiple SMBHs are more likely in high mass haloes because of higher merger rate. Note that a halo with $M_0 = 10^{15} M_\odot$ will have $M_{z=6} \sim 10^{12} M_\odot$.

5.5.2 Evolution of nuclei with multiple SMBHs

We now run the simulation along merger trees of haloes drawn from the Millennium simulation as described in Section 5.4. These simulations are described in Tables 5.1 and 5.2. We randomly select 8 haloes with mass M_0 around $10^{14} M_\odot$ at $z = 0$. These correspond to the typical haloes ($M \approx M_*$) in the present epoch. We also randomly select 17 haloes whose present-day mass M_0 is in excess of $10^{15} M_\odot$. These are rare, high mass haloes that are expected to host the redshift 6 SDSS quasars [Li et al., 2007a]. Additionally, we have also simulated 11 haloes with present-day mass similar to the Milky Way halo ($M_0 \sim 10^{12} M_\odot$). Using the prescriptions described in the previous section, and using the N-body integrator, these simulations tell us about the effect of multiple mergers of galactic nuclei with SMBHs.

Figure 5.6 shows results from a typical simulation run, for a halo of mass $1.29 \times 10^{14} M_\odot$. We plot here the number of BHs in the bulge in the main branch of the galaxy's merger tree at various redshifts. It is seen that the central bulge has more than one SMBH for a wide redshift range ($2 \lesssim z \lesssim 6$; about 2.5 Gyr). For $3 \lesssim z \lesssim 5$ (about 1 Gyr) the bulge holds more than 2 BHs. The maximum number of BHs interacting within the bulge in this simulation is 6. Lastly, the number of BHs reduces to one well before $z = 0$ due to coalescences and ejections. Note that at the highest redshifts ($z \gtrsim 6$) there are no BHs in the central bulge. This is simply an artifact of the limited numerical resolution of the Millennium simulation, because of which the halo merger tree is not resolved at these redshifts. To ensure that this does not affect our results for $z \lesssim 6$, we set up initial conditions at $z \sim 6$ such that the BHs are on the $M - \sigma$ relation, and by using a Hernquist bulge with inner slope -1 . In the absence of gas, the systems with multiple SMBHs form generically, in high mass haloes with frequent of major mergers. It is evident than such systems are usually short-lived and most often these nuclei contain a single SMBH at $z = 0$. Most SMBHs escape into the halo, where they join a population of wandering black holes or escape the halo completely.

Similar results from a few other simulation runs for haloes with mass $M_0 \sim 10^{14} M_\odot$ at $z = 0$ are shown in Figure 5.9. Most of these runs have features similar to the run described above. Multiple BH systems form generically and last for 2 – 3 Gyr. Importantly, most of these galaxies end up with a single SMBH in their central bulge. This is in contrast with expectations from some simple arguments in earlier work [Hut & Rees, 1992]. About 5% of galaxies in our simulations end up with no BHs in their centres at $z = 0$. Tables 5.1 and 5.2 summarize these features of all our simulations. The last columns of these tables show the cumulative number of BHs that were ejected out of the galactic nucleus throughout the run either due to recoil associated with emission of gravitational waves or due to many-body interaction between the BHs. We find that for most triple and quadruple SMBH systems in our calculation, gravitational wave recoil is the dominant mechanism for SMBH escape. Many-body interaction between SMBHs was the dominant cause only when the number of black holes was more than

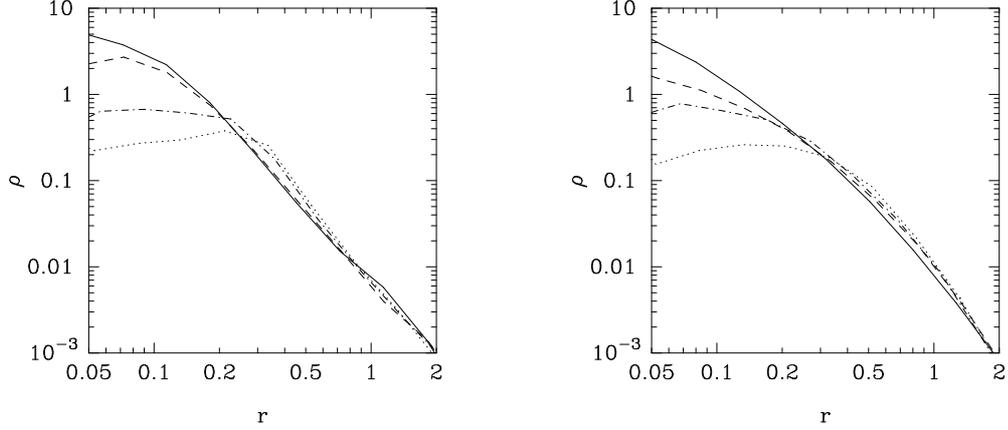


Figure 5.14: Evolution of density profile for simulations H3 and H5 in N-body units. The solid line is the original Hernquist profile with an inner logarithmic slope of $\gamma \approx -1$. Dashed line shows the profile after one SMBH binary coalescence, dot-dashed line after the second coalescence and the dotted line after the third coalescence. These plots are shown in N-body units to scale out the doubling of the half-mass radius. See text for details.

four. Consequently, for low-mass galaxies in which the number of BHs is small, almost all escapes were because of gravitational wave recoil. Whereas in our low mass galaxy simulations, larger number of coalescence usually results in large escapers, in the high mass galaxy simulations, coalescence often does not lead to escape. In high mass galaxies, BH-BH interaction is the dominant mechanism behind escaping SMBHs. Figure 5.5 summarizes this. The right hand panel shows that most ejections happen at high redshifts. Typical ejection velocities are seen in the left hand panel. Ejection velocities are spread out up to 200 km s^{-1} , which is the GW recoil kick in our simulations. Note that this plot does not show kicks with very high velocities, which we describe below.

With the prescription that we have adopted in this chapter, we find that SMBH coalescence happens in each one of our simulations. Tables 5.1 and 5.2 give the number of BH coalescences occurring in our simulations. Due to the limitation on the particle number, our simulations implement BH coalescence by replacing a bound binary BH by a single BH whose mass is equal to the total mass of the binary. As an example, Figure 5.10 shows the merger of two bulges beginning from initial conditions at redshift 6.7 in the run H5. In Figure 5.10, the hardening radius is $a_h = 0.5 \text{ pc}$ at $t_h = 500 \text{ Myr}$. We find the the BHs remain associated with their host cusps until cusp coalescence. It is known that by increasing the effective mass of the BHs, this increases the rate of coalescence of the BHs by as much as ~ 6 times compared to the dynamical friction time scale. We also see the homology of density structure before and after the merger, as reported previously in the literature [Milosavljević & Merritt, 2001]. However, one

prominent difference from previous works is in the evolution of the density profile in the later stages of the merger. In our simulations, each coalescence event is followed by recoil of the remnant at 200 km s^{-1} , which at high redshift, usually results in the escape of the SMBH from the galaxy. At relatively low redshifts, the recoiled SMBH returns to the nucleus in few hundreds Myr. Because of this recoil, the remnant BH is detached from its cusp immediately. At the recoil speed implemented here, this happens at a much smaller time scale than the local crossing time scale. As a result, the only effect of the remnant on the cusp is due to subsequent core passages.

Usually, most coalescences are assumed to take place due to BH hardening via BH-star encounters. In gas-free systems, this leads to the final parsec problem. In our simulations, we find that in high mass haloes, roughly half of the SMBH coalescences are due to three-body scattering with intruder SMBHs. This is expected, since in spite of higher major merger rate, high mass galaxies in our model are still left with at most two SMBHs at $z = 0$. The dominant mechanism of coalescence is then three body interactions. Figure 5.9 shows an example of the evolution of a multiple BH system that undergoes three coalescences due to BH-BH dynamics. We find violent oscillations of the cusp-BH system as shown in Figure 5.10. This has significant impact on the density distribution of the core, and also results in off-centre BHs, which slowly return to the centre of the cusp due to dynamical friction.

About 10% of SMBH ejections in our simulations occur at very high speeds of $\gtrsim 2000 \text{ km s}^{-1}$. In haloes with $M_0 \approx 10^{15} M_\odot$ these SMBHs will linger in the outskirts of the halo for $2 - 10 \text{ Gyr}$ as can be seen by comparing with the bulge escape speeds in Figure 5.8. The SMBHs in the wandering phase that are introduced via this mechanism have markedly different properties than the BHs introduced due to galaxies that have not yet reached the host galaxy's center so as to have a close encounter [Volonteri, Haardt & Madau, 2003]. The main difference is that our ejected black holes are much more massive than those in the other category. Moreover, the velocity of ejected SMBHs will typically be higher than black holes in the other category, which have already experienced significant dynamical friction. Three of the 30 BH ejections in our runs are ejected binaries.

5.5.3 Likelihood of nuclei with multiple SMBHs at high redshift

From the results of our simulations, we can estimate the likelihood of galactic nuclei with multiple black holes at high redshifts. The histograms in Figures 5.12 and 5.13 show fraction of runs with multiple SMBHs at each redshift for haloes with present-day masses of $\sim 10^{14} M_\odot$ and $\sim 10^{15} M_\odot$, respectively. The three panels from left to right describe the occurrence of systems with more than 2, 3 and 4 black holes respectively. At each redshift, this number can be interpreted as the likelihood of occurrence of such systems at that redshift.

Systems with more than 2 SMBHs are generically expected in the central galaxies

of haloes with $M_0 \gtrsim 10^{14} M_\odot$ at around $z \gtrsim 3$. On the other hand, few galaxies hold multiple black holes at redshifts $z \lesssim 2$ because the galaxy merger rate is low at these redshifts and the BHs have sufficient time to coalescence. This is consistent with the expectation from our heuristic analysis of Section 5.2. In other words, multiple black hole systems are numerous at around redshifts of 6, when there are many major mergers in the system. Our numerical simulations show that such systems can exist in sufficiently long-lived configurations of SMBHs separated on pc–kpc scale. Note that these histograms show the likelihood of such systems to be zero at redshifts $z \gtrsim 10$. However, this is simply because the Millennium simulation merger trees do not resolve progenitors at these redshifts. As mentioned before, we have minimized the effect of this shortcoming on our results by requiring that the SMBHs always follow the $M - \sigma$ relation initially.

High mass galaxies ($M_0 \approx 10^{15} M_\odot$) are more likely to have multiple BHs in their nuclei at higher redshift. About 60% of these galaxies have more than 2 BHs between redshifts $z \approx 2$ and 10. This fraction is less than 40% for the low mass galaxies ($M_0 \approx 10^{14} M_\odot$). The likelihood of occurrence of more than 3 and 4 BHs is similar, about 30%, in the two categories of simulation. However, for the high mass galaxies this likelihood is spread out over a wider range in redshift, again due to the higher rate of major mergers.

It is extremely rare for Milky Way-sized galaxies (halo mass $M_0 \approx 10^{12} M_\odot$) to have more than three SMBHs in their nuclei at any moment in their assembly history. Indeed, in our simulations of these galaxies, only one run shows a triple BH system. The main reason behind this is the smaller number of major mergers for these galaxies. Moreover, it is easier for SMBHs to escape the nuclei of predominantly small mass progenitors of these galaxies.

5.5.4 Effects on the stellar distribution

Most bulges and early-type galaxies have a shallow cusp near their centre. The mass distribution in this region can be described as a power law $\rho \propto r^{-\gamma}$. Most galaxies have slope $0.5 \lesssim \gamma \lesssim 2.0$ [Ferrarese et al., 2006; Merritt & Szell, 2006]. We expect the constituent SMBH in the bulge to affect the mass distribution within its radius of influence. Only two galaxies, the Milky Way [Genzel et al., 2003] and M32 [Lauer et al., 1998], have been resolved at these small distances. Both these galaxies have $\gamma \approx 1.5$ in their innermost regions.

It is commonly postulated that cores can form in elliptical galaxies and spiral bulges due to mass ejection by a hard binary SMBH (e.g. Milosavljević & Merritt 2001). However, the mass ejected by a hard binary is of the order of the black hole mass. In other words, the mass deficiency M_{def} , which is the difference between the mass of the initial and final density distribution in a region around the centre, is roughly M_{bh} , the total mass of the SMBH binary. The possibility of enhanced mass deficit because of repeated

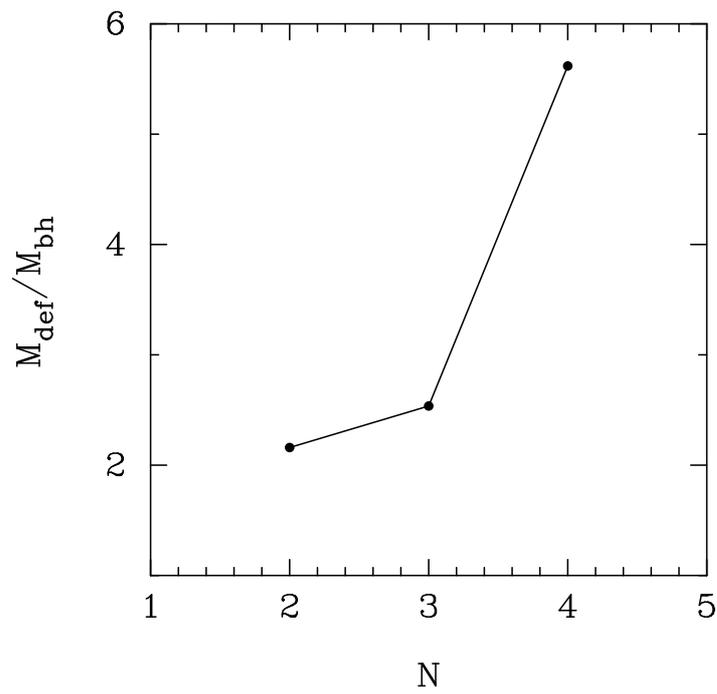


Figure 5.15: Mass deficiency versus number of coalescences averaged over ten simulation runs. The presence of multiple SMBHs generally leads to larger mass deficiency compared to a single hard SMBH binary.

core passages of recoiled black holes [Gualandris & Merritt, 2008] and due to repeated mergers [Merritt, 2006] has been considered in the literature. Our simulations allow us to understand the effect of both of these factors in addition to the mass deficit produced by simultaneous presence of multiple SMBH in the galactic bulge.

Figure 5.5.2 shows the cusp evolution in two of our simulations, each of which has four SMBHs and three coalescences. Density profiles after each coalescence is shown. Strong core formation is clearly seen. We calculate $M_{\text{def}}/M_{\text{bh}}$ for ten such runs and show the average result in Figure. Clearly $M_{\text{def}}/M_{\text{bh}}$ is much larger when multiple SMBHs are present. Values of $M_{\text{def}}/M_{\text{bh}} \approx 5$ have been observed in large elliptical galaxies [Graham, 2004; Ferrarese et al., 2006; Hopkins & Hernquist, 2010]. Our model explains the occurrence of such systems. Since the star-star relaxation time in large elliptical galaxies is expected to be $\sim 10^{10}$ yr, we can expect them to carry the signature of core formation at high redshift due to multiple SMBHs. At lower redshift our simulations are applicable to spiral bulges, which have a much lower relaxation time scale ($\sim 10^9$ yr). Indeed in the runs where a single black hole is left for $z \lesssim 2$, we find the formation of a Bahcall-Wolf cusp. This is consistent with the observed structure of the Milky Way bulge.

The above considerations regarding cores in galaxy luminosity profile are also applicable to dark matter cores. The ejection of dark matter particles by the black holes will produce a core similar in size to the stellar core.

5.6 Observational Signatures

From the results of our simulations described above, we expect about 30% of the galaxies within haloes with a present-day mass of $M_0 \approx 10^{14} M_{\odot}$ to contain more than two SMBHs at redshifts $2 \lesssim z \lesssim 6$. For more massive haloes with $M_0 \gtrsim 10^{15} M_{\odot}$, this fraction is almost 60%. However, since few such systems have been unambiguously observed so far, we consider some observational signatures that would indicate their existence⁴. Apart from their effect on the stellar mass distribution, multiple SMBH systems lead to an enhanced rate of tidal disruption of stars, modified gravitational wave signals compared to isolated BH binaries, and slingshot ejection of SMBHs from galaxies at high speeds.

