MASSIVE BLACK HOLE REMNANTS OF THE FIRST STARS AND THEIR SIGNIFICANCE IN PRESENT-DAY GALACTIC HALOS

Ranty Rajat Islam

St Cross College, Oxford

May 2003

THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
AT THE UNIVERSITY OF OXFORD

Für Mama und Papa

Declaration

I declare that no part of this thesis has been accepted or is currently being submitted for any degree or diploma or certificate or any other qualification in this University or elsewhere.

This thesis is the result of my own work unless otherwise stated.

Some of this work has appeared in refereed journals:

Knebe A., Islam R.R., Silk J., 2001, MNRAS, 326, 109 (parts of chapter 2)
Islam R.R., Taylor J.E., Silk J., 2003, MNRAS, 340, 647 (parts of chapter 5 & 6)

MASSIVE BLACK HOLE REMNANTS OF THE FIRST STARS AND THEIR SIGNIFICANCE IN PRESENT-DAY GALACTIC HALOS

Ranty Rajat Islam

St Cross College, Oxford

Trinity Term 2003

THESIS SUBMITTED INT FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
AT THE UNIVERSITY OF OXFORD

Abstract

We investigate the possibility that present-day galaxies and their dark matter halos contain a population of massive black holes (MBHs) that form by hierarchical merging of the black hole remnants of the first stars in the Universe.

Some of the MBHs may be large enough or close enough to the centre of the galactic host that they merge within a Hubble time. We estimate to what extent this process could contribute to the mass of the super-massive black holes (SMBHs) observed in galactic centres today. The relation between SMBH and galactic bulge mass in our model displays the same slope as that found in observations. Many MBHs will not reach the centre of the host halo, however, but continue to orbit within it. In doing so MBHs may remain associated with remnants of the satellite halo systems of which they were previously a part. Using a semi-analytical approach that explicitly accounts for dynamical friction, tidal disruption and encounters with galactic disks, we follow the hierarchical merging of MBH systems and their subsequent dynamical evolution inside the respective host halos. In this context two types of dynamical processes are examined in more detail.

We predict the mass and abundance of MBHs in present-day galactic halos and also estimate the MBH mass accretion rates considering two different accretion scenarios. On this basis we determine the bolometric, optical and X-ray luminosity functions for the accreting MBHs using thin disk and advection dominated accretion flow models. Our predicted MBH X-ray emissions are then compared with observations of ultra-luminous X-ray sources in galaxies. We find that the slope and normalisation of the predicted X-ray luminosity function are consistent with the observations.

We also estimate the rate of gravitational wave events received from MBH mergers across all redshifts. At the end of their lives the first stars may explode in supernovae that are associated with gamma ray bursts (GRBs). Provided these are in principle detectable we have estimated the expected rate of events observed.

Acknowledgments

I am fortunate to have had the opportunity to be supervised by Joe Silk. I thank him for all his patience, continual encouragement and ideas which were a constant source of inspiration. Most of all I am grateful for the freedom I was given to develop and pursue my own ideas. This was initially intimidating, sometimes frustrating, frequently challenging but ultimately one of the most valuable experiences that I take away from my time in Oxford. Thanks are also due to James Taylor for all his support and the many things that I have learnt from hours of discussions we have had over the past two years and not least for providing the halo merger tree runs on which the numerical work, presented in these pages , is based. I also thank Greg Bryan for many stimulating discussions on first star formation and all his help with setting up the hydro simulations.

Life in Astro wouldn't have been half the fun were it not for the numerous distractions created by the wonderful people that I was fortunate enough to share an office with. I thank all of them and in particular Asim Mahmood, Martin Beer, Chris Blake, Becky Sigmon, Ewan Mitchell and Sugata Kaviraj. Thanks also go to Louise Griffiths, and Neil Clifford who was always at hand when help was needed with computing facilities. My stay in Oxford would not have been possible without the efforts of Geoff Smith. I owe him my gratitude for arranging my funding and college. I gratefully acknowledge receipt of a University graduate studentship.

Many other people made life in Oxford a wonderful time. In particular I would like to thank Phyllis Ferguson, Ximena Cobo-Santillan, Raeli Bronstein, Cortney Marabetta, Edward Conway, Raechel Ker, Pedro Ruiz-Castell, Sue Walters and Robert Behlsen.