From the results of scattering experiments, Chen et al. [2009] found that the stellar tidal disruption rates due to three-body interactions between a hard, unequal-mass SMBH binary with fixed separation and a bound stellar cusp is higher by several orders

⁴Some systems with triple active galactic nuclei (AGNs) were reported so far. Examples are NGC 6166 and 7720 [Tonry, 1984] and SDSSJ1027+1749 [Liu, Shen & Strauss, 2011]. The first two objects are cD galaxies at $z \approx 0.03$ and the latter is at $z \approx 0.06$. All three are kpc-scale triples. It is possible that NGC 6166 is simply a superposition of a central cD galaxy and two low-luminosity elliptical galaxies [Lauer et al., 1998].

of magnitude than the corresponding rates for a single SMBH. In particular, they find that the stellar tidal disruption rate is about 1 yr^{-1} for an isothermal stellar cusp with $\sigma = 100 \text{ km s}^{-1}$ containing an SMBH binary of total mass $10^7 M_{\odot}$. In comparison, the corresponding rate for a single $10^7 M_{\odot}$ black hole is about 10^{-4} yr^{-1} . The duration of the tidal disruption phase is about 10^5 yr . This enhancement in the tidal disruption is due to the Kozai-Lidov effect and due to chaotic resonant scattering [Chen et al., 2011]. Tidal disruption of a star results in about half of the stellar mass being inserted in bound elliptical orbits. When it falls back in the black hole, this mass gives rise to a bright UV/X-ray emission (“tidal flare”) lasting for a few years. One such event may have already been recently observed in the form of high-energy transients that can be modeled as sudden accretion events onto an SMBH [Levan et al., 2011; Bloom et al., 2011; Zauderer et al., 2011].

We expect similar enhancement in the rate of stellar tidal disruption in systems with multiple black holes. Firstly, the presence of multiple SMBHs increases the combined tidal disruption cross section of the black holes. (Although this will only enhance the tidal disruption rate by a factor of a few.) Secondly, even before they closely interact, the presence of a third SMBH affects the tidal disruption event rate onto an SMBH binary by scattering stars into the binary’s loss cone at a rate that increases as inverse square of its separation from the binary [Hoffman & Loeb, 2007]. Thirdly, as we saw above, multiple SMBH systems are likely to contain recoiled black holes, which have been kicked either due to anisotropic gravitational wave emission after coalescence, or due to the gravitational slingshot. Sudden recoil promptly fills the loss cone of these black holes. The resultant enhancement in the tidal disruption event rate can be substantial, increasing it up to 0.1 yr^{-1} [Stone & Loeb, 2011]. Furthermore, if their recoil velocity is not too high, these recoiled SMBHs oscillate around the stellar core with decreasing amplitude due to dynamical friction. This motion results in their repeated passages through the stellar core, thereby increasing the stellar tidal disruption event rate.

Another observational signature of systems with multiple SMBHs is gravitational waves (GWs). The GW emission from binary and triple SMBHs has been studied in the literature [Wyithe & Loeb, 2003a; Sesana et al., 2004; Amaro-Seoane et al., 2010]. Space-based detectors like the Laser Interferometer Space Antenna (LISA) are expected to be sensitive in the frequency range $\sim 10^{-4}$ – 10^{-1} Hz . This corresponds to the inspiral of SMBH systems with total mass $\sim 10^4 - 10^{10} M_{\odot}$. Pulsar timing arrays (PTAs) like the Parkes PTA [Manchester, 2008] and the European PTA [Janssen et al., 2008] and ground-based detectors like the North American Nanohertz Observatory for Gravitational Waves [Jenet et al., 2009] are sensitive to even lower frequencies of $\sim 10^{-8}$ – 10^{-6} Hz .

Yunes, Coleman Miller & Thornburg [2011] studied modifications due to the presence of a secondary SMBH in the waveform of an extreme mass-ratio inspiral (EMRI) of a stellar mass objects into an SMBH. They find that a $10^6 M_{\odot}$ SMBH will produce

detectable modifications if it is within a few tenths of a parsec from the EMRI system, although this distance increases for higher mass SMBHs. In this chapter, we have quantified the presence of such ‘massive perturbers.’ The resultant modifications to gravitational waveforms will be a distinct signature of multiple-SMBH systems. Furthermore, such systems often contain binaries that have phases of very high eccentricities, created via mechanisms like the Kozai-Lidov effect [Hoffman & Loeb, 2007]. Such binaries are expected to emit intense bursts of high-frequency gravitational waves at the orbital periapsis [Amaro-Seoane et al., 2010]. As a result, sources that would normally emit outside of the frequency windows of planned gravitational wave searches may be shifted into observable range. For example, Amaro-Seoane et al. [2010] find that a few to a hundred gravitational wave bursts could be produced at a detectable (1 ns) level within the PTA frequency range if the fraction of SMBH triplets is ≥ 0.1 .

Presence of triple SMBHs also has important implications for gravitational wave searches using matched-filtering by possibly requiring additional waveform templates [Amaro-Seoane & Freitag, 2011].

Lastly, an observable signature of these systems will be the presence of wandering SMBHs in the large haloes [Hoffman & Loeb, 2007]. We have shown that about 10% of the SMBHs are ejected at velocities $> 2000 \text{ km s}^{-1}$ due to the slingshot mechanism. This high-speed black holes will spend $1 - 10 \text{ Gyr}$ in the outskirts of the halo. However, it is not clear whether detecting this population of wandering black holes will be possible.

5.7 Conclusions

In this chapter, we have addressed the formation of galactic nuclei with multiple SMBHs. We performed accurate N-body simulations of mergers of galactic nuclei with SMBHs in a cosmological setting. Our calculation uniquely incorporated cosmological mergers of galaxies with an accurate treatment of dynamical interactions between SMBHs and stars, which we achieved using the direct summation N-body code, NBODY6. The need for such simulations has been recognized in the literature [Merritt & Milosavljević, 2005]. Our main conclusions are as follows:

- In the absence of gas, high mass galaxies ($M_0 \gtrsim 10^{14} M_\odot$ at $z = 0$) are generically expected to have had multiple SMBHs in their nuclei during their assembly history. Our simulations suggest that $\sim 30\%$ galaxies within haloes with a present-day mass of $M_0 \approx 10^{14} M_\odot$ ($M_{z=6} \approx 10^{11} M_\odot$) contain more than two SMBHs at redshifts $2 \lesssim z \lesssim 6$. For more massive haloes, with $M_0 \gtrsim 10^{15} M_\odot$ ($M_{z=6} \approx 10^{12} M_\odot$), this fraction is almost 60%. This is in contrast to lower-mass galaxies ($M_0 \approx 10^{12} M_\odot$; $M_{z=6} \approx 10^{10} M_\odot$), which rarely host more than two SMBHs in their nuclei at any moment in their assembly history.

- High mass galaxies as well as their low mass counterparts are rarely expected to retain more than two SMBHs in their nuclei at the present epoch. SMBH coalescence and ejection reduces the number of SMBHs on the time scale of a Gyr. Furthermore, major mergers are rare at lower redshift. We also find that the number of SMBHs in galactic nuclei is rarely reduced to zero at $z = 0$. Less than 5% of our high-mass runs resulted in such galaxies.
- SMBH coalescence is common at high redshifts. Subsequent recoil due to anisotropic gravitational wave emission often results in escaping SMBHs. Some of these SMBHs add to the wandering population of black holes in the galactic halo. In a few cases, this process also results in galactic nuclei with no SMBH near their centres. BH-BH interaction also leads to ejected SMBHs via the slingshot mechanism. While most of ejected SMBHs have velocities $\lesssim 500 \text{ km s}^{-1}$, about 10% SMBHs are ejected at very high velocities exceeding 2000 km s^{-1} . We also find binary SMBH ejection in $\lesssim 10\%$ of the cases.
- Multiple SMBHs have a strong effect on the stellar distribution due to three-body interactions and core passages. The resulting mass deficit is usually much larger than that due to a single SMBH binary because of resonant BH-BH interactions and GW recoil of the BH remnant. We observe long-term oscillations of the BH-core system that could explain observations of offset AGNs. This has implications for recent observations by Civano et al. [2010] of a $z = 0.359$ system that potentially contains a recoiled BH.
- The presence of multiple SMBHs will have important effects on the rate of tidal disruption of stars in galactic nuclei due to enhanced tidal disruption cross section, scattering of stars by other BHs, prompt loss cone refilling due to GW recoil and gravitational slingshot. Similarly, the presence of more than two BHs in a hierarchical triple is expected to leave a signature in the GW emission from the inner binary. This signature could be observed with future GW observatories, such as LISA. Finally, we also expect such systems to give rise to a distinct population of wandering SMBHs that could travel in large haloes over long time scales of a few Gyrs.

The presence of gas could alter the above picture to some extent. However, simulations of binary BHs in gaseous environment have not reached sufficient resolution to confirm this. Moreover, we expect that at high redshifts, AGN activity triggered by galaxy mergers could efficiently drive gas away from the shallow potential wells of the galaxy. Our work can also be extended by calculating late stages of binary SMBH evolution more consistently. New regularization techniques to do this are now available [Aarseth, 2003]. Furthermore, multiple SMBH systems can also form in additional ways, for example by fragmentation of disks [Goodman & Tan, 2004]. However, these

systems would evolve by migration [Kocsis, Yunes & Loeb, 2011] on a much shorter time scale than considered here.

Post-reionization cosmological H I distribution in a hierarchical galaxy formation model

In previous chapters, we studied the evolution of the neutral fraction of IGM at high redshifts. We saw that the predictions of a model, in which the neutral fraction evolves from unity at redshifts $z \sim 20$ to less than 10^{-4} at $z \sim 6$ due to the build-up of a UV background sourced by galaxies, agree well with a variety of observations. In this chapter, we extend this study of neutral gas in the universe to post-reionization redshifts ($z < 6$). At these redshifts, as we will discuss below, cold, neutral gas predominantly resides in the ISM of galaxies. As the raw material for star formation, such gas—mainly consisting of neutral hydrogen, molecular hydrogen, and helium—plays an important role in the evolution of galaxies. Yet little is known about how much cold gas there is in the universe at these redshifts and how it is distributed in haloes of different mass. These are the two governing questions of this chapter. Thus, this study complements our work on the average evolution of galaxies and IGM presented before by modelling properties of galaxies, like their metallicity and luminosity, in more detail, and by moving beyond the global average to understand the spatial distribution of cold gas.

Our knowledge of the cold gas content of the local universe mainly comes from radio observations from the rest-frame 21 cm line corresponding to a hyperfine transition in the Hydrogen atom. (We discussed properties of this transition in Chapter 1.) Observations of the 21 cm line allows us to measure the density, temperature, and velocity dispersion of HI along our line of sight. The H I Parkes All Sky Survey (HIPASS) attempts a comprehensive census of H I in the local universe [Barnes et al., 2001; Zwaan et al., 2003, 2005]. This survey revealed that the galaxy population seen in H I is essentially the same as that seen in optical and infrared, but weighted towards gas-rich systems, which are mainly late type [Zwaan et al., 2005]. It showed that the mass function of

H I in the local universe (the number density of galaxies with a given H I mass per unit comoving volume) can be described well by a Schechter function. Lastly, it also constrained the cosmological density of H I in the local universe to $\Omega_{\text{HI}} \sim 3.5 \times 10^{-4} h_{75}^{-1}$ (assuming a dimensionless Hubble parameter $h_{75} = h/0.75 = 1$). This is approximately one-tenth the value of the cosmic stellar mass density at $z = 0$. Beyond these advances, HIPASS is being extended to lower mass galaxies by the ALFALFA survey [Giovanelli et al., 2005]. In their first results, the updated Schechter fit to the H I mass function obtained by this survey was found to predict an order of magnitude more galaxies at the high mass end than HIPASS. Also, their estimate of Ω_{HI} is 16% larger than HIPASS [Martin et al., 2010].

Nonetheless, these advances in observing H I are confined to the local universe ($z \lesssim 0.5$). Detection of the 21cm line at higher redshifts requires too great a sensitivity for reasonable observing time.¹ As a result, we know little about H I at high redshifts. At these redshifts, our main window on H I is the absorption of the Ly- α line by H I clouds along the line of sight to distant quasars [e.g. Lanzetta et al., 1991, Wolfe et al., 1995, Storrie-Lombardi, Irwin & McMahon, 1996, Péroux et al., 2005, and Wolfe, Gawiser & Prochaska, 2005]. These observations show that most of the neutral hydrogen at higher redshifts resides in relatively rare damped Ly- α systems (DLAs). By making a quantitative estimate of the total H I content in DLAs, they further indicate that at $1 \leq z \leq 5$, the neutral hydrogen content of the universe is almost constant with a density parameter of $\Omega_{\text{HI}} \sim 0.001$. It has been proposed that the 21cm absorption of H I illuminated by background sources will act as an unbiased probe of DLAs [Kanekar & Briggs, 2004; Kanekar et al., 2009]. For example, a detection of 21cm absorption and molecular hydrogen absorption lines in the same DLA has been made [Srianand et al., 2010]. However, the role of 21cm in detecting H I at high redshifts is decidedly limited.

However, this situation is expected to change dramatically over the next decade, with the application of existing facilities like the Giant Metrewave Radio Telescope [GMRT; Swarup, 1984] and the Ooty Radio Telescope [ORT; Prasad & Subrahmanya, 2011], and construction of newer ones such as the Murchison Widefield Array (MWA), the Low Frequency Array [LOFAR], the Australian Square Kilometer Array Pathfinder [ASKAP; Johnston et al., 2008], the South African Square Kilometer Array Pathfinder [MeerKAT; Booth et al., 2009] and the Square Kilometer Array [SKA; Schilizzi, Dewdney & Lazio 2008] itself. The GMRT can observe redshifted 21cm emission from a few selected redshift windows whereas other instruments have continuous coverage over a range of redshifts. Using ASKAP, the WALLABY survey will reach a median redshift of $z \sim 0.05$, compared with $z \sim 0.009$ reached by HIPASS [Johnston et al., 2008]. The SKA will have sufficient sensitivity and angular resolution to map H I in galaxies up to

¹Lah et al. [2009] have used a stacking technique to co-add signal from multiple optically-selected galaxies to reach $z = 0.34$. See Zwaan [2000]; Chengalur, Braun & Wieringa [2001]; Lah et al. [2007]; Khandai et al. [2011].

redshift of $z \gtrsim 3$ [Blake et al., 2004].

Clearly, results from these next generation H I surveys will have profound impact on our understanding of galaxy formation and evolution. For this reason, it is important to understand what the currently favoured galaxy formation models tell us about H I in the high redshift universe. For example, these models can tell us how Ω_{HI} evolves. They can also predict the H I mass function at different redshifts and in different environments. These predictions will potentially help us understand the physical significance of observational data. They can also provide important input into the design of new telescopes.

Modelling H I at high redshifts is difficult, and several approaches have been followed. These fall into three categories: (1) empirical modelling, which relies upon observations of H I, (2) numerical modelling, in which gas dynamical cosmological simulations are used to model H I content of galaxies from first principles, and (3) semi-analytic modelling, in which physically motivated analytic prescription are used with results of numerical simulations of large scale structure to understand H I. Despite the lack of observational data, empirical modelling has made some progress by making assumptions about the evolution of the H I mass function over a broad range of redshifts [Abdalla & Rawlings, 2005; Abdalla, Blake & Rawlings, 2010]. This extrapolation is constrained by the total H I density, which can be compared with observations of DLAs. The empirical approach does not predict clustering of H I sources. Numerical modelling incorporating gas dynamics is computationally expensive and typically restricted to small volumes. This makes it impossible to accurately follow the growth of structure up to $z = 0$. The H I mass function obtained by these studies is in poor agreement with HIPASS observations [Popping et al., 2009]. Clustering predictions are limited to scales smaller than a few Mpc. Of course, gas dynamic simulations still do not have enough resolution to follow various processes involved in galaxy formation, and themselves resort to semi-analytical rules to treat sub-resolution physics.

As a result, we are left with the possibility of using semi-analytic modelling of galaxy formation, which is currently the most favoured route to making predictions for H I [Baugh, 2006; Power, Baugh & Lacey, 2010; Kim et al., 2011]. This category of models treats various astrophysical processes involved in galaxy formation with the help of simple but physically motivated prescriptions. At any given time, the amount of cold gas in a galaxy is determined by the rates of various processes that deplete and replenish it. These processes are implemented halo-by-halo on the results of gravity-only cosmological N-body simulations. The greatest advantage of this approach is its speed. These calculations can cover haloes in a large cosmological volumes in a short time, thereby allowing a detailed study of the parameter space. Baugh et al. [2004] first presented predictions for the mass function of cold gas in galaxies using one particular semi-analytic model. They later compared their results with three other models, which differed from their original models in the details of various physical prescriptions [Power, Baugh &

Lacey, 2010]. Despite the differing prescriptions, the predictions of all these models show generic features. For example, it is found that there is surprisingly little variation in the H I mass function in the post-reionization universe. As another example of this class of models, Kim et al. [2011] studied the halo occupation distribution of H I galaxies to find that satellite galaxies make little contribution to the abundance or clustering of cold-gas-selected samples, and are far less important than they are in optically selected samples.

In this chapter, we use a calibrated semi-analytic model of galaxy formation to study the distribution of H I in the post-reionization universe. Several attempts have been made in recent years to model the H I distribution in the post-reionization universe [Scott & Rees, 1990; Subramanian & Padmanabhan, 1993; Kumar, Padmanabhan & Subramanian, 1995; Bagla, Nath & Padmanabhan, 1997; Bharadwaj, Nath & Sethi, 2001; Bharadwaj & Sethi, 2001; Bagla & White, 2003; Bharadwaj & Srikant, 2004; Bharadwaj & Ali, 2005; Loeb & Wyithe, 2008; Wyithe, Loeb & Geil, 2008; Pritchard & Loeb, 2008; Wyithe & Brown, 2010; Bagla, Khandai & Datta, 2010]. We study the relation between large scale H I distribution and galaxy formation by using semi-analytic modelling.

6.1 Semi-analytic model for galaxy formation

In this section, we summarise our implementation of a semi-analytic model of galaxy formation. Elements of this model were developed by White & Frenk [1991]; Kauffmann, White & Guiderdoni [1993]; Kauffmann et al. [1999]; Springel et al. [2001]; De Lucia, Kauffmann & White [2004]; Croton et al. [2006]. See those papers for more general information on semi-analytic techniques.

There are two main parts in the semi-analytic modelling of galaxy formation: (1) the measurement of dark matter merging trees from a sequence of outputs from dark matter-only cosmological N-body simulations, and (2) the implementation of actual semi-recipes for the physics of galaxy formation on top of these merger trees. We now describe both of these in order.

6.1.1 N-body simulations

We use gravity-only simulations run with the TREEPM code [Bagla, 2002; Bagla & Ray, 2003; Khandai & Bagla, 2009]. A set of four simulations with number of particles $N = 512^3$ were run. Details of these are given in Table 6.1.1. Parameters for cosmology and the power spectrum of density fluctuations take values that best fit the WMAP 5-yr data. These values are $\Omega_{\text{nr}} = 0.26$, $\Omega_{\Lambda} = 0.74$, $\Omega_{\text{b}}h^2 = 0.02273$, $n_{\text{s}} = 0.96$ and $\sigma_8 = 0.79$ [Komatsu et al., 2009].

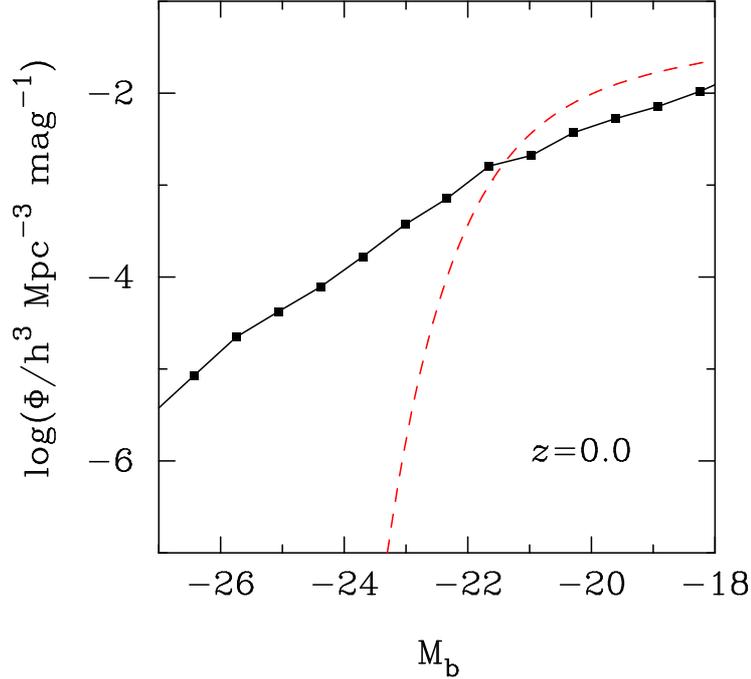


Figure 6.1: The B -band luminosity function in our simulations is shown by the black solid line with square symbols. The red dashed curve is the best-fit Schechter function to the B -band luminosity function from the 2dFGRS [Norberg et al., 2002].

Snapshot outputs, in which the positions and velocities of all particles is stored, are drawn from these simulations at various redshifts. Table 6.1.1 lists the number of snapshots drawn from each simulation. For every simulation output, we compile a list of virialised dark matter haloes using the friend-of-friends (FOF) algorithm [Davis et al., 1985]. We adopt a linking length of $l = 0.2$ times the mean inter-particle separation. This procedure yields several catalogues of FOF haloes for each simulation, one catalogue per snapshot. Only haloes containing at least 15 particles are included in these catalogues as these are shown to be stable. Thus the lowest luminosity galaxy in our $153.6 \text{ h}^{-1} \text{ Mpc}$ simulation has mass $\sim 10^{11} \text{ h}^{-1} \text{ M}_{\odot}$. Since the Milky Way has a mass of $10^{12} \text{ h}^{-1} \text{ M}_{\odot}$ in these models, the lowest bright galaxy is about tenth as bright. This is comparable to the Large Magellanic Cloud, and similar to the resolution used by Kauffmann et al. [1999]. Mass resolution is much better for the other three simulations, as shown in Table 6.1.1, but those simulations cannot be run up to $z = 0$ due to finite box effects [Bagla & Prasad, 2006; Bagla, Prasad & Khandai, 2009]. At $z \simeq 0$, we need a simulation box with $L_{\text{box}} \geq 140 \text{ h}^{-1} \text{ Mpc}$ for the finite box size effects to be negligible [Bagla & Ray, 2005; Bagla & Prasad, 2006]. The mass resolution, on the other hand, decreases as the cube of simulation volume. Using different simulations to study

the H I distribution at different redshifts lets us balance the requirements of high mass resolution and a sufficiently large box.

Once the halo catalogues are ready, we relate haloes in each snapshot to their progenitors in the previous snapshot and to their descendants in the subsequent snapshot. At this stage, we choose to define a progenitor of a halo as simply a halo in the previous snapshot that has contributed at least one particle to it. (We implement more stringent criteria for progenitor-descendant relationship later, when we implement the galaxy formation model.) Accretion of dark matter particles that do not originate in a halo, but are obtained from the diffuse inter-halo medium is also measured. About 5% of our haloes undergo fragmentation [Fakhouri & Ma, 2009], that is, have more than one descendants. This is an artifact of the FOF algorithm; we ignore all such haloes from our analysis.

Along with the merger trees, this stage of modelling also involves computing a set of physical quantities for each halo in the halo catalogues. We compute

- Index of the *central particle* of each halo, which is the most-bound particle in the halo.
- The *virial mass* of the halo, M_{vir} , which is given by the number of particles in each halo times the particle mass of the simulation.
- The *virial radius* of the halo, R_{vir} , which is given by $(GM_{\text{vir}}/100H^2)^{1/3}$.
- The *circular velocity* of the halo at virial radius, given by $V_c = (GM_{\text{vir}}/R_{\text{vir}})^{1/2}$.

We will describe the use of this information in the next subsection. Note that we do not include subhaloes in our analysis [Springel et al., 2001]. It is known that including sub-haloes remarkably improves the luminosity function obtained from the semi-analytic model. We will return to this point when we derive the galaxy luminosity function from our model.

6.1.2 Galaxy formation

We now implement the semi-analytic model of galaxy formation, by supplementing the catalogues of dark matter haloes with the notion of a galaxy population with physical properties given by semi-analytic prescriptions. In this picture, each dark matter halo contains exactly one ‘central galaxy,’ whose position is given by the central particle of the halo. Apart from the central galaxy, a halo can also have one or more ‘satellite’ galaxies, where the positions of each one of them is given by one of the particles of the halo. Satellite galaxies are galaxies that had been central galaxies themselves in the past, but their haloes have merged at some previous time with the larger halo that they now reside in. Satellite galaxies orbit in their halo and are assumed to merge with the central galaxy on a dynamical friction time-scale.

L_{box}	N_{part}	m_{part}	z^f	n_{snap}
23.04	512^3	6.7×10^6	5.0	24
51.20	512^3	7.0×10^7	3.0	29
76.80	512^3	2.3×10^8	1.0	19
153.6	512^3	7.5×10^9	0.0	23

Table 6.1: Λ CDM N-body simulations used in this chapter. First and second columns list the size of the box ($h^{-1}\text{Mpc}$) and the number of particles used. Third column gives the mass resolution of the simulations ($h^{-1}\text{M}_{\odot}$). Fourth column gives the redshift at which the simulation was terminated. Last column gives the number of snapshot outputs drawn from each simulation for calculating properties of galaxies.

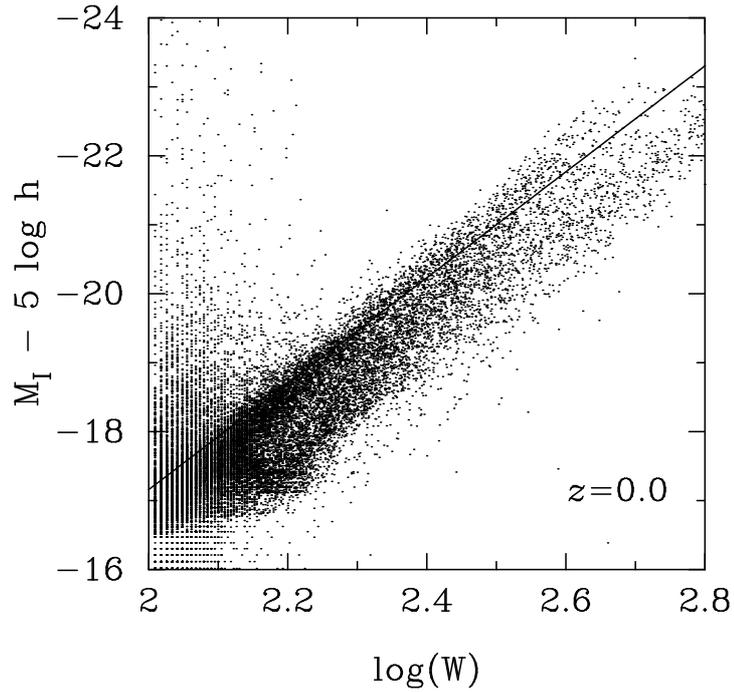


Figure 6.2: Tully-Fisher relation for galaxies in our calculation compared with observational result of Giovanelli et al. 1997.

Thus, in any given output snapshot, we deal with a population of central galaxies and satellite galaxies, each attached to the position of a simulation particle. Starting at the first output snapshot at high redshift, when the first haloes have formed, we initialize the galaxy population with a set of central galaxies, one for each halo, with stellar mass, cold gas mass, and luminosity set to zero. The physical properties of these galaxies are then evolved to the next output time, where we obtain a new galaxy population based on a combination of semi-analytic prescriptions and the merging history of dark matter haloes. Propagating this scheme forward in time, from output to output, we obtain the galaxy population at the present time, and at all output times at higher redshift.

We now describe the prescriptions used in this evolution. We begin by improving the definition of the progenitor-descendent relationship used in deriving the merger trees. According to the improved criterion, a halo H_B at redshift z_B is defined to be a progenitor of a halo H_A at redshift $z_A < z_B$ if (1) at least half of the particles of H_B are contained in H_A , and (2) the central particle of H_B is contained in H_A . At low redshifts ($z \lesssim 5$), this change typically reduces the number of haloes in merger trees by about 20%.

Next, beginning with the galaxy population at redshift z_B , we generate the galaxies of the new population at redshift z_A based on the merging history of the haloes. Using the halo catalogue at redshift z_B , we create an ‘initial’ population of galaxies at redshift z_A as follows:

- Each galaxy at z_B is assigned to its new halo at z_A .
- Each halo at z_A selects as its central galaxy, the central galaxy of its most massive progenitor. Central galaxies of other progenitors become satellite galaxies of the new halo.
- If a halo has no progenitors, a new central galaxy is created at the position of its most-bound particle.

Once this initial population of galaxies is created at redshift z_A , the properties of these galaxies are evolved for the time between the two snapshots according to our physical prescriptions to obtain the ‘final population’ of galaxies at redshift z_A .

We model the following physical processes: (1) radiative cooling of hot gas on to central galaxies, (2) transformation of cold gas into stars by star formation, (3) reheating of cold gas, or its ejection out of the halo, by supernova feedback, (4) orbital decay of satellite galaxies and their merging with central galaxies, (5) photometric evolution of galaxies, and (6) evolution of the metallicity of stars and the ISM of galaxies. Our prescriptions for these processes are essentially that developed by Kauffmann et al. 1999, Springel et al. 2001, De Lucia, Kauffmann & White 2004, and Croton et al. 2006. The main features of these prescriptions are as follows.

Gas cooling

Cooling of hot gas is modelled according to standard ideas presented in White & Frenk [1991]. According to this model, each dark matter halo collapses with an amount of baryons equal to the cosmic baryon-to-dark-matter ratio. (We take this to be $f_b = 0.17$ following WMAP 5-yr data.) These initial baryons are in the form of a diffuse gas with primordial composition, to which metals are added only after star formation has taken place. Upon entering the dark matter halo, the gas also shock heats to the virial temperature of the halo. We assume that the hot gas within a dark matter halo is distributed as an isothermal sphere with density profile $\rho_g(r)$ given by

$$\rho_g(r) = \frac{m_{\text{hot}}}{4\pi R_{\text{vir}} r^2}. \quad (6.1)$$

We can then define a local cooling time of the gas as the ratio of its specific thermal energy to the cooling rate per unit volume,

$$t_{\text{cool}}(r) = \frac{3}{2} \frac{\rho_g(r)}{\mu m_p n_e^2(r)} \frac{kT}{\Lambda(t, Z)}. \quad (6.2)$$

Here $n_e(r)$ is the electron density, μm_p is the mean particle mass and $\Lambda(t, Z)$ is the cooling rate. The cooling rate is also dependent on the gas metallicity. We model this dependence using the collisional ionization cooling curves of Sutherland & Dopita [1993]. At high temperatures bremsstrahlung cooling dominates, whereas at low temperatures cooling due to metals is the biggest contributor. The effect of the metallicity is thus to enhance cooling, thereby increasing star formation in galaxies. This effect is especially dominant in low mass galaxies [De Lucia, Kauffmann & White, 2004]. The virial temperature is given by

$$T = 35.9(V_c/\text{km s}^{-1})^2 \text{ K}. \quad (6.3)$$

Cooling of hot gas in haloes, depends on the redshifts and on the depth of the halo potential well. In low mass haloes at high redshift, the shocked gas cools rapidly and settles onto a central object. On the other hand, in high mass haloes at low redshift, gas spreads in the form of a quasi-static hot atmosphere that extends to the halo virial radius. We take this distinction into account by following Kauffmann et al. [1999] and defining a cooling radius r_{cool} as the radius within which the cooling time is equal to the halo dynamical time, R_{vir}/V_c . For small haloes at high redshifts, the cooling radius is larger than the virial radius. For such systems, the hot gas never attains hydrostatic equilibrium and cools rapidly on to the central object in a dynamical time. As a result, we have,

$$\dot{M}_{\text{cool}}(V_c, z) = \frac{M_{\text{hot}} V_c}{R_{\text{vir}}}. \quad (6.4)$$

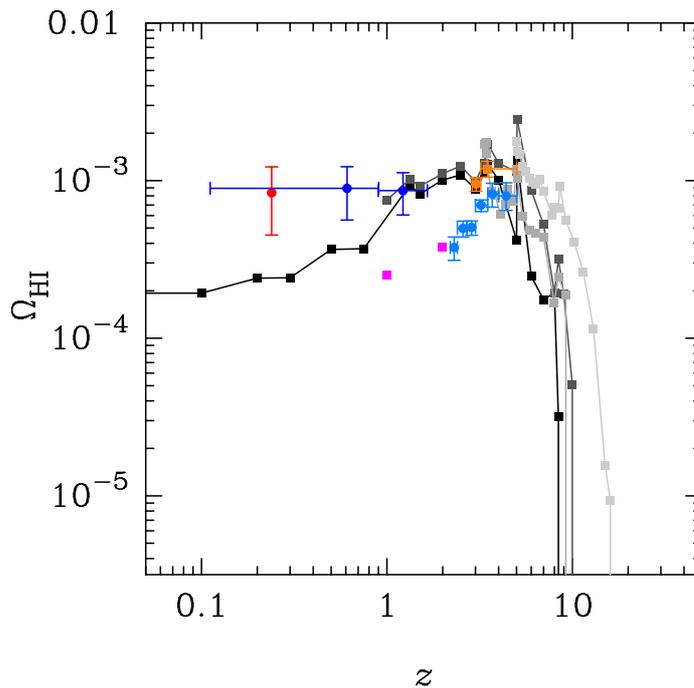


Figure 6.3: Cosmic H I density at various epochs. Black squares show results from our simulations, magenta squares are from the semi-analytic model of Duffy et al. [2011]. All round points are observations. Red points are from [Lah et al., 2009], blue points are from Rao, Turnshek & Nestor [2006], orange points are from [Noterdaeme et al., 2009], and the blue-cyan points are from [Prochaska & Wolfe, 2009].