Finally, I like to thank the people without whom none of this would have been possible nor worthwhile. If this thesis has any value it is due to them. This is Mohammad Al-Nassery whose faith and positive view of the world has always been an inspiration to me; my parents for their unwavering support as long as I can think; Shahana and Rana for being my two best friends; and finally Andrea who bravely put up with me, but without whom I would have given up long ago.

Contents

Declaration iv
Abstract vi
Acknowledgments vii
List of Tables xi
List of Figures xiii

1 Introduction **1**
1.1 A tale of cosmic elephants 2
1.2 The model 3
1.3 Structure of the thesis 4

2 Basics of Structure Formation **7**
2.1 The Paradigm 7
  2.1.1 Universal Expansion and the Cosmological Principle 8
  2.1.2 Origin and Description of Density Fluctuations 13
  2.1.3 Cold Dark Matter models 17
2.2 Gravitational Growth of density Perturbations 20
  2.2.1 Adiabatic vs iso-curvature perturbations 20
  2.2.2 Linear Regime 21
  2.2.3 Non-Linear Growth and Collapse of Objects 22
  2.2.4 Abundance of Non-Linear Objects 24
2.3 Evolution of the Baryonic Component 25
  2.3.1 Jeans Mass and baryonic collapse 25
  2.3.2 Baryonic Cooling and the Formation of Galaxies 27
2.4 Non-standard cosmologies - an example 28
  2.4.1 A primordial feature in the matter power spectrum 28
  2.4.2 Modified power spectra and their evolution 29
  2.4.3 Primordial features and the evolution of clusters 31

3 Massive black holes and galaxies **35**
3.1 SMBHs in galactic centres - observations 36
  3.1.1 The SMBH – bulge connection 37
  3.1.2 SMBH demographics 39
3.2 SMBHs in galactic centres - theoretical approaches 40
  3.2.1 SMBH binaries at galactic centres 44
3.3 Intermediate mass black holes 45
  3.3.1 The significance of IMBHs 45
  3.3.2 Formation scenarios 46

4 Massive Black Holes from Population III remnants 48
4.1 Baryonic collapse at z $\gtrsim$ 15 – analytical estimates . . . 49
4.1.1 Cooling of primordial molecular clouds . . . 49
4.1.2 Collapse of 1, 2, 3 σ peaks . . . 54
4.1.3 Qualitative features of the primordial stellar IMF . . . 57
4.2 Baryonic collapse at $z \gtrsim 15$ - hydrodynamical simulations . . . 58
4.2.1 Initial conditions and parameters . . . 58
4.2.2 First results . . . 59
4.3 Massive population III stars and black holes . . . 59
4.3.1 Formation of the first stars . . . 59
4.3.2 Radiative feedback from the first stars and its effect on the host cloud . . . 60
4.3.3 Supernova feedback from the first stars and its effect on the host cloud . . . 62
4.3.4 Effect of feedback on neighbouring clouds . . . 62
4.3.5 When does the epoch of first star formation end? . . . 66
4.3.6 End of stellar evolution and MBH formation . . . 68
4.4 Summary . . . 68

5 Hierarchical merging of galaxies and MBHs 70
5.1 Semi-analytical treatment of hierarchical merging and satellite dynamics . . . 71
5.1.1 Monte-Carlo merger histories . . . 71
5.1.2 Dynamical evolution of satellites and MBHs . . . 73
5.1.3 Inclusion of MBHs . . . 77
5.2 Tidal stripping and erosion of satellite density profiles . . . 78
5.2.1 Assumptions and initial conditions . . . 79
5.2.2 Satellite properties and their evolution . . . 80
5.2.3 Satellite moving on circular orbits at fixed host distance . . . 85
5.3 (S)MBH mergers and the galactic centre . . . 89
5.3.1 MBH inspiral and the effect on the galactic centre . . . 89
5.3.2 An upper limit to MBH merger efficiency . . . 95
5.4 Summary overview of modelling procedure and participating processes . . . 98
5.5 Summary . . . 99