For high mass haloes at late times, the cooling radius is smaller than the halo virial radius. In this case, the the hot atmosphere attains hydrostatic equilibrium and the cooling flow can be described by the continuity equation

$$\dot{M}_{\text{cool}}(V_c, z) = 4\pi\rho_g(r_{\text{cool}})r_{\text{cool}}^2 \frac{dr_{\text{cool}}}{dt}, \quad (6.5)$$

which gives us

$$\dot{M}_{\text{cool}}(V_c, z) = 0.5m_{\text{hot}} \frac{r_{\text{cool}}V_{\text{vir}}}{R_{\text{vir}}^2}. \quad (6.6)$$

Note that unlike previous models, we do *not* impose an ad hoc cutoff on cooling in our model in high mass haloes. This has significance for the H I power spectrum we obtain later in this chapter.

Star formation

We use the prescription presented by De Lucia, Kauffmann & White [2004] as our star formation recipe. In this prescription it is assumed that the star formation rate is proportional to the amount of available cold gas.

$$\dot{M}_* = (1 - R)\alpha M_{\text{cool}}/t_{\text{dyn}}, \quad (6.7)$$

where M_{cool} is the cold gas mass. The parameter α is the star formation efficiency. For a central galaxy, dynamical time t_{dyn} is given by

$$t_{\text{dyn}} = 0.1R_{\text{vir}}/V_c. \quad (6.8)$$

For a satellite galaxy, dynamical time value is fixed at that obtained when it was last a central galaxy. Following De Lucia, Kauffmann & White [2004], we assume that

$$\alpha = \alpha_0 \left(\frac{V_{\text{vir}}}{220 \text{ km s}^{-1}} \right)^n, \quad (6.9)$$

where α_0 and n are free parameters. We also assume that a fraction R of mass in stars is immediately returned to the cold gas reservoir.

Reheating and feedback

Supernova explosions play an important role in the evolution of the ISM. The amount of cold gas reheated by supernova explosions can be estimated using energy conservation arguments. Our prescription for supernova feedback is as used by Croton et al. [2006]. We assume that the amount of cold gas reheated by supernovae is given by

$$\Delta m_{\text{reheated}} = \epsilon_{\text{gal}} \Delta m_*, \quad (6.10)$$

where Δm_* is the amount of stars formed in a finite time interval, and ϵ_{gal} is a parameter that we set equal to 3.5 [Martin, 1999]. The total energy released in this interval can be approximated by

$$\Delta E_{\text{SN}} = 0.5\epsilon_{\text{halo}} \Delta m_* V_{\text{SN}}^2, \quad (6.11)$$

where $0.5V_{\text{SN}}^2$ is the mean energy of the supernova ejecta per unit mass of stars formed and ϵ_{halo} is a parameter that controls the efficiency with which this ejecta can heat cold gas. We adopt $V_{\text{SN}}^2 = 630 \text{ km s}^{-1}$ and $\epsilon_{\text{halo}} = 0.35$. If the reheated gas is added to the hot gas halo, then halo's thermal energy changes by

$$\Delta E_{\text{hot}} = 0.5\Delta m_{\text{reheated}} V_{\text{vir}}^2. \quad (6.12)$$

Thus the excess energy in the hot halo is

$$\Delta E_{\text{excess}} = \Delta E_{\text{SN}} - \Delta E_{\text{hot}}. \quad (6.13)$$

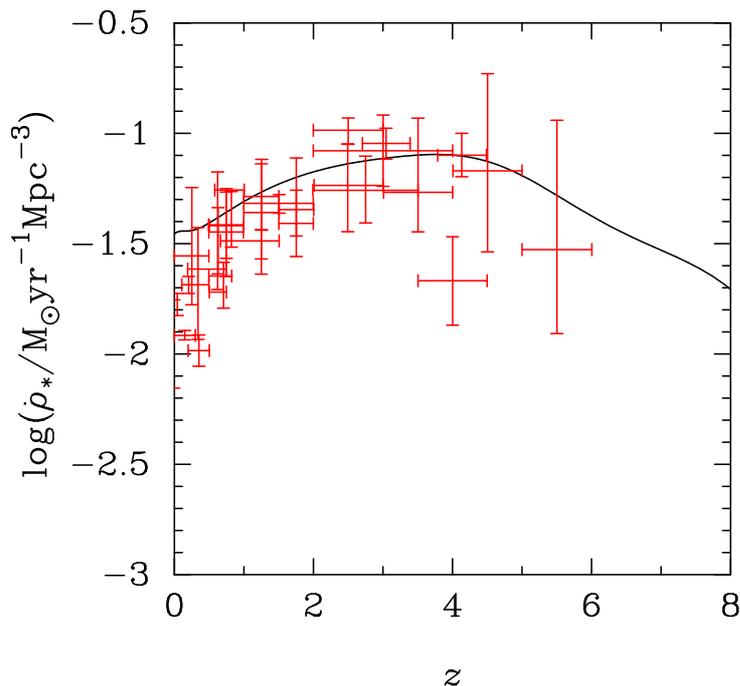


Figure 6.4: Black solid curve shows the evolution of the cosmic star formation rate density in our model. Data points are from a compilation of observations in Springel & Hernquist [2003b].

When $\Delta E_{\text{excess}} > 0$, we assume that some hot gas escapes the halo. The amount of hot gas ejected is given by

$$\Delta m_{\text{ejected}} = \frac{\Delta E_{\text{excess}}}{E_{\text{hot}}} m_{\text{hot}} = \left(\epsilon_{\text{halo}} \frac{V_{\text{SN}}^2}{V_{\text{vir}}^2} - \epsilon_{\text{gal}} \right) \Delta m_*, \quad (6.14)$$

where $E_{\text{hot}} = 0.5 m_{\text{hot}} V_{\text{vir}}^2$ is the total thermal energy of the hot gas. In this prescription, all hot gas can be ejected for small haloes at high redshifts. On the other hand, for large haloes, no hot gas is ejected. This treatment corresponds to the ‘ejection’ scheme of Kauffmann et al. [1999]. We assume that ejected gas can be reincorporated in the hot gas reservoir of a halo, if that halo grows by a factor of more than two between snapshots.

Mergers

Mergers of galaxies are common in the hierarchical picture of galaxy formation. In our model, we do not calculate the morphological evolution of galaxies. However, galaxy

mergers still have an important effect on the properties of the central galaxies due to effects like starbursts.

We treat merger of a satellite galaxy with a central galaxy by using a dynamical friction time scale suggested by the N-body studies of Navarro, Frenk & White [1995]. In this picture, satellite galaxies lose angular momentum and merge with central galaxies due to dynamical friction. The corresponding time scale is given by

$$T_{\text{df}} = \frac{1}{2} \frac{f(\epsilon)}{GC \ln(\Lambda)} \frac{V_c r_c^2}{M_{\text{sat}}} \quad (6.15)$$

where m_{sat} is the mass of a satellite orbiting at radius r_c in an isothermal halo of circular velocity V_c . The circularity of the orbit (defined as the ratio between the angular momentum of the current orbit relative to that of a circular orbit of equal energy) is denoted by ϵ and $f(\epsilon) = \epsilon^{0.78}$ incorporates the effect of the orbital eccentricity of the satellite on the dynamical friction time scale. Following Kauffmann et al. [1999], we always set $r_c = R_{\text{vir}}$. We also approximate the Coulomb logarithm as $\ln \Lambda = 1 + m_{\text{vir}}/m_{\text{sat}}$. We set the constant C to 0.43 and use the average value of $f(\epsilon)$, which is 0.5 [Lacey & Cole, 1993].

When a small satellite galaxy merges with a large central galaxy, its stars, cold gas and metals are simply added to the central galaxy. However, if the ratio of baryonic masses $m_{\text{central}}/m_{\text{sat}}$ is greater than 0.3, we implement a starburst in which a fraction e_{burst} of cold gas in the merger remnant is instantaneously consumed. Following Croton et al. [2006], we write this fraction as

$$e_{\text{burst}} = \beta_{\text{burst}} (m_{\text{sat}}/m_{\text{central}})^{\alpha_{\text{burst}}}, \quad (6.16)$$

where the parameters α_{burst} and β_{burst} take values of 0.7 and 0.56 respectively [Somerville, Primack & Faber, 2001; Cox et al., 2004].

Spectrophotometric evolution

We use stellar population synthesis models to calculate photometric properties of galaxies in our model. In these models, a stellar population is chosen according to an initial mass function and then each star is evolved along theoretical stellar evolutionary tracks. We use the STARBURST99 code [Leitherer et al., 1999; Vázquez & Leitherer, 2005] with a $0.1 - 100 M_{\odot}$ Kroupa IMF [Kroupa, 2002] with various values of initial metallicity (from 0.0004 to 0.05) to generate look-up tables. The SED of a galaxy is then computed from its star formation history $\dot{m}_*(t)$ as

$$S_{\nu}(t) = \int_0^t F_{\nu}(t-t', Z) \dot{m}_*(t') dt', \quad (6.17)$$

where $F_{\nu}(t)$ is the SED of a single-age population of star, which is obtained from the look-up tables by interpolating in t and Z . We also convolve with the standard UBVRI

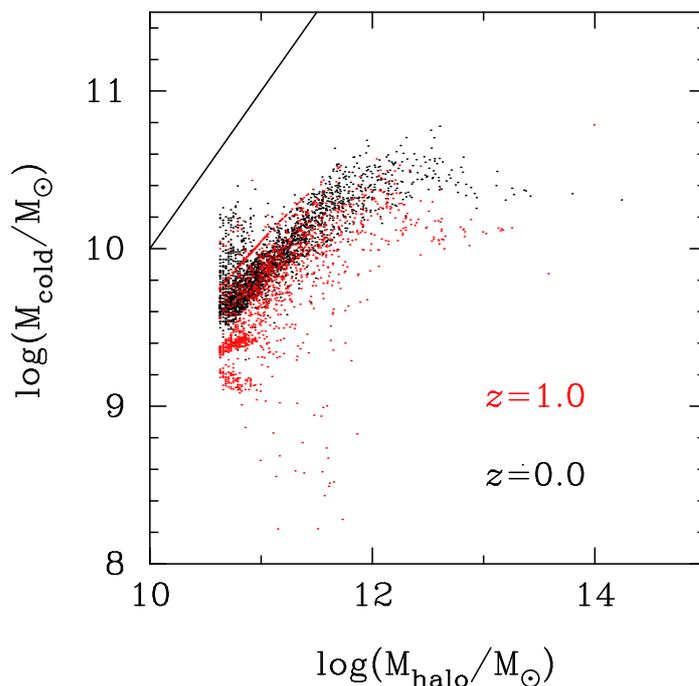


Figure 6.5: Total cold gas mass of galaxies as a function of the mass of their host dark matter halo. Points show individual model galaxies. The black line shows the total mass of the halo.

filters to obtain colours and luminosities in the desired bands. We do not incorporate the effects of dust in our model. Finally, note that stellar population synthesis also gives us the metal yields of stellar populations.

Metallicity evolution

We perform a self-consistent metallicity evolution of stars and the ISM of our model galaxies. Our metal evolution recipe is essentially that implemented in De Lucia, Kauffmann & White [2004]. Metallicity of newly formed stars is same as the cold gas that they formed out of. Furthermore, we assume that yield Y of metals is produced per unit solar mass of stars. These metals are instantaneously returned to the cold phase. Subsequent evolution depends on the exchange of metals between different gas phases. Metals can be exchanged between the cold and hot gas phases via reheating and cooling, and by reincorporation of ejected gas. Thus, for central galaxies, the stellar metallicity evolves as

$$\dot{M}_{\text{stars}}^Z = \dot{M}_* Z_{\text{cold}}. \quad (6.18)$$

Metallicity of the hot gas phase evolves as

$$\dot{M}_{\text{hot}}^Z = -\dot{M}_{\text{cool}}Z_{\text{hot}} + \dot{M}_{\text{back}}Z_{\text{ejected}} + \sum_{\text{gal}}[\dot{M}_{\text{reheated}}Z_{\text{cold}}]. \quad (6.19)$$

The cold gas metallicity evolves as

$$\dot{M}_{\text{cold}}^Z = \dot{M}_{\text{cool}}Z_{\text{hot}} - \dot{M}_{\text{stars}}^Z + Y\dot{M}_* - \dot{M}_{\text{out}}Z_{\text{cold}}, \quad (6.20)$$

and that of the ejected gas is given by

$$\dot{M}_{\text{ejected}}^Z = \dot{M}_{\text{out}}Z_{\text{cold}} - \dot{M}_{\text{back}}Z_{\text{ejected}}. \quad (6.21)$$

In all of above $Z_{\text{cold}} = M_{\text{cold}}^Z/M_{\text{cold}}$, $Z_{\text{hot}} = M_{\text{hot}}^Z/M_{\text{hot}}$, and $Z_{\text{ejected}} = M_{\text{ejected}}^Z/M_{\text{ejected}}$ are the metallicities of various phases. For satellite galaxies, equations are similar, except that the hot gas and ejected components are absent. Also, \dot{M}_{out} is the rate at which cold gas is ejected out of the halo, \dot{M}_{back} is the rate at which ejected gas is reincorporated into the halo, and \sum_{gal} denotes a summation over all satellite galaxies.

6.1.3 Calibration

As mentioned before, we propagate the properties of galaxies in our model using above prescriptions from one simulation output to the next and obtain the galaxy populations at all redshifts at which simulation outputs are available. If a new halo forms while going from a snapshot to the next, its hot gas mass is initialized to the value $f_b M_{\text{vir}}$. For every other halo, the initial hot gas mass available for cooling is given by

$$M_{\text{hot}} = f_b M_{\text{vir}} - \sum_{\text{gal}} [M_* + M_{\text{cool}}] - \sum_{\text{prog}} M_{\text{eject}}, \quad (6.22)$$

where \sum_{gal} denotes sum over all galaxies in the halo, and the last summation is only over those satellites that have newly entered the halo and whose mass is greater than $m_{\text{vir}}/2$.

We calibrate free parameters in our model to reproduce observed properties of the Milky Way. This is similar to previous semi-analytic models. We also check that we get the right normalization and slope for the Tully-Fisher relation. The free parameters in our model and their values are as follows:

- Parameters determining star formation efficiency: α_0 and n . We set these to 0.09 and 2.2 respectively. The parameter α_0 influences the gas fraction of galaxies and n controls the dependence of the gas fraction on galaxy mass.
- The supernova feedback reheating efficiency ϵ_{gal} and the ejection efficiency ϵ_{halo} are set to 3.5 and 0.35 respectively. These two parameters control the slope of the Tully-Fisher relation. Increased feedback makes the Tully-Fisher relation steeper by reducing the luminosity of low mass galaxies.

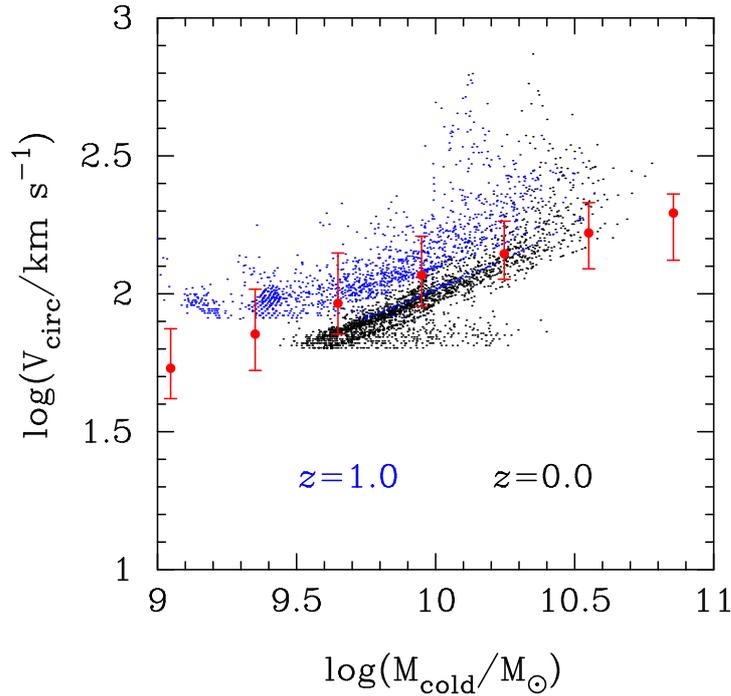


Figure 6.6: Predicted circular velocity-cold gas mass relation at $z = 0$ and 1. The red data points from the semi-analytical model of De Lucia & Blaizot [2007] are shown for comparison. Here, circular velocity is measured at the virial radius of the dark matter halo.

- The yield Y of metals produced per solar mass of star formation is given by the population synthesis. For our model, this value is 0.03. This has a marginal effect on the slope of the Tully-Fisher relation by enhancing the cooling in low mass galaxies.
- The ‘retention parameter’ R is also available from population synthesis. It has a value of 0.3 in our model. This parameter has a slight influence on galaxy luminosities.

Having specified our semi-analytic model and its implementation, we now proceed to study its results in the next section.

6.2 Results

We begin by comparing results of our model with observations of galaxies in the local universe ($z \sim 0$). We then proceed to understand the evolution of the H I content of

the universe. Finally we study the large-scale distribution of H I at various redshifts by looking at the power spectrum of fluctuation in the H I as predicted by our galaxy formation model.

6.2.1 Comparison with local observations

We select all objects in the simulation with circular velocities in the range 219-221 km s⁻¹ as Milky-Way-type galaxies. While doing this we assume that the circular velocity of a galaxy is ~ 25 per cent larger than the circular velocity of its halo. This follows earlier galaxy formation models [De Lucia, Kauffmann & White, 2004], and is motivated by detailed modelling of the structure of disk galaxies embedded in cold dark matter haloes with the universal NFW profile [Navarro, Frenk & White, 1995; Mo, Mao & White, 1998]. In our 153.6 h⁻¹Mpc run, we find 230 such galaxies at $z = 0$. We measure the stellar masses, gas masses, star formation rates, and metallicities of these galaxies. We find average total gas mass $1.2 \times 10^{10} M_{\odot}$, average stellar mass $8.0 \times 10^9 M_{\odot}$, average star formation rate $5.3 M_{\odot} \text{ yr}^{-1}$, and average metallicity of 0.98 in multiple of the solar metallicity. We also measure the *B*-band and *I*-band absolute magnitude, which comes to be -21.1 and -23.2 in the average. Due to our calibration, these values match the observed values of the Milky Way galaxy closely. Scatter around these quantities is relatively low. Note that since we do not calculate galaxy morphologies in our model, we are unable to select galaxies based on morphology.

Figure 6.1 shows the *B*-band luminosity function of all galaxies at $z = 0$ in our model. This luminosity function is shown in comparison with the observed *B*-band luminosity function of 110500 galaxies from the 2dF Galaxy Redshift Survey [2dFGRS; Norberg et al., 2002]. (This 2dFGRS luminosity function is in excellent agreement with the SDSS data when the latter are normalised in the same way, and the magnitudes are translated to the *B* band.) The model luminosity function exceeds the observed one by more than an order of magnitude at the bright end. We find that instead of declining exponentially at the bright end, the luminosity function exhibits a gentler turn down. This problem with the semi-analytical models has been noted before by Kauffmann et al. [1999]. One possible explanation of this mismatch is a problem with the FOF halo finding algorithm. Springel et al. [2001] show that including subhaloes leads to a significant improvement in cluster luminosity functions because of a more realistic merger rate. The improved luminosity function is found to be in good agreement with observations. Another possible explanation of the mismatch between the model and observed luminosity functions is the absence of AGN feedback in our model. Croton et al. [2006] find that suppression of cooling flows due to AGN feedback dramatically improves the match with the 2dFGRS data by reducing the luminosities of bright galaxies by up to two magnitudes. We also find a deficit at the faint end of the luminosity function, which is not seen by Croton et al. [2006]. This is due to the enhanced supernova feedback in small mass haloes in our model, which we need to match the slope and the zero-point

of the Tully-Fisher relation in absence of dust extinction.

In Figure 6.2, we show the Tully-Fisher relation obtained for our model galaxies. The solid line shows the relation measured by Giovanelli et al. [1997]. The slope of the model result matches the observations well. In our model, as discussed above, we follow Kauffmann et al. [1999] and calibrate our model so that the I -band magnitude of the central galaxy in a halo of circular velocity $V_c = 220 \text{ km s}^{-1}$ is $-22.1 + 5 \log h$. This agrees with the zero-point of the Tully-Fisher relation derived by Giovanelli et al. [1997]. There is large scatter in the Tully-Fisher relation in our simulations. This is because the lack of morphological evolution of galaxies in our model prevents us from isolating Sb/Sc type galaxies with specified bulge magnitude for this comparison. For this reason, Figure 6.2 shows a random sample of *all* central galaxies in our model. Part of the scatter is also because of that in the properties of galaxies in haloes of same circular velocity. The absence of a treatment of dust extinction in our calculation does not affect this comparison since the result of Giovanelli et al. [1997] is already corrected for internal extinction. The slope of the Tully-Fisher relation is strongly dependent on the adopted feedback prescriptions and, with more detailed modelling, can be used to observationally test different feedback models.

Thus we see that our model gives a good fit to the observations of the Tully-Fisher relation but does a poor job in reproducing the observed luminosity function. Obtaining a good fit to both of these observations has been known to be difficult in the Λ CDM model [Cole et al., 2000]. Croton et al. [2006] solve this problem by making an unrealistic approximation: For central galaxies, they take the rotational velocity to simply equal to the halo virial velocity, but for satellite galaxies the rotational velocity is taken the the virial velocity of haloes in which they were the central galaxies. McGaugh et al. [2000] have pointed out that the observed Tully-Fisher relation itself may deviate from the power law at the bright end. They found that combining gas mass in a galaxy with its stellar mass brings the Tully-Fisher relation closer to the power law. This agrees with the finding of Croton et al. [2006].

6.2.2 Global HI distribution

As we saw above, high column density features like the DLAs in the spectra of distant quasars indicate that at $1 \leq z \leq 5$, the neutral hydrogen content of the universe is almost constant with a density parameter of $\Omega_{\text{HI}} \sim 0.001$. Figure 6.3 shows the evolution of Ω_{HI} in our model. Results from all four simulation runs are shown together (grey symbols). These are compared with various observational estimates of this quantity. We find that Ω_{HI} rises rapidly from $z = 0$ to $z = 1$. Subsequently, its value peaks at around 10^{-3} at $z \sim 4$, before dropping off at higher redshifts. One issue here is that our semi-analytic model predicts only the total mass of cold gas, which includes helium and both atomic and molecular hydrogen. We have assumed that 76% of cold gas is in the form of

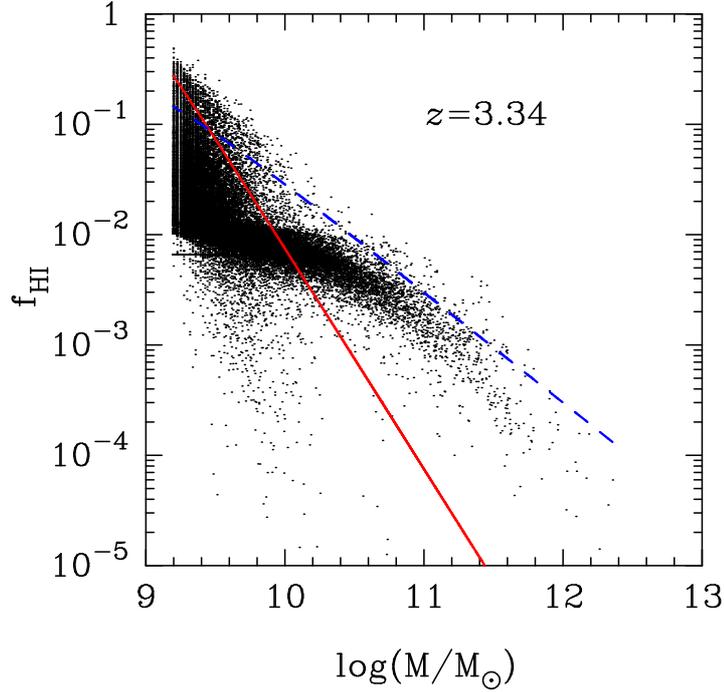


Figure 6.7: Cold gas fraction at $z = 3.34$ in our simulations compared the with prescriptions 2 and 3 of Bagla, Khandai & Datta [2010], shown by the red solid and blue dashed lines respectively. Upper mass cut-off in the prescription is not shown. High mass haloes contain a lot of H I in the simulations.

H I [Power, Baugh & Lacey, 2010].² Our results agree quite well with observational estimates. It is interesting to note here that a comparison for four currently favoured semi-analytical models by Power, Baugh & Lacey [2010] shows that all models underpredict Ω_{HI} at $z > 0$. Also, our model predicts a more rapid evolution in Ω_{HI} compared to a recent calculation using gas dynamical simulations by Duffy et al. [2011].

Figure 6.4 shows the global star formation history in our model. Our results are broadly consistent with observational data.

Figure 6.5 shows the cold gas mass of our model galaxies as a function of the mass of their host dark matter halo. We find that there is a strong correlation between the cold gas mass of a galaxy and its host halo mass. In haloes more massive than $10^{12} M_{\odot}$, the correlation breaks down but haloes continue to hold cold gas. It is instructive to compare

²Prochaska & Wolfe [2009]; Obreschkow & Rawlings [2009] have developed an empirical model of based on observations and theoretical arguments in which the $\text{H}_2/\text{H I}$ ratio could vary from galaxy to galaxy. [Obreschkow & Rawlings, 2009] applied this model to the galaxy formation model of De Lucia & Blaizot [2007]. They find that at $1 \leq z \leq 5$, galaxies contain similar amount of H I as today, but substantially larger amount of H_2 . We do not consider molecular hydrogen here.

this result with that obtained in the model of Bower et al. [2006], in which a dramatic break occurs at halo masses of $10^{12} M_{\odot}$ due to cooling flow suppression resulting from AGN feedback. Another interesting feature is that galaxies with the largest mass of cold gas do not lie in haloes with the highest mass. Instead, it is the haloes with mass of $10^{12} M_{\odot}$ that contain such galaxies.

Figure 6.6 compares the halo circular velocity with the halo cold gas mass content. This quantity is useful because it indicates how the velocity width—which is related to the rotation velocities of galactic disks, which in turn is related to the halo circular velocity—is likely to scale with H I mass. This is important for H I surveys. Also, this result is closely related to the previous one, which compared cold gas mass to the halo mass, since the halo mass can be related to the circular velocity. We find that our result agrees remarkably well with the semi-analytical model of De Lucia & Blaizot [2007]. The level of agreement is quite good considering the differences in the implementation of physical ingredients of the two models. One important difference between the two predictions is that in our model a lot of high circular velocity haloes contain cold gas. This is partly due to a lack of cutoff in circular velocity as in De Lucia & Blaizot [2007] and partly due to the absence of AGN feedback, which is usually the strongest remover of cold gas in these high mass haloes. A comparison of the results of De Lucia & Blaizot [2007] by Power, Baugh & Lacey [2010] suggests that our model also agrees with other galaxy formation models, with possible differences only at the low cold gas mass limit.

Finally, in Figure 6.7, we compare the cold gas fraction f_{HI} of haloes in our simulations with the prescriptions used in Bagla, Khandai & Datta [2010]. We find the form of the dependence of f_{HI} on halo mass M_{vir} is quite different from the prescription. Moreover, in the prescriptions it was assumed that haloes above a mass of $10^{11.5} M_{\odot}$ do not hold H I. Clearly this is not valid in our simulations. As we will see below, this fact has important consequences for the H I power spectra.

6.2.3 Fluctuations in HI

We now consider the power spectrum of fluctuations in H I density in our model. An advantage of the 21cm line is that it lets us separate the fluctuating component of the 21cm brightness temperature both in sky and in frequency. This promise of such ‘tomographic’ observations, and their usefulness for understanding galaxy formation, motivates our study. Although the global 21cm background contains useful information about high redshift universe, the main problem with the global signal is its slow evolution in the post-reionization era. H I fluctuations, on the other hand, grow more rapidly, especially at small scales, partly due to the evolution of density fluctuations on the matter distribution. This property is expected to help in separating the 21cm signal from foregrounds [Furlanetto, Oh & Briggs, 2006]. Traditionally, the power spectrum has been the most-favoured statistical quantity in the study of 21cm signal fluctuations.

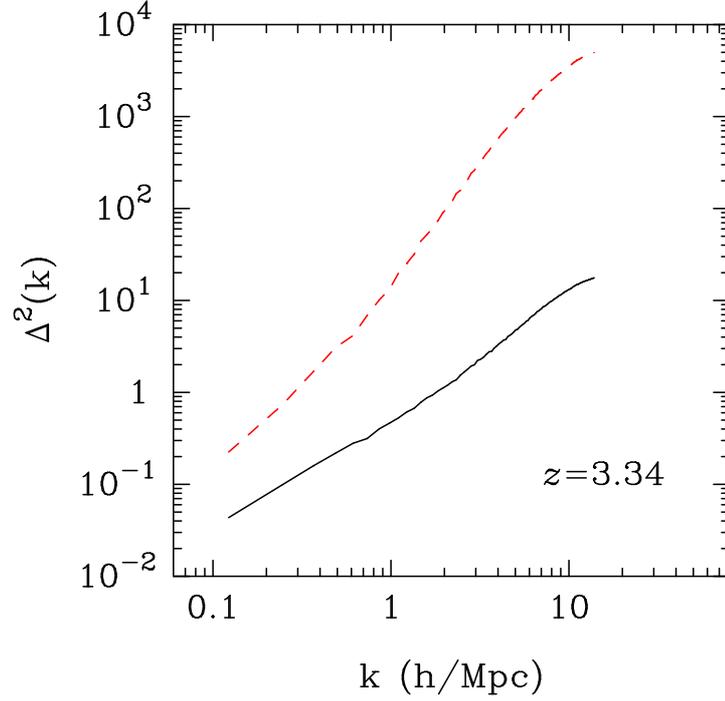


Figure 6.8: Power spectra in our model at $z = 3.34$. Black solid is dark matter and red dashed is H I. This is from the $51.2 \text{ h}^{-1} \text{ Mpc}$ box.

While other statistical quantities are also useful, the use of power spectrum is partly motivated by the success of CMB and large scale structure studies in constraining cosmological parameters using the power spectrum.

As we discussed in Chapter 1, the spin temperature of a hydrogen cloud couples to the gas temperature through collisional coupling with other atoms, electrons, ions and through Ly- α pumping via the Wouthuysen-Field effect [Purcell & Field, 1956; Field, 1959b; Furlanetto, Oh & Briggs, 2006; Wouthuysen, 1952]. Observations of the 21cm absorption by DLAs indicate that the spin temperature is orders of magnitude higher than the temperature of the CMBR at corresponding redshifts [Chengalur & Kanekar, 2000; Kanekar et al., 2009]. As a result, the 21cm brightness temperature is proportional to the density of neutral hydrogen [Furlanetto, Oh & Briggs, 2006; Bagla, Khandai & Datta, 2010]. With this in mind, we study the fluctuations in the H I density in this section. We choose to focus on the redshift of the GMRT, which are representative of the range of redshifts in the post-reionization universe. In particular, we focus on the redshifts $z = 5.04, 3.34$, and 1.33 . As discussed above, we use different simulation run at each redshift to balance the requirements of a high mass resolution and sufficiently large box size.

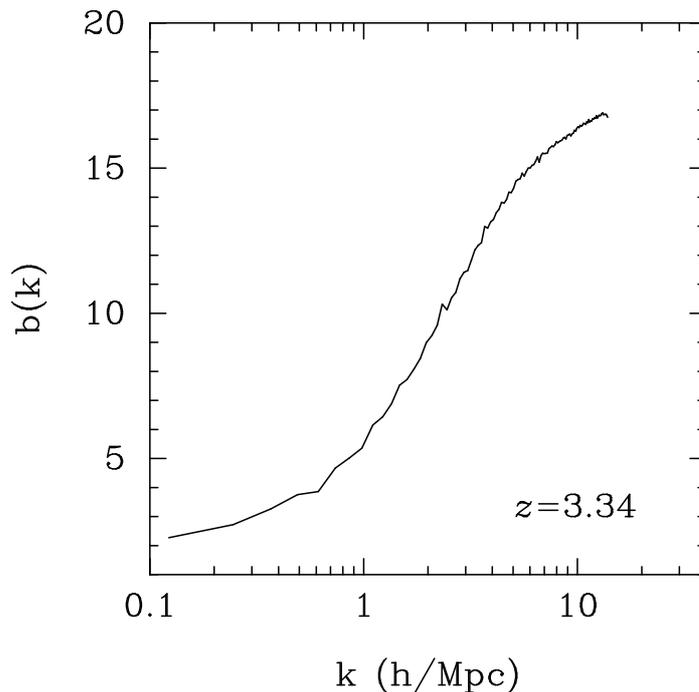


Figure 6.9: H I bias for the power spectra shown in Figure 6.8

We use the cloud-in-cell scheme to distribute particles in our simulation on a regular grid for the purpose of computing densities and the power spectrum. In order to compute the H I power spectrum, we first assign an H I fraction to each particle in the N-body simulation output. This assignment is done according to the results of our semi-analytic model of galaxy formation. As we saw in the previous section and in Figure 6.7, the semi-analytic model gives us the fractional H I mass in every halo of the simulation. We use this information by assigning to each particle in a given halo an H I mass that is a fraction f_{HI} of its total mass, where f_{HI} is the fractional halo mass of that particular halo.³ This is in contrast to the H I mass assignment used in earlier work [Bagla, Khandai & Datta, 2010] in two respects: (1) Two haloes with same total mass can have different H I mass in our scheme, unlike the previous scheme. This introduces a large scatter in H I as is evident in Figure 6.7. (2) Secondly, we do not have an in-built, arbitrary, high mass cutoff on f_{HI} .