6 MBHs in present-day Galactic Halos 101
6.1 Abundance and Mass of MBHs in galactic halos . . . 101
6.1.1 Abundance of MBHs in galactic halos . . . 101
6.1.2 Constraints on initial MBH mass function . . . 110
6.2 MBH mass accretion rates . . . 111
6.2.1 Bondi-Hoyle accretion . . . 111
6.2.2 Accretion environment: ISM vs. baryonic core remnants . . . 112
6.2.3 Distribution of mass accretion rates . . . 113
6.3 Abundance and mass of SMBHs . . . 118
6.3.1 SMBH from hierarchical merging of remnant MBHs . . . 118
6.3.2 SMBH growth from gas accretion . . . 120
6.4 Summary . . . 122

7 Observations I: Present Day Abundance of MBHs 124
7.1 Optical and X-ray emissions from accreting MBHs 125
7.2 Bolometric luminosity 126
  7.2.1 Bolometric luminosity for accretion from the host ISM 126
  7.2.2 Bolometric luminosity for accretion from baryonic core remnants 128
7.3 Accretion processes 131
  7.3.1 Accretion geometries: spherical vs. disk accretion 131
  7.3.2 Accretion mechanisms I: Thin disk accretion 132
  7.3.3 Accretion mechanisms II: Advection dominated accretion flows (ADAFs) . 133
7.4 Optical emission 138
7.5 X-rays 143
7.6 Optical, UV and X-ray signatures of sources in Earth-centred volumes 143
7.7 Comparison with observations 145
  7.7.1 Observations of ultra-luminous X-ray sources (ULXs) 146
  7.7.2 Baryonic core accretion at the centre of dwarf galaxies and globular clusters ?154
  7.7.3 Sources in and around the Milky-Way 155
7.8 Discussion and summary 157

8 Observations II: Redshift Evolution of MBH Abundance 161
8.1 Gravitational waves from SMBH-MBH Mergers 161
  8.1.1 Gravitational Wave Amplitude 162
  8.1.2 Frequency of gravitational Waves 163
  8.1.3 MBH Merger Efficiency 164
  8.1.4 Rate of SMBH-MBH mergers 172
  8.1.5 Detections and the distribution of strain amplitudes 176
8.2 Observations of GRBs from exploding population III stars at high z 178
  8.2.1 Total number of GRBs 179
  8.2.2 Rate of GRBs received 180
8.3 Other detections 183
  8.3.1 Gamma rays from MBH mergers 183
  8.3.2 Flares from the tidal disruption of stars 184
  8.3.3 Microlensing 185
8.4 Summary 185

9 Summary and Outlook 188
9.1 First star formation and seed MBHs 188
9.2 Hierarchical merging of MBHs 189
9.3 Observational signatures 191
9.4 A note on recent data from WMAP and QSO observations 193
9.5 Outlook and future work 194

Index of symbols used 196

References 198

List of Tables

2.1 ΛCDM cosmological model used throughout this work 20
2.2 Non-standard cosmologies: Specifications of features in the power spectrum 29

4.1 Number density of peaks *collapsing* at redshifts $z = 25$ and $= 20$ 66

5.1 Initial conditions: Seed MBH masses and heights of collapsing peaks 78
5.2 Parameters of the host and satellite halos. 86

6.1 Total number of MBHs in halo 103
6.2 Total mass in halo MBHs and mass of central SMBH 105
6.3 MBH abundance in Earth-centred volumes 109

7.1 Overview of X-ray observations 148

List of Figures

2.1 Input spectra for simulations. 30
2.2 Evolution of cluster abundance: SCDM vs. OCDM 32
2.3 Evolution of cluster abundance: SCDM models with added features 33
2.4 Analytical estimate of cluster abundance evolution for OCDM and SCDM 34

3.1 SMBH – bulge relation: SMBH vs bulge magnitude and velocity dispersion 38

4.1 Gas cooling rates for early objects 51
4.2 Halo cooling mass, virial masses and virial radii 52
4.3 Halo virial temperatures & binding energies 53
4.4 Fraction of mass contained in density peaks of height ν_{pk} and above 55
4.5 Correlation functions of high σ peaks 64
4.6 Average fraction of peaks that lie within a comoving distance r_{fp} of another one . 65

5.1 Halo progenitor masses vs redshift 72
5.2 Satellite tidal radius vs distance from host 81
5.3 Tidal erosion of satellite density and circular velocity profiles – analytical 86
5.4 Time evolution of satellite tidal radius and bound mass 87
5.5 Tidal erosion of satellite density and circular velocity profiles – comparison with simulations 88
5.6 Flow diagram of the computational model 100

6.1 Galactic halo MBH abundance - for seed masses 260 $M_\odot$ 102
6.2 Galactic abundance of all/'naked' MBHs for different seed masses 104
6.3 Radial MBH distribution in halos - model A 106
6.4 Cumulative radial distribution of MBHs in halos - models A,C,D 108
6.5 Abundance of MBH accretion rates in galactic halo 114
6.6 Redshift dependence of baryonic core accretion rates 115
6.7 Accretion rates vs MBH mass for all halos and models A, C, D 117
6.8 Central SMBH mass vs. bulge mass 119

7.1 Bolometric Luminosity function for ISM and baryonic core accretion - model C . . 127
7.2 Projected bolometric luminosity function - models A, C and D 129
7.3 Emission spectra from ADAF and thin disk accreting MBHs 137
7.4 B - band luminosity function - model A 139
7.5 Projected B and V - band luminosity function - models A, C and D 141
7.6 Projected X-ray luminosity function - models A, C and D 142
7.7 Luminosities of sources in Earth-centred volumes - models A, C and D 144
7.8 B,V-band luminosities of sources in Earth-centred volumes - models A, C and D . 145

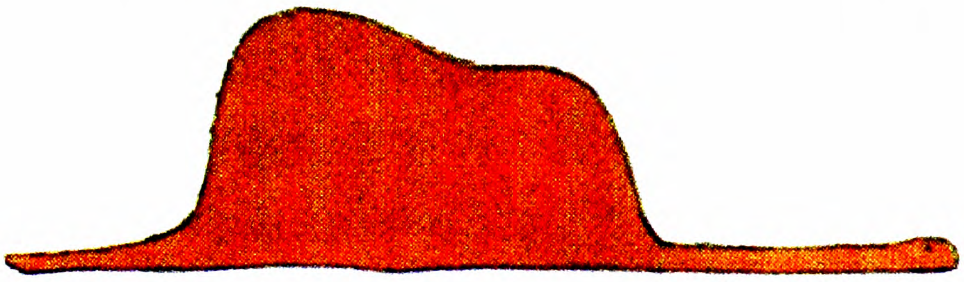
7.9 Comparison with observations of X-ray point sources 150
7.10 Comparison with observed off-centre ULXs in galaxies - luminosity functions . . . 151
7.11 Comparison with observed off-centre ULXs in galaxies - radial distribution 152
7.12 Cumulative X-ray luminosity function for Milky-Way sized halo 156

8.1 Host mass versus central SMBH mass 167
8.2 Ratio of dynamical friction time to orbital period 169
8.3 Redshift dependent MBH merger rate per halo 171
8.4 Redshift dependent MBH merger rate per halo - contribution from different strain amplitudes 173
8.5 Redshift dependent rate of received MBH merger events - contribution from different strain amplitudes 175
8.6 Total number of received events - distribution of strain amplitudes 177
8.7 First stars as GRB sources: number density of MBC halos 181
8.8 Number of GRB events received per year originating from redshift z to $z +$ dz . . 182

Chapter 1

Introduction

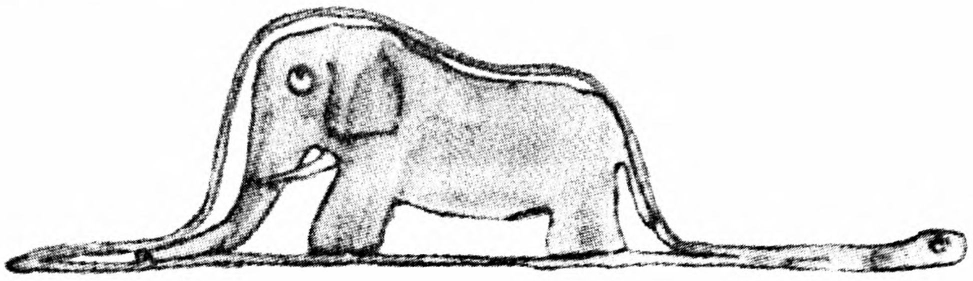
In August 1998 at a conference in Munich, the Japanese cosmologist Naoshi Sugiyama concluded his talk on the Cosmic Microwave Background radiation showing this slide