We compute the power spectrum in both real and redshift space. For redshift-space calculation, we use the peculiar velocity of particles in halo as obtained from the N-

³An important caveat here is that due to our stricter definition of the progenitor-descendant relationship in halo merger trees, the number of haloes in the semi-analytic model is less than in the FOF halo catalogues. For the FOF haloes that are absent in the semi-analytic scheme, we set $f_{\text{HI}} = 0$.

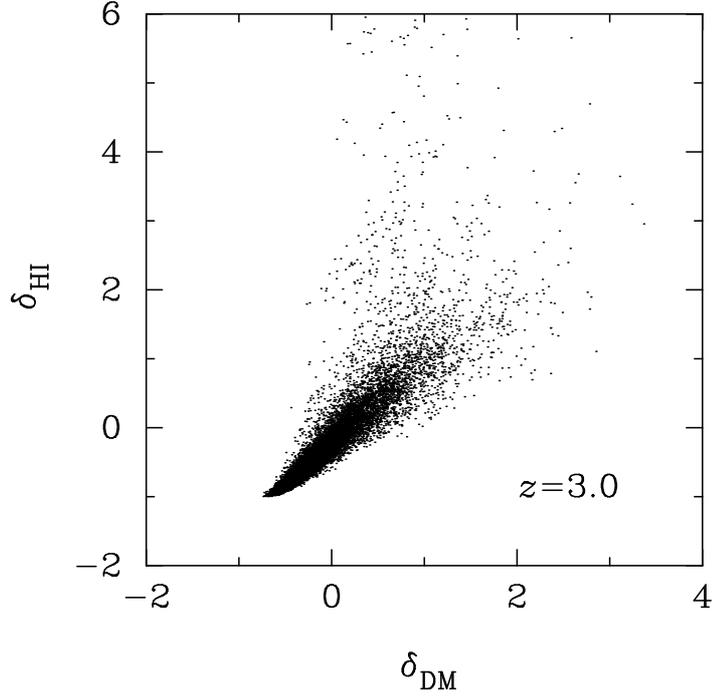


Figure 6.10: Scatter plot of H I and dark matter density contrast smoothed on $3 h^{-1}\text{Mpc}$ scale. The $51.2 h^{-1}\text{Mpc}$ box is used here.

body simulation. Thanks to our high mass resolution, we do not have to add the internal velocity dispersion in haloes by hand [Kumar, Padmanabhan & Subramanian, 1995; Bagla, Nath & Padmanabhan, 1997; Bagla & White, 2003; Bagla, Khandai & Datta, 2010]. Instead of expressing the H I power spectrum in terms of the brightness temperature, we use the usual dimensionless form here, for convenience. We also calculate the real and redshift space H I bias, which are given by

$$b(k) = \left[\frac{P_{\text{HI}}(k)}{P_{\text{DM}}(k)} \right]^{1/2} \quad (6.23)$$

and

$$b^s(k) = \left[\frac{P_{\text{HI}}^s(k)}{P_{\text{DM}}^s(k)} \right]^{1/2}. \quad (6.24)$$

Figure 6.8 shows the real space power spectrum of fluctuations at $z = 3.34$ in our simulations. The solid black line shows the non-linear dark matter power spectrum and the red dashed line shows its H I counterpart. Both the dark matter and H I power spectra have been computed using the simulation with $L_{\text{box}} = 51.2 h^{-1} \text{Mpc}$. The corresponding H I bias $b(k)$ is shown in Figure 6.9. Clearly, the bias is much higher

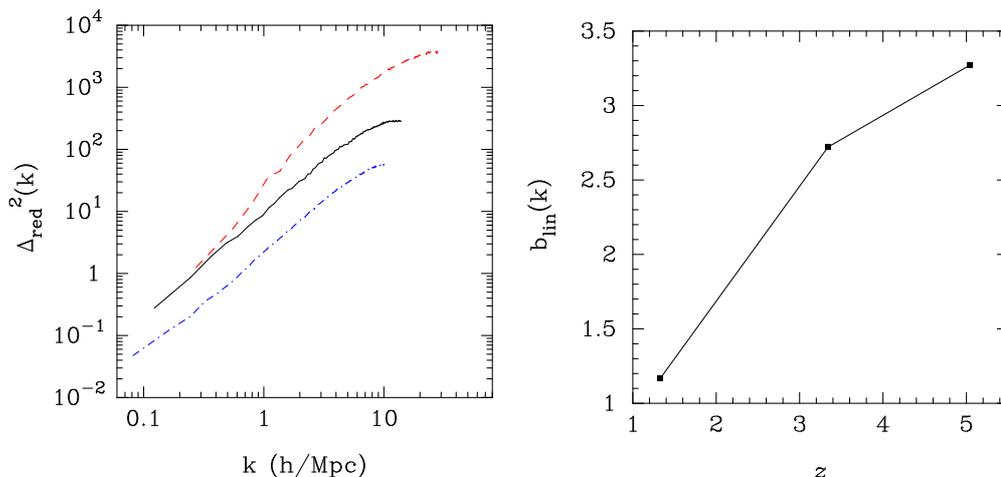


Figure 6.11: Left panel shows the evolution of redshift-space power spectra. Blue dot-dashed line is at $z = 1.3$ from the $76.8 \text{ h}^{-1}\text{Mpc}$ box, black solid line is at $z = 3.34$ from the $51.2 \text{ h}^{-1}\text{Mpc}$ box, and dashed red line is at $z = 5.1$ from the $23.04 \text{ h}^{-1}\text{Mpc}$ box. Evolution of the H I linear bias is shown in the right panel.

at high redshifts [Fry, 1996; Mo & White, 1996; Bagla, 1998a,b; Mo, Mao & White, 1998; Baugh et al., 1999; Magliocchetti et al., 2000; Benson et al., 2000; Roukema & Valls-Gabaud, 2000; Sheth, Mo & Tormen, 2001; Wyithe & Brown, 2010; Bagla, Khandai & Datta, 2010]. The bias is also scale dependent, and leads to a significant enhancement in the H I power spectrum at small scales. Although the linear bias in this model is comparable to that in the previous prescription-based work, the bias in our model increases more rapidly as we go towards smaller scales. The value of the bias at small scales is higher than the previous estimate of Bagla, Khandai & Datta [2010] by about a factor of four. This is not surprising if we note the fact that the bias and its scale-dependence has a strong dependence on the characteristic mass of haloes with H I. Since high mass haloes do hold H I in our semi-analytic model, the bias in this model is quite high. It is suggested that a high circular velocity cutoff has to be incorporated in semi-analytical models of galaxy formation in order to avoid the ‘cooling flow problem’ in galaxy cluster [De Lucia, Kauffmann & White, 2004]. Such a cutoff could reduce the H I bias. Similarly, AGN feedback can reduce the bias by pushing cold gas out of their parent haloes. Conversely, observations of clustering in H I sources at high redshift could be used to constrain masses of haloes that contain H I [Wyithe, 2008], thereby constraining feedback mechanisms in our model.

In Figure 6.10, we study the stochasticity of bias in the H I distribution [Dekel & Lahav, 1999]. This figure shows a scatter plot of the density contrast δ_{HI} and δ_{DM} smoothed at a scale of $3 \text{ h}^{-1} \text{ Mpc}$ at a random subset of points in our simulation grid.

We find that the scatter about the average $\delta_{\text{HI}} - \delta_{\text{DM}}$ trend is significant, and increases as we go to points with high dark matter over-density. This result is in agreement with what we find in earlier work [Bagla, Khandai & Datta, 2010].

We study the evolution of the redshift space power spectrum of H I distribution through the GMRT redshift windows in Figure 6.11. The curves show the dimensionless redshift space power spectrum $\Delta_{\text{red}}^2(k)$ at $z = 1.3$ (dot-dashed line), $z = 3.34$ (solid line), and $z = 5.1$ dashed line. The enhancement at small scales is less strong than the real space power spectrum, due to velocity dispersion within haloes at small scales. On the other hand, power is enhanced at large scales due to the Kaiser effect. The evolution of linear bias, that is the bias at large scales, through these redshifts is shown in Figure 6.11. We see that the bias decreases from about 3.4 at $z = 5.1$ to around 1.2 at $z = 1.3$.

6.3 Discussion and conclusions

In this chapter, we used a semi-analytical model of galaxy formation to predict the abundance and large scale distribution of H I in the post-reionization universe. Our galaxy formation model is calibrated such that it matches a wide variety of observations of properties of galaxies in the local universe. Main results of this chapter can be summarised as follows:

- Predictions for the H I density parameter Ω_{HI} are in good agreement with observations in the local universe. Our model predicts a peak value of $\Omega_{\text{HI}} \sim 10^{-3}$ at around $z \sim 5$.
- H I distribution is strongly biased at high redshift. This enhances the H I power spectrum significantly as compared to the dark matter power spectrum.
- Bias decreases rapidly with decreasing redshift. This leads to a decrease in the H I density fluctuation power spectrum with decreasing redshift with a corresponding decrease in the 21cm brightness temperature power spectrum.
- Small scale bias is higher in our model than in previous work. This is because of a lack of AGN feedback and an ad hoc H I cut-off.
- The enhanced H I power spectrum in our model improves the prospects for a detection with instruments like GMRT and MWA [Bagla, Khandai & Datta, 2010].
- Rare bright H I peaks are enhanced in our model as compared to earlier work. This augurs well for detection with instruments like the GMRT [Bagla, Khandai & Datta, 2010].

The main caveat in our model is an absence of AGN feedback. We expect AGN feedback to explain more observations of galaxy properties than what the present model can do, e.g., exponential cut-off at the bright end of the galaxy luminosity function and the low observed mass drop-out rate cooling flows.

Bibliography

- Aarseth S. J., 1999, *PASP*, 111, 1333
- , 2003, *Gravitational N-Body Simulations*. Cambridge, UK: Cambridge University Press
- Abdalla F. B., Blake C., Rawlings S., 2010, *MNRAS*, 401, 743
- Abdalla F. B., Rawlings S., 2005, *MNRAS*, 360, 27
- Abel T., Norman M. L., Madau P., 1999, *ApJ*, 523, 66
- Adelman-McCarthy J. K. et al., 2008, *ApJS*, 175, 297
- Ahmad A., Cohen L., 1973, *Journal of Computational Physics*, 12, 389
- Amaro-Seoane P., Freitag M. D., 2011, *MNRAS*, 412, 551
- Amaro-Seoane P. et al., 2010, *MNRAS*, 402, 2308
- Armitage P. J., Natarajan P., 2002, *ApJ*, 567, L9
- Bagla J. S., 1998a, *MNRAS*, 297, 251
- , 1998b, *MNRAS*, 299, 417
- , 2002, *Journal of Astrophysics and Astronomy*, 23, 185
- Bagla J. S., Khandai N., Datta K. K., 2010, *MNRAS*, 407, 567
- Bagla J. S., Khandai N., Kulkarni G., 2009, *ArXiv e-prints*, 0908.2702
- Bagla J. S., Kulkarni G., Padmanabhan T., 2009, *MNRAS*, 397, 971
- Bagla J. S., Nath B., Padmanabhan T., 1997, *MNRAS*, 289, 671

-
- Bagla J. S., Padmanabhan T., 1997, *Pramana*, 49, 161
- Bagla J. S., Prasad J., 2006, *MNRAS*, 370, 993
- Bagla J. S., Prasad J., Khandai N., 2009, *MNRAS*, 395, 918
- Bagla J. S., Ray S., 2003, *New Astron.*, 8, 665
- , 2005, *MNRAS*, 358, 1076
- Bagla J. S., White M., 2003, in *Astronomical Society of the Pacific Conference Series*, Vol. 289, *The Proceedings of the IAU 8th Asian-Pacific Regional Meeting, Volume I*, S. Ikeuchi, J. Hearnshaw, & T. Hanawa, ed., pp. 251–254
- Bagla J. S., Yadav J., Seshadri T. R., 2008, *MNRAS*, 390, 829
- Bahcall J. N., Greenstein J. L., Sargent W. L. W., 1968, *ApJ*, 153, 689
- Baker J. G. et al., 2006a, *Physical Review Letters*, 96, 111102
- , 2006b, *ApJ*, 653, L93
- Balogh M. L. et al., 2001, *MNRAS*, 326, 1228
- Barkana R., Loeb A., 2001, *Phys. Rep.*, 349, 125
- Barnes D. G. et al., 2001, *MNRAS*, 322, 486
- Barnes J., Hut P., 1986, *Nature*, 324, 446
- Bartelmann M., 1996, *A&A*, 313, 697
- Bartelmann M., King L. J., Schneider P., 2001, *A&A*, 378, 361
- Baugh C. M., 2006, *Reports on Progress in Physics*, 69, 3101
- Baugh C. M. et al., 1999, *MNRAS*, 305, L21
- , 2004, *New Astron. Reviews*, 48, 1239
- Becker G. D., Rauch M., Sargent W. L. W., 2009, *ApJ*, 698, 1010
- Becker R. H. et al., 2001, *AJ*, 122, 2850
- Begelman M. C., Blandford R. D., Rees M. J., 1980, *Nature*, 287, 307
- Bekenstein J. D., 1973, *ApJ*, 183, 657

- Bennett C. L. et al., 1996, *ApJ*, 464, L1+
- Benson A. J., 2005, *MNRAS*, 358, 551
- Benson A. J. et al., 2000, *MNRAS*, 311, 793
- Benson A. J., Kamionkowski M., Hassani S. H., 2005, *MNRAS*, 357, 847
- Berczik P. et al., 2006, *ApJ*, 642, L21
- Bernardeau F. et al., 2002, *Phys. Rep.*, 367, 1
- Bharadwaj S., Ali S. S., 2005, *MNRAS*, 356, 1519
- Bharadwaj S., Nath B. B., Sethi S. K., 2001, *Journal of Astrophysics and Astronomy*, 22, 21
- Bharadwaj S., Sethi S. K., 1998, *ApJS*, 114, 37
- , 2001, *Journal of Astrophysics and Astronomy*, 22, 293
- Bharadwaj S., Srikant P. S., 2004, *Journal of Astrophysics and Astronomy*, 25, 67
- Bi H., Davidsen A. F., 1997, *ApJ*, 479, 523
- Binney J., 1977, *ApJ*, 215, 483
- Binney J., Tremaine S., 2008, *Galactic Dynamics: Second Edition*. Princeton, NJ USA: Princeton University Press
- Blaes O., Lee M. H., Socrates A., 2002, *ApJ*, 578, 775
- Blain A. W., Longair M. S., 1993, *MNRAS*, 264, 509
- Blake C. A. et al., 2004, *New Astron. Reviews*, 48, 1063
- Blecha L. et al., 2011, *MNRAS*, 412, 2154
- Bloom J. S. et al., 2011, *ArXiv e-prints*, 1104.3257
- Bogdanović T., Reynolds C. S., Miller M. C., 2007, *ApJ*, 661, L147
- Bolton J., Meiksin A., White M., 2004, *MNRAS*, 348, L43
- Bolton J. S., Haehnelt M. G., 2007, *MNRAS*, 382, 325
- Bond J. R. et al., 1991, *ApJ*, 379, 440