and left the audience to guess for a while at the meaning of what looks like a hat to ordinary people (or like a pathological spectral line profile to ordinary astrophysicists). The author of the talk did in fact suggest that the picture shares some similarities with the curve representing the spectrum of fluctuations in the cosmic microwave background (CMB) radiation that we have begun to observe from all over the sky. This radiation is believed to be the faint echo of the Big Bang that is supposed to have brought our Universe into existence some 13,000 million years ago.

What is essential is invisible to the eye – and what is true in life also holds for the Universe. Theoretical and observational efforts, especially over the last ten years or so, have revealed that there is certainly more to the CMB than meets the eye. The ripples in the CMB across the sky contain possibly the most succinct record of the many things that happened in our Universe when it was very young. Fundamental properties of our Universe, such as its size, age and 'weight' are now relatively well known, thanks to the observation and interpretation of the background radiation. The latest impressions of the CMB as collected by the WMAP satellite reveal a number of distinctive features that tell the story of the early Universe in unprecedented detail.

Regarding the conference slide someone pointed out a distinctive feature at the 'right end of the plot' and it was soon agreed that this was not what it seemed to be – certainly not a spectral line profile. However, despite the careful observation in this case it is questionable whether any amount of scientific intuition could have revealed the true nature of what hides underneath the curve on the previous page:



The conference speaker lifted this 'plot' of an elephant-digesting snake from a fine piece of research into the meaning of life as conducted by Antoine de Saint-Exupéry in his tale of the *Little Prince.*

1.1 A tale of cosmic elephants

The CMB is certainly one very important example of how our present Universe keeps a memory of its tumultuous past. In a way the background radiation adorns the stage for a play that has by now been going on for a very long time. In this thesis we attempt to relate a story about the possible origin of one of the more experienced and – judging by their size – somewhat overweight actors in this play.

These are Super-Massive Black Holes (SMBHs), compact dark objects some 10 to 1000 times larger in size than our sun but several hundred million times more massive. From observations of our Milky-Way and other galaxies it now seems certain that SMBHs are present at many if not most galaxy centres. It is not inappropriate to think of these SMBHs as the cosmic equivalents of the elephant from Saint-Exupéry's *Little Prince.* They are massive, tucked away in the dense inner parts of galactic centres, yet capable of making a big noise. SMBHs have not only been observed in nearby galaxies. Accretion of large amounts of material onto a lot more hungry SMBH specimens is thought to be the driving force behind the extremely bright emissions we observe from very distant galaxies. As yet, there is no generally agreed theory as to where these SMBHs come from. However, there is a growing consensus that SMBHs and their host galaxies have in some way evolved and grown together. This gives rise to the possibility that SMBHs have in a sense retained – again in true elephant fashion – an extensive memory of their past evolution. The mass and distribution of, as well as the activity around SMBHs may tell us more about the traumatising events that their host galaxies went through before assuming their currently observed appearance.

In this thesis we propose a model for the origin of the observed SMBHs. This model makes a number of predictions, the most important of which is the presence of a number of much lighter BHs (but still as massive as 100 times the mass of the sun) – baby elephants if one likes – in the neighbourhood of galaxies.

1.2 The model

We are carrying further a recent idea suggesting that SMBHs have grown from the merging of much lighter *MBHs* (Massive Black Holes) which are the remnants of the first stars in the Universe.

This idea is motivated by recent theoretical advances which indicate that *a)* the first stars in the Universe may have formed very shortly after the Big Bang, *b)* they were possibly 100 to 1000 times more massive than the sun, and *c)* that these stars could have turned into black holes of similar mass at the end of their lives.

Theoretical arguments suggest that the first stars could have formed inside dense clouds of material very early on, when the Universe was only about 1/100 of its present age. Stars produce most of the large variety of chemical elements we know, including carbon and oxygen, without which life on Earth simply would not exist. By definition, the first stars formed and lived in an environment where these elements were not yet present. Theoretical work and large computer simulations indicate that the absence of these elements may have modified certain processes in the early Universe in such a way that caused the first stars to become very massive. At the end of their life times these stars are likely to have left behind a black hole with a mass similar to that of the original star, i.e. some 100 or 1000 solar masses.

Our current view of how galaxies formed is that they have grown through the hierarchical merging of smaller 'proto' galaxies, which in turn have formed from still smaller groups of material and so on. The beginning of this hierarchy may have been the same clouds in which the first stars formed. The implication is then, that the black hole remnants of the first stars have also merged hierarchically to form ever more massive black holes and ultimately build the SMBHs observed today.

In this thesis we describe the results from an investigation that explicitly follows the hierarchical merging of the black hole remnants and their host clouds. We find that MBH merging can account for a significant part of the mass of presently observed SMBHs. More importantly, this merging model predicts the existence of hundreds or thousands of light MBHs throughout and around galaxies. A major part of our investigation is to estimate any observable trace of these MBHs. As they cruise through galaxies they accrete any material that crosses their path. This material may even be tightly wrapped around the MBHs and travel with them, much like a 'portable' fuel supply. These supplies could just be remnants of the dense clouds of which the MBHs were previously a part. We predict the amounts of visible light and X-rays emitted by the material as it gets sucked into the MBHs. It appears that some of these emissions may have already been seen. The MBH merging process itself may be observed through the gravitational radiation that is emitted when any two MBHs merge. We have estimated how many merger events per year could be detected in this way. It is possible that there were huge explosions (gamma ray bursts) when the first stars turned into MBHs. Provided traces of these can be seen today, we estimate how many of these would be observed per year.

Hopefully, observations in the future will help decide whether or not this model provides a reasonable explanation for where the cosmic elephants come from.

1.3 Structure of the thesis

In chapter 2 we give a brief account of cosmological structure formation with particular emphasis on the hierarchical nature of the process. This is followed in chapter 3 by a review of observations and theoretical models to explain the presence of SMBHs as well as their relation to their host galaxies.

First stars and massive black holes

In chapter 4 we present analytical calculations to estimate the number and clustering of mini-halos where the first stars formed as well as the redshift of when this occurred. We also address the question of whether and to what extent the first stars forming in some halos could have suppressed star-formation in nearby mini-halos.

Hierarchical merging of massive black holes

In chapter 5 we present the semi-analytical approach we have used to investigate the merging process. This is a code which combines a hierarchical merger tree algorithm with explicit prescriptions for the dynamical evolution of merged halos inside larger host halos. In this context we also discuss some of the particular dynamical processes at work.

A main result is the determination of the abundances and accretion rates of MBHs in present-day galactic halos. This is the subject of chapter 6.

Observational signatures

Observational signatures of MBHs in our model come in two categories. In chapter 7 we use two accretion models to compute the optical and X-ray signatures of accreting MBHs in galaxies and their dark matter halos today. We then compare these with observations of luminous X-ray sources in nearby galaxies.

At high redshifts MBH formation and evolution could be probed via gravitational waves from MBH mergers and possibly by gamma ray bursts associated with the collapse of the first stars into black holes. Our estimates for both are presented in chapter 8.

In chapter 9 we sum up our results and discuss them briefly in the light of the most recent observations of the cosmic microwave background, before concluding with a few remarks on extensions of our work and future prospects.

A word on nomenclature – the acronyms BH, MBH, IMBH, SMBH appear frequently throughout this work and denote black holes in different mass ranges. **BH** refers to a stellar mass black hole of up to 100 $M_\odot$. **IMBH** stands for intermediate mass black hole in the mass range from $10^2 - 10^5$ $M_\odot$. This is used primarily to distinguish it from black holes in other mass ranges, otherwise we simply refer to IMBHs as **MBHs**. Finally, as mentioned above, **SMBHs** denote super-massive black holes in the mass range $10^6 - 10^9$ $M_\odot$, like those found in the nuclei of nearby galaxies.