-
- Booth R. S. et al., 2009, ArXiv e-prints
- Borriello A., Salucci P., 2001, MNRAS, 323, 285
- Bouchet F. R., Adam J.-C., Pellat R., 1985, A&A, 144, 413
- Bouchet F. R., Kandrup H. E., 1985, ApJ, 299, 1
- Bouwens R., Illingworth G., 2006, New Astronomy Review, 50, 152
- Bouwens R. J. et al., 2009, ApJ, 705, 936
- , 2007, ApJ, 670, 928
- , 2008, ApJ, 686, 230
- , 2010a, ApJ, 725, 1587
- , 2011, Nature, 469, 504
- , 2010b, ArXiv e-prints, 1006.4360
- , 2010c, ApJ, 709, L133
- , 2010d, ApJ, 708, L69
- Bower R. G. et al., 2006, MNRAS, 370, 645
- Boylan-Kolchin M., Ma C., Quataert E., 2008, MNRAS, 383, 93
- Boylan-Kolchin M. et al., 2009, MNRAS, 398, 1150
- Bradley L. D. et al., 2008, ApJ, 678, 647
- Bromm V., 2004, PASP, 116, 103
- Bromm V., Larson R. B., 2004, ARA&A, 42, 79
- Bromm V., Loeb A., 2003, ApJ, 596, 34
- , 2007, in American Institute of Physics Conference Series, Vol. 937, Supernova 1987A: 20 Years After: Supernovae and Gamma-Ray Bursters, S. Immler, K. Weiler, & R. McCray, ed., pp. 532–541
- Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
- Bullock J. S., Kravtsov A. V., Weinberg D. H., 2000, ApJ, 539, 517

- Bundy K., Treu T., Ellis R. S., 2007, *ApJ*, 665, L5
- Bunker A. J. et al., 2004, *MNRAS*, 355, 374
- , 2010, *MNRAS*, 1, 1378
- Burigana C. et al., 2008, *MNRAS*, 385, 404
- Campanelli M. et al., 2007a, *ApJ*, 659, L5
- , 2007b, *Physical Review Letters*, 98, 231102
- Castellano M. et al., 2010, *A&A*, 511, A20+
- Cen R., 2003, *ApJ*, 591, L5
- , 2005, *ArXiv Astrophysics e-prints*, 0507014
- Cen R., Haiman Z., Mesinger A., 2005, *ApJ*, 621, 89
- Cen R., Ostriker J. P., 1999, *ApJ*, 514, 1
- Chand H. et al., 2004, *A&A*, 417, 853
- Chandrasekhar S., 1943a, *ApJ*, 97, 255
- , 1943b, *Reviews of Modern Physics*, 15, 1
- Chatterjee P., Hernquist L., Loeb A., 2002, *ApJ*, 572, 371
- Chen X. et al., 2009, *ApJ*, 697, L149
- , 2011, *ApJ*, 729, 13
- Chengalur J. N., Braun R., Wieringa M., 2001, *A&A*, 372, 768
- Chengalur J. N., Kanekar N., 2000, *MNRAS*, 318, 303
- Chiu W. A., Ostriker J. P., 2000, *ApJ*, 534, 507
- Choudhury T. R., 2009, *ArXiv e-prints*, 0904.4596
- Choudhury T. R., Ferrara A., 2005, *MNRAS*, 361, 577
- , 2006, *MNRAS*, 371, L55
- , 2007, *MNRAS*, 380, L6

-
- Choudhury T. R., Ferrara A., Gallerani S., 2008, MNRAS, 385, L58
- Civano F. et al., 2010, ApJ, 717, 209
- Cohn J. D., Bagla J. S., White M., 2001, MNRAS, 325, 1053
- Cohn J. D., White M., 2008, MNRAS, 385, 2025
- Cole S. et al., 2000, MNRAS, 319, 168
- , 2001, MNRAS, 326, 255
- Colombo L. P. L., Pierpaoli E., 2009, New Astron., 14, 269
- Colpi M., Dotti M., 2009, ArXiv e-prints, 0906.4339
- Colpi M., Mayer L., Governato F., 1999, ApJ, 525, 720
- Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
- Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
- Cowie L. L. et al., 1995, AJ, 109, 1522
- Cox T. J. et al., 2004, ApJ, 607, L87
- Croft R. A. C. et al., 1998, ApJ, 495, 44
- , 1999, ApJ, 520, 1
- Croton D. J. et al., 2006, MNRAS, 365, 11
- Daigne F. et al., 2006, ApJ, 647, 773
- , 2004, ApJ, 617, 693
- Danforth C. W., Shull J. M., 2008, ApJ, 679, 194
- Davidson A. F., Kriss G. A., Zheng W., 1996, Nature, 380, 47
- Davis M. et al., 1985, ApJ, 292, 371
- de Bernardis P. et al., 2000, Nature, 404, 955
- De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
- De Lucia G., Kauffmann G., White S. D. M., 2004, MNRAS, 349, 1101

- De Simone A., Maggioro M., Riotto A., 2011, ArXiv e-prints, 1102.0046
- Dekel A., Lahav O., 1999, ApJ, 520, 24
- Dekel A., Silk J., 1986, ApJ, 303, 39
- Dijkstra M., Haiman Z., Loeb A., 2004, ApJ, 613, 646
- Dijkstra M., Wyithe J. S. B., Haiman Z., 2007, MNRAS, 379, 253
- Duffy A. R. et al., 2011, ArXiv e-prints
- Dunkley J. et al., 2009, ApJS, 180, 306
- Efstathiou G., 1992, MNRAS, 256, 43P
- Efstathiou G., Bond J. R., White S. D. M., 1992, MNRAS, 258, 1P
- Efstathiou G. et al., 1988, MNRAS, 235, 715
- Eisenstein D. J., Loeb A., 1995, ApJ, 443, 11
- Eke V. R. et al., 2006, MNRAS, 370, 1147
- , 2004, MNRAS, 355, 769
- Ellison S. L. et al., 2000, AJ, 120, 1175
- Escala A. et al., 2004, ApJ, 607, 765
- , 2005, ApJ, 630, 152
- Faber S. M. et al., 1997, AJ, 114, 1771
- Fakhouri O., Ma C., 2009, MNRAS, 394, 1825
- Fakhouri O., Ma C., Boylan-Kolchin M., 2010, MNRAS, 406, 2267
- Fan X., 2006, New Astron. Reviews, 50, 665
- Fan X., Carilli C. L., Keating B., 2006, ARA&A, 44, 415
- Fan X. et al., 2001, AJ, 122, 2833
- , 2006, AJ, 132, 117
- Fardal M. A. et al., 2001, ApJ, 562, 605

-
- Ferrarese L., 2002, *ApJ*, 578, 90
- Ferrarese L. et al., 2006, *ApJS*, 164, 334
- Ferrarese L., Ford H., 2005, *Space Sci. Rev.*, 116, 523
- Ferrarese L., Merritt D., 2000, *ApJ*, 539, L9
- Field G. B., 1959a, *ApJ*, 129, 525
- , 1959b, *ApJ*, 129, 536
- Finkelstein S. L. et al., 2010, *ApJ*, 719, 1250
- Forero-Romero J. E. et al., 2010, *MNRAS*, 403, L31
- Freidmann A., 1922, *Z. Phys.*, 10, 377
- Freitag M., Amaro-Seoane P., Kalogera V., 2006, *ApJ*, 649, 91
- Fry J. N., 1996, *ApJ*, 461, L65+
- Frye B., Broadhurst T., Benítez N., 2002, *ApJ*, 568, 558
- Furlanetto S. R., Loeb A., 2003, *ApJ*, 588, 18
- Furlanetto S. R., Oh S. P., Briggs F. H., 2006, *Phys. Rep.*, 433, 181
- Gallerani S. et al., 2008a, *MNRAS*, 386, 359
- , 2008b, *MNRAS*, 388, L84
- Gebhardt K. et al., 2000, *ApJ*, 539, L13
- Geil P. M., Wyithe J. S. B., 2009, *MNRAS*, 399, 1877
- Genzel R. et al., 2003, *ApJ*, 594, 812
- Giocoli C. et al., 2007, *MNRAS*, 376, 977
- Giovanelli R. et al., 1997, *ApJ*, 477, L1+
- , 2005, *AJ*, 130, 2613
- Gnedin N. Y., 2000, *ApJ*, 542, 535
- Gnedin N. Y., Hui L., 1998, *MNRAS*, 296, 44

- Gnedin O. Y., 2003, *ApJ*, 589, 752
- González J. A. et al., 2007, *Physical Review Letters*, 98, 231101
- Goodman J., 1995, *Phys. Rev. D*, 52, 1821
- Goodman J., Tan J. C., 2004, *ApJ*, 608, 108
- Graham A. W., 2004, *ApJ*, 613, L33
- Gualandris A., Merritt D., 2008, *ApJ*, 678, 780
- Gültekin K. et al., 2009, *ApJ*, 698, 198
- Gunn J. E., Gott, III J. R., 1972, *ApJ*, 176, 1
- Gunn J. E., Peterson B. A., 1965, *ApJ*, 142, 1633
- Haehnelt M. G., Kauffmann G., 2002, *MNRAS*, 336, L61
- Haehnelt M. G. et al., 2001, *ApJ*, 549, L151
- Haiman Z., 2000, in *The First Stars*, A. Weiss, T. G. Abel, & V. Hill, ed., pp. 219–+
- Haiman Z., Cen R., 2005, *ApJ*, 623, 627
- Haiman Z., Holder G. P., 2003, *ApJ*, 595, 1
- Haiman Z., Loeb A., 1997, *ApJ*, 483, 21
- , 1999a, *ApJ*, 519, 479
- , 1999b, *ApJ*, 521, L9
- Heap S. R. et al., 2000, *ApJ*, 534, 69
- Heath D. J., 1977, *MNRAS*, 179, 351
- Heggie D. C., Mathieu R. D., 1986, in *Lecture Notes in Physics*, Berlin Springer Verlag, Vol. 267, *The Use of Supercomputers in Stellar Dynamics*, P. Hut & S. L. W. McMillan, ed., pp. 233–+
- Heinämäki P., 2001, *A&A*, 371, 795
- Henry A. L. et al., 2008, *ApJ*, 680, L97
- , 2007, *ApJ*, 656, L1

-
- , 2009, *ApJ*, 697, 1128
- Herrmann F. et al., 2007, *Classical and Quantum Gravity*, 24, 33
- Hickey S. et al., 2010, *MNRAS*, 404, 212
- Hockney R. W., Eastwood J. W., 1988, *Computer simulation using particles*. Bristol: Hilger
- Hoffman L., Loeb A., 2007, *MNRAS*, 377, 957
- Hopkins P. F., Hernquist L., 2010, *MNRAS*, 407, 447
- Hopkins P. F. et al., 2006, *ApJS*, 163, 1
- , 2007, *ApJ*, 669, 45
- Hoyle F., 1953, *ApJ*, 118, 513
- Hui L., Gnedin N. Y., Zhang Y., 1997, *ApJ*, 486, 599
- Hut P., Rees M. J., 1992, *MNRAS*, 259, 27P
- Iliev I. T. et al., 2007, *MNRAS*, 376, 534
- , 2008, *MNRAS*, 391, 63
- Iwasawa M., Funato Y., Makino J., 2006, *ApJ*, 651, 1059
- Iye M. et al., 2006, *Nature*, 443, 186
- Janknecht E. et al., 2006, *A&A*, 458, 427
- Janssen G. H. et al., 2008, in *American Institute of Physics Conference Series*, Vol. 983, 40 Years of Pulsars: Millisecond Pulsars, Magnetars and More, C. Bassa, Z. Wang, A. Cumming, & V. M. Kaspi, ed., pp. 633–635
- Jarosik N. et al., 2011, *ApJS*, 192, 14
- Jenet F. et al., 2009, *ArXiv e-prints*, 0909.1058
- Jenkins A. et al., 2001, *MNRAS*, 321, 372
- Jiang C. Y. et al., 2008, *ApJ*, 675, 1095
- Jimenez R., Verde L., Oh S. P., 2003, *MNRAS*, 339, 243
- Johnston S. et al., 2008, *Experimental Astronomy*, 22, 151

- Kanekar N., Briggs F. H., 2004, *New Astron. Reviews*, 48, 1259
- Kanekar N. et al., 2009, *MNRAS*, 396, 385
- Kauffmann G. et al., 1999, *MNRAS*, 303, 188
- Kauffmann G., Haehnelt M., 2000, *MNRAS*, 311, 576
- Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, 264, 201
- Kay S. T., 2004, *MNRAS*, 347, L13
- Kereš D. et al., 2005, *MNRAS*, 363, 2
- Khan F., Just A., Merritt D., 2011, *ArXiv e-prints*, 1103.0272
- Khandai N., Bagla J. S., 2009, *Research in Astronomy and Astrophysics*, 9, 861
- Khandai N. et al., 2011, *MNRAS*, 415, 2580
- Khochfar S., Burkert A., 2006, *A&A*, 445, 403
- Kim H. et al., 2010, *ArXiv e-prints*, 1003.0008
- Kim H.-S. et al., 2011, *MNRAS*, 414, 2367
- Kim S. et al., 2009, *ApJ*, 695, 809
- Kim T.-S. et al., 2002, *MNRAS*, 335, 555
- Kitayama T., Suto Y., 1996, *MNRAS*, 280, 638
- Kleyna J. T. et al., 2003, *ApJ*, 588, L21
- Klypin A. A., Shandarin S. F., 1983, *MNRAS*, 204, 891
- Kocsis B., Yunes N., Loeb A., 2011, *ArXiv e-prints*, 1104.2322
- Kolb E. W., Turner M. S., 1990, *The early universe*. Westview Press
- Komatsu E. et al., 2009, *ApJS*, 180, 330
- , 2011, *ApJS*, 192, 18
- Komiya Y. et al., 2007, *ApJ*, 658, 367
- Kroupa P., 2002, in *Astronomical Society of the Pacific Conference Series*, Vol. 285, *Modes of Star Formation and the Origin of Field Populations*, E. K. Grebel & W. Brandner, ed., pp. 86–+

-
- Kulkarni G., Choudhury T. R., 2011, MNRAS, 412, 2781
- Kulkarni S. R., Hut P., McMillan S., 1993, Nature, 364, 421
- Kumar A., Padmanabhan T., Subramanian K., 1995, MNRAS, 272, 544
- Lacey C., Cole S., 1993, MNRAS, 262, 627
- , 1994, MNRAS, 271, 676
- Lah P. et al., 2007, MNRAS, 376, 1357
- , 2009, MNRAS, 399, 1447
- Lahav O. et al., 2002, MNRAS, 333, 961
- Lanzetta K. M. et al., 1991, ApJS, 77, 1
- Larson D. et al., 2011, ApJS, 192, 16
- Larson R. B., 1974, MNRAS, 169, 229
- Lauer T. R. et al., 1998, AJ, 116, 2263
- Leitherer C. et al., 1999, ApJS, 123, 3
- Levan A. J. et al., 2011, ArXiv e-prints, 1104.3356
- Li Y. et al., 2007a, ApJ, 665, 187
- , 2007b, MNRAS, 379, 689
- Lilly S. J. et al., 1996, ApJ, 460, L1+
- Linder E. V., Jenkins A., 2003, MNRAS, 346, 573
- Liu X., Shen Y., Strauss M. A., 2011, ArXiv e-prints, 1104.3391
- Loeb A., 2010, Phys. Rev. D, 81, 047503
- Loeb A., Barkana R., 2001, ARA&A, 39, 19
- Loeb A., Rasio F. A., 1994, ApJ, 432, 52
- Loeb A., Wyithe J. S. B., 2008, Physical Review Letters, 100, 161301
- Lousto C. O., Zlochower Y., 2009, Phys. Rev. D, 79, 064018

- Lukić Z. et al., 2007, *ApJ*, 671, 1160
- Macciò A. V. et al., 2004, *Phys. Rev. D*, 69, 123516
- Madau P. et al., 1996, *MNRAS*, 283, 1388
- Madau P., Ferrara A., Rees M. J., 2001, *ApJ*, 555, 92
- Madau P., Haardt F., Rees M. J., 1999, *ApJ*, 514, 648
- Magliocchetti M. et al., 2000, *MNRAS*, 314, 546
- Magorrian J. et al., 1998, *AJ*, 115, 2285
- Makino J., 1997, *ApJ*, 478, 58
- Makino J., Aarseth S. J., 1992, *PASJ*, 44, 141
- Makino J., Ebisuzaki T., 1996, *ApJ*, 465, 527
- Makino J., Funato Y., 2004, *ApJ*, 602, 93
- Malhotra S., Rhoads J. E., 2004, *ApJ*, 617, L5
- Manchester R. N., 2008, in *American Institute of Physics Conference Series*, Vol. 983, 40 Years of Pulsars: Millisecond Pulsars, Magnetars and More, C. Bassa, Z. Wang, A. Cumming, & V. M. Kaspi, ed., pp. 584–592
- Mapelli M., Ferrara A., 2005, *MNRAS*, 364, 2
- Marconi A., Hunt L. K., 2003, *ApJ*, 589, L21
- Martin A. M. et al., 2010, *ApJ*, 723, 1359
- Martin C. L., 1999, *ApJ*, 513, 156
- McGaugh S. S. et al., 2000, *ApJ*, 533, L99
- McKee C. F., Ostriker E. C., 2007, *ARA&A*, 45, 565
- McLure R. J. et al., 2009, *MNRAS*, 395, 2196
- McLure R. J., Dunlop J. S., 2002, *MNRAS*, 331, 795
- McLure R. J. et al., 2010, *MNRAS*, 403, 960
- Meiksin A. A., 2009, *Reviews of Modern Physics*, 81, 1405