Chapter 2

Basics of Structure Formation

In this chapter we give a brief general introduction to cosmological structure formation. While the aim is to provide a self-contained account of the relevant concepts and physics the focus will be on those aspects that will naturally lead on to the topics to be dealt with subsequently.

In section 2.1 we start by describing the 'paradigm' that underlies theory and interpretation of observations in cosmology today. In doing so we distinguish the term 'paradigm' as the overarching conceptual framework within which researchers deal with the Universe, from the term 'Cosmological Model'. The latter is part of the former and we use it as a collective term for a set of parameters that quantify or otherwise specify a range of constituents of the paradigm, such as cosmological expansion, matter density etc. In particular we introduce the family of Cold Dark Matter (CDM) Cosmological Models which form the basis of our work. Section 2.2 addresses the gravitational growth of density inhomogeneities in the linear and non-linear regimes up to final collapse and how to estimate the abundance of collapsed objects. In CDM models gravitational growth is dominated by dark matter and the presence and behaviour of baryons only become important inside dark matter structures that have already collapsed or are in the process of doing so. The evolution of the baryonic component in these circumstances is the subject of section 2.3. Finally, in section 2.4 we present a particular type of non-standard cosmological model.

2.1 The Paradigm

One of the many remarkable things about the Universe is that when viewed as a physical system we as observers are inevitably a part of it. This places the Universe in unique opposition to any other physical system. To be more precise this is the case as long as we view a system as embedded in a continuous space-time manifold and as having physical properties that in principle can be determined individually and uniquely independent of any given observer [1]. In Newtonian mechanics Euclidean 'absolute' space endowed with a single absolute time throughout provided the immutable background structure for everything that was present or happened within it. Properties like position and velocity of individual entities inside were defined with respect to that of others or the background stars considered at rest relative to absolute space (c.f Newton's water bucket experiment). In this way Newton stated an explicit preference for a particular inertial frame of reference, but exactly how absolute space is being defined here remains somewhat obscure especially on cosmological scales. Ernst Mach abandoned absolute space and instead linked the inertial properties of objects to the total gravitating mass in the Universe. Einstein termed this 'Mach's Principle'. By linking the seemingly intrinsic properties of local objects within a system to a global property attached to it, Mach introduced a degree of 'self-referentiality', which proved very useful conceptually when dealing with the Universe as a whole where no recourse to an absolute and independent external observer is possible. This idea had a strong impact on Einstein and is ultimately reflected in his general theory of relativity in which the notions of matter and space (-time) are intimately linked based on his 'Equivalence Principle'.

2.1.1 Universal Expansion and the Cosmological Principle

Einstein believed that Mach's principle could lead to contradictions if no further assumptions were made about the underlying geometry of the Universe. Imagine a Universe containing all its mass within a localised region. If this Universe was infinite and asymptotically flat, then in principle a body could move away arbitrarily far from this material region, but according to the theory of relativity would still keep its inertial properties. This contradicts Mach's principle.

---

[1] This is precisely not how a physical system is characterised in the Quantum Interpretation. Here properties are not intrinsic or unique; they only 'come into being' as 'observables' when a measurement is performed, and as such are neither independent of the outside observer undertaking the measurement, nor as it turns out of one another. In this sense our inability to view the Universe as independent observers is only the most obvious manifestation of the interdependence that is present on all levels between everything within it. Furthermore the values returned through observables reveal a quantum system to be quantised and to the extent they make reference to space and time also to be non-localised. In fact, both call into question the applicability of the concepts of space and time themselves to quantum systems and make it difficult to reconcile the quantum interpretation with a continuous space-time manifold. Except when dealing with the origin of the density perturbations that gave rise to structure in the Universe we will consider the quantum interpretation as not applicable throughout this work.

Einstein realised that this could be remedied if one assumed that the Universe is closed and thus finite.