-
- Menou K., Haiman Z., Narayanan V. K., 2001, *ApJ*, 558, 535
- Merritt D., 2000, in *Astronomical Society of the Pacific Conference Series*, Vol. 197, *Dynamics of Galaxies: from the Early Universe to the Present*, F. Combes, G. A. Mamon, & V. Charmandaris, ed., pp. 221–+
- , 2006, *ApJ*, 648, 976
- Merritt D., Fridman T., 1996, *ApJ*, 460, 136
- Merritt D., Mikkola S., Szell A., 2007, *ApJ*, 671, 53
- Merritt D., Milosavljević M., 2005, *Living Reviews in Relativity*, 8, 8
- Merritt D., Poon M. Y., 2004, *ApJ*, 606, 788
- Merritt D., Szell A., 2006, *ApJ*, 648, 890
- Merz H., Pen U.-L., Trac H., 2005, *New Astron.*, 10, 393
- Mesinger A., Dijkstra M., 2008, *MNRAS*, 390, 1071
- Mikkola S., Aarseth S. J., 1990, *Celestial Mechanics and Dynamical Astronomy*, 47, 375
- Mikkola S., Valtonen M. J., 1990, *ApJ*, 348, 412
- Miller L. et al., 2006, *A&A*, 459, 43
- Miller R. H., 1983, *ApJ*, 270, 390
- Milosavljević M., Merritt D., 2001, *ApJ*, 563, 34
- , 2003a, *ApJ*, 596, 860
- , 2003b, in *American Institute of Physics Conference Series*, Vol. 686, *The Astrophysics of Gravitational Wave Sources*, J. M. Centrella, ed., pp. 201–210
- Miralda-Escudé J., Gould A., 2000, *ApJ*, 545, 847
- Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, *ApJ*, 530, 1
- Miralda-Escudé J., Ostriker J. P., 1990, *ApJ*, 350, 1
- Mitra S. et al., 2011, *ArXiv e-prints*, 1103.5828
- Mo H. J., Mao S., White S. D. M., 1998, *MNRAS*, 295, 319

- Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
- Monaco P., Fontanot F., Taffoni G., 2007, MNRAS, 375, 1189
- Monaghan J. J., 1992, ARA&A, 30, 543
- Morales M. F., Bowman J. D., Hewitt J. N., 2006, ApJ, 648, 767
- Moreno J., Giocoli C., Sheth R. K., 2008, MNRAS, 391, 1729
- , 2009, MNRAS, 397, 299
- Mortlock D. J. et al., 2011, Nature, 474, 616
- Muñoz J. A., Loeb A., 2008, MNRAS, 385, 2175
- Nagai D., Kravtsov A. V., 2005, ApJ, 618, 557
- Navarro J. F., Frenk C. S., White S. D. M., 1995, MNRAS, 275, 720
- Neistein E., Dekel A., 2008, MNRAS, 388, 1792
- Neistein E., Macciò A. V., Dekel A., 2010, MNRAS, 403, 984
- Norberg P. et al., 2002, MNRAS, 336, 907
- Noterdaeme P. et al., 2009, A&A, 505, 1087
- Obreschkow D., Rawlings S., 2009, MNRAS, 394, 1857
- Oesch P. A. et al., 2010, ApJ, 709, L16
- , 2009, ApJ, 690, 1350
- Oguri M., Taruya A., Suto Y., 2001, ApJ, 559, 572
- Oh S. P., Mack K. J., 2003, MNRAS, 346, 871
- Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
- Olive K. A., Steigman G., Walker T. P., 2000, Phys. Rep., 333, 389
- O'Meara J. M. et al., 2006, ApJ, 649, L61
- Ota K. et al., 2008, ApJ, 677, 12
- Ouchi M. et al., 2009a, ApJ, 706, 1136

-
- , 2009b, *ApJ*, 696, 1164
- Padmanabhan T., 2002, *Theoretical Astrophysics, Volume III: Galaxies and Cosmology*. Cambridge, UK: Cambridge University Press
- Peacock J. A., 1999, *Cosmological Physics*. Cambridge, UK: Cambridge University Press
- Peebles P. J. E., 1980, *The large-scale structure of the universe*. Princeton NJ: Princeton University Press
- , 1993, *Principles of Physical Cosmology*. Princeton NJ: Princeton University Press
- Peng C. Y. et al., 2006, *ApJ*, 640, 114
- Penton S. V., Shull J. M., Stocke J. T., 2000, *ApJ*, 544, 150
- Penton S. V., Stocke J. T., Shull J. M., 2004, *ApJS*, 152, 29
- Percival W., Miller L., 1999, *MNRAS*, 309, 823
- Percival W. J., 2001, *MNRAS*, 327, 1313
- Percival W. J., Miller L., Peacock J. A., 2000, *MNRAS*, 318, 273
- Peres A., 1962, *Physical Review*, 128, 2471
- Péroux C. et al., 2005, *MNRAS*, 363, 479
- Persic M., Salucci P., Stel F., 1996, *MNRAS*, 281, 27
- Peters P. C., 1964, *Phys. Rev.*, 136, 1224
- Pettini M. et al., 2001, *ApJ*, 554, 981
- Pieri M. M., Schaye J., Aguirre A., 2006, *ApJ*, 638, 45
- Pierpaoli E., 2004, *Physical Review Letters*, 92, 031301
- Popping A. et al., 2009, *A&A*, 504, 15
- Porciani C. et al., 1998, *MNRAS*, 298, 1097
- Power C., Baugh C. M., Lacey C. G., 2010, *MNRAS*, 406, 43
- Prasad P., Subrahmanya C. R., 2011, *ArXiv e-prints*
- Press W. H., Schechter P., 1974, *ApJ*, 187, 425

- Pretorius F., 2005, *Physical Review Letters*, 95, 121101
- Pritchard J. R., Furlanetto S. R., 2007, *MNRAS*, 376, 1680
- Pritchard J. R., Loeb A., 2008, *Phys. Rev. D*, 78, 103511
- Prochaska J. X., Herbert-Fort S., Wolfe A. M., 2005, *ApJ*, 635, 123
- Prochaska J. X., Wolfe A. M., 2009, *ApJ*, 696, 1543
- Purcell E. M., Field G. B., 1956, *ApJ*, 124, 542
- Quinlan G. D., 1996, *New Astron.*, 1, 35
- Quinlan G. D., Hernquist L., 1997, *New Astron.*, 2, 533
- Rao S. M., Turnshek D. A., Nestor D. B., 2006, *ApJ*, 636, 610
- Reed D. et al., 2003, *MNRAS*, 346, 565
- Reed D. S. et al., 2007, *MNRAS*, 374, 2
- Rees M. J., Ostriker J. P., 1977, *MNRAS*, 179, 541
- Richard J. et al., 2008, *ApJ*, 685, 705
- Richstone D. et al., 1998, *Nature*, 395, A14+
- Ripamonti E., 2007, *MNRAS*, 376, 709
- Roukema B., Valls-Gabaud D., 2000, in *Astronomical Society of the Pacific Conference Series*, Vol. 200, *Clustering at High Redshift*, A. Mazure, O. Le Fèvre, & V. Le Brun, ed., pp. 24–+
- Ryan-Weber E. V. et al., 2009, *MNRAS*, 395, 1476
- Ryu D. et al., 1993, *ApJ*, 414, 1
- Salucci P., Burkert A., 2000, *ApJ*, 537, L9
- Salvaterra R. et al., 2009, *Nature*, 461, 1258
- Samui S., Srianand R., Subramanian K., 2007, *MNRAS*, 377, 285
- , 2009, *MNRAS*, 398, 2061
- Samui S., Subramanian K., Srianand R., 2008, *MNRAS*, 385, 783

-
- , 2009, *New Astron.*, 14, 591
- Sargent W. L. W. et al., 1980, *ApJS*, 42, 41
- Sarkar P. et al., 2009, *MNRAS*, 399, L128
- Sasaki S., 1994, *PASJ*, 46, 427
- Saslaw W. C., Valtonen M. J., Aarseth S. J., 1974, *ApJ*, 190, 253
- Scannapieco E., 2005, *ApJ*, 624, L1
- Scannapieco E., Ferrara A., Madau P., 2002, *ApJ*, 574, 590
- Scannapieco E. et al., 2006, *MNRAS*, 365, 615
- Schaye J. et al., 2003, *ApJ*, 596, 768
- Schilizzi R. T., Dewdney P. E. F., Lazio T. J. W., 2008, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 7012, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*
- Schleicher D. R. G., Banerjee R., Klessen R. S., 2008, *Phys. Rev. D*, 78, 083005
- Schneider R. et al., 2008, *MNRAS*, 384, 1525
- , 2006, *MNRAS*, 369, 825
- Scott D., Rees M. J., 1990, *MNRAS*, 247, 510
- Seljak U., Slosar A., McDonald P., 2006, *JCAP*, 10, 14
- Sesana A. et al., 2004, *ApJ*, 611, 623
- Sethi S. K., 2005, *MNRAS*, 363, 818
- Shen Y., 2009, *ApJ*, 704, 89
- Sheth R. K., Mo H. J., Tormen G., 2001, *MNRAS*, 323, 1
- Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119
- , 2002, *MNRAS*, 329, 61
- Sigurdsson S., Hernquist L., 1993, *Nature*, 364, 423
- Sijacki D. et al., 2007, *MNRAS*, 380, 877

- Silk J., 1977, *ApJ*, 211, 638
- Simcoe R. A. et al., 2011, *ArXiv e-prints*, 1104.4117
- Simcoe R. A., Sargent W. L. W., Rauch M., 2004, *ApJ*, 606, 92
- Sobral D. et al., 2009, *MNRAS*, 398, L68
- Somerville R. S., Primack J. R., Faber S. M., 2001, *MNRAS*, 320, 504
- Songaila A., 1997, *ApJ*, 490, L1+
- , 2001, *ApJ*, 561, L153
- Spergel D. N. et al., 2003, *ApJS*, 148, 175
- Springel V., 2010, *ARA&A*, 48, 391
- Springel V., Hernquist L., 2003a, *MNRAS*, 339, 289
- , 2003b, *MNRAS*, 339, 312
- Springel V. et al., 2005, *Nature*, 435, 629
- , 2001, *MNRAS*, 328, 726
- Srianand R. et al., 2004, *Physical Review Letters*, 92, 121302
- , 2010, *MNRAS*, 405, 1888
- Stark D. P. et al., 2007, *ApJ*, 663, 10
- Steidel C. C. et al., 2000, *ApJ*, 532, 170
- Stern D. et al., 2005, *ApJ*, 619, 12
- Stone N., Loeb A., 2011, *MNRAS*, 412, 75
- Storrie-Lombardi L. J., Irwin M. J., McMahon R. G., 1996, *MNRAS*, 282, 1330
- Subramanian K., Padmanabhan T., 1993, *MNRAS*, 265, 101
- Sunyaev R. A., Zeldovich Y. B., 1972, *Comments on Astrophysics and Space Physics*, 4, 173
- Sutherland R. S., Dopita M. A., 1993, *ApJS*, 88, 253

-
- Swarup G., 1984, Giant meter-wavelength radio telescope—proposal, radio astronomy centre, tifr, india
- Swaters R. A., Madore B. F., Trewhella M., 2000, *ApJ*, 531, L107
- Taffoni G. et al., 2003, *MNRAS*, 341, 434
- Tanaka T., Haiman Z., 2009, *ApJ*, 696, 1798
- Tanvir N. R. et al., 2009, *Nature*, 461, 1254
- Taylor J. E., Babul A., 2001, *ApJ*, 559, 716
- Tegmark M. et al., 1997, *ApJ*, 474, 1
- Thoul A. A., Weinberg D. H., 1996, *ApJ*, 465, 608
- Tinker J. et al., 2008, *ApJ*, 688, 709
- Tonry J. L., 1984, *ApJ*, 279, 13
- Tremaine S. et al., 2002, *ApJ*, 574, 740
- Trenti M., Stiavelli M., 2009, *ApJ*, 694, 879
- Trimble V., 1987, *ARA&A*, 25, 425
- Tseliakhovich D., Hirata C., 2010, *Phys. Rev. D*, 82, 083520
- Tumlinson J., 2007, *ApJ*, 665, 1361
- Tyson J. A., Kochanski G. P., dell'Antonio I. P., 1998, *ApJ*, 498, L107+
- Vale A., Ostriker J. P., 2004, *MNRAS*, 353, 189
- Valtonen M. J., 1976, *A&A*, 46, 435
- Valtonen M. J. et al., 1994, *ApJS*, 95, 69
- van den Bosch F. C. et al., 1999, *ApJ*, 515, 50
- , 2000, *AJ*, 119, 1579
- Vázquez G. A., Leitherer C., 2005, *ApJ*, 621, 695
- Venkatesan A., Truran J. W., 2003, *ApJ*, 594, L1
- Verde L. et al., 2002, *MNRAS*, 335, 432

- Volonteri M., 2007, *ApJ*, 663, L5
- Volonteri M., Haardt F., Madau P., 2003, *ApJ*, 582, 559
- Volonteri M., Rees M. J., 2006, *ApJ*, 650, 669
- Warren M. S. et al., 2006, *ApJ*, 646, 881
- Weinberg S., 1972, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. Wiley-CVH
- Wetzel A. R., Cohn J. D., White M., 2009, *MNRAS*, 395, 1376
- Weymann R. J. et al., 1998, *ApJ*, 506, 1
- White M., 2002, *ApJS*, 143, 241
- White R. L. et al., 2003, *AJ*, 126, 1
- , 2005, *AJ*, 129, 2102
- White S. D. M., Frenk C. S., 1991, *ApJ*, 379, 52
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Windhorst R. A. et al., 2006, in *Bulletin of the American Astronomical Society*, Vol. 38, *Bulletin of the American Astronomical Society*, pp. 1188–+
- Wolfe A. M., Gawiser E., Prochaska J. X., 2005, *ARA&A*, 43, 861
- Wolfe A. M. et al., 1995, *ApJ*, 454, 698
- Wouthuysen S. A., 1952, *AJ*, 57, 31
- Wright E. L. et al., 1992, *ApJ*, 396, L13
- Wu K. K. S., Lahav O., Rees M. J., 1999, *Nature*, 397, 225
- Wyithe J. S. B., 2008, *MNRAS*, 388, 1889
- Wyithe J. S. B., Bolton J. S., Haehnelt M. G., 2008, *MNRAS*, 383, 691
- Wyithe J. S. B., Brown M. J. I., 2010, *MNRAS*, 404, 876
- Wyithe J. S. B., Loeb A., 2003a, *ApJ*, 590, 691
- , 2003b, *ApJ*, 595, 614

-
- , 2003c, *ApJ*, 588, L69
- , 2005, *ApJ*, 625, 1
- , 2007, *MNRAS*, 375, 1034
- Wyithe J. S. B., Loeb A., Geil P. M., 2008, *MNRAS*, 383, 1195
- Xu G., Ostriker J. P., 1994, *ApJ*, 437, 184
- Yadav J. et al., 2005, *MNRAS*, 364, 601
- Yan H., Windhorst R. A., 2004, *ApJ*, 600, L1
- Yang X., Mo H. J., van den Bosch F. C., 2003, *MNRAS*, 339, 1057
- Yang X. et al., 2005, *MNRAS*, 356, 1293
- Yoo J., Miralda-Escudé J., 2004, *ApJ*, 614, L25
- Yu Q., 2002, *MNRAS*, 331, 935
- Yunes N., Coleman Miller M., Thornburg J., 2011, *Phys. Rev. D*, 83, 044030
- Zauderer B. A. et al., 2011, *ArXiv e-prints*, 1106.3568
- Zentner A. R., 2007, *International Journal of Modern Physics D*, 16, 763
- Zentner A. R. et al., 2005, *ApJ*, 624, 505
- Zhang B., 2009, *Nature*, 461, 1221
- Zheng W. et al., 2009, *ApJ*, 697, 1907
- Zinnecker H., Yorke H. W., 2007, *ARA&A*, 45, 481
- Zwaan M. A., 2000, PhD thesis, PhD Thesis, Groningen: Rijksuniversiteit, 2000 152 p. Proefschrift, Rijksuniversiteit Groningen, 2000
- Zwaan M. A. et al., 2005, *MNRAS*, 359, L30
- , 2003, *AJ*, 125, 2842