To be able to solve the field equations of his general theory of relativity at all Einstein introduced further simplifying assumptions, namely that the Universe be homogeneous and isotropic. In addition, he assumed that the Universe is static. It appears that it is very difficult to do without the first two assumptions which constitute what is known as the 'Cosmological Principle'. This states that the Universe is homogeneous and isotropic at any given cosmic time. The assumption of a static Universe, however, appears to have been based on philosophy rather than mathematical necessity. In fact it constrains the solutions to the field equations to an extent that he had to invoke an additional entity, the cosmological constant, to ensure the solution would describe a static Universe. de Sitter showed that solutions to the Einstein equations could be obtained even for empty space when no matter was present at all. According to Mach's principle this renders the concept of inertial properties meaningless, but the de Sitter model also had the characteristic that it was not static but expanded exponentially, raising the possibility that expansion could be a general possibility to consider. Einstein's initial aversion against a static Universe was not entirely unfounded also on mathematical grounds; a non-static Universe could in principle involve the crossing of world-lines in which case the solution of the field equations would get very complicated, too. In the uniformly expanding de Sitter Universe this difficulty did not occur, world lines would not cross but move away from one another uniformly.

This possibility gained firm support when E. Hubble observed that stars appeared to be receding faster from us the further away they were. In particular, their average recession velocity v seemed to be directly proportional to their distance r

$$v = Hr \tag{2.1}$$

where H is the Hubble constant.

We can now introduce a comoving coordinate system that accounts for the expansion. In this system a comoving distance x is obtained by dividing the physical distance r by the expansion factor a, to which we will subsequently refer as the cosmic scale factor.

$$x \equiv \frac{r}{a} \tag{2.2}$$

Differentiating this with respect to time and substituting into eq. (2.1) we get

$$Hr = a\dot{x} + x\frac{\dot{a}}{a^2} \tag{2.3}$$

For a star that follows the Hubble expansion - or Hubble flow - exactly, $\dot{x} = 0$ by definition and thus

$$H = \frac{\dot{a}}{a} \tag{2.4}$$

The Hubble constant is also commonly expressed in dimensionless form as $h = H/100$ km s^{-1} Mpc^{-1}.

In the 1930's Robertson and Walker derived in explicit mathematical form the geometry underlying a homogeneous, isotropic and potentially uniformly expanding Universe. This is the well-known Robertson-Walker (RW) metric

$$\mathrm{d}s^2 = c^2\ \mathrm{d}t^2 - S(t)^2 \left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2(\ \mathrm{d}\theta^2 + sin^2\theta\ \mathrm{d}\phi^2)\right] \tag{2.5}$$

Here $\mathrm{d}s^2$ is the line element, c is the speed of light, t cosmic time, $S(t)$ the time dependent cosmic expansion rate and r, θ and ϕ are comoving spherical coordinates. k parametrises the spatial curvature and is $k = -1, 0, 1$ for a Universe that is open, flat or closed respectively. For this metric the Einstein field equations yield two non-trivial equations, which allow us to determine the form of the time dependence in $S(t)$ that governs the dynamics of the cosmological expansion. These are

$$2\frac{\ddot{S}}{S} + \frac{\dot{S}^2 + kc^2}{S^2} = \frac{8\pi G}{c^2}T_i^i \tag{2.6}$$

$$\frac{\dot{S}^2 + kc^2}{S^2} = \frac{8\pi G}{3c^2}T_0^0 \tag{2.7}$$

where G is the gravitational constant and dots denote differentiation with respect to time. T_i^i and T_0^0 are the diagonal spatial and time components of the stress-energy tensor T_μ^ν. For the RW metric these take the form

$$T_i^i = -p \tag{2.8}$$

$$T^0_0 = \rho c^2 \tag{2.9}$$

where p is the pressure and ρ the matter density. We can combine equations (2.6) and (2.7) to obtain

$$\frac{\ddot{S}}{S} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) \tag{2.10}$$

A. Friedmann considered a particular set of models of a Universe filled with pressure-less matter and no electro-magnetic radiation. In this case $p = 0$. If in addition we define the cosmic scale factor mentioned above to be unity at the present time, i.e. $a(t) = S(t)/S(t_0)$, equations (2.6), (2.7) and (2.10) then become

$$2\frac{\ddot{a}}{a} + \frac{\dot{a}^2 + kc^2}{a^2} = 0 \tag{2.11}$$

$$\frac{\dot{a}^2 + kc^2}{a^2} = \frac{8\pi G\rho}{3} \tag{2.12}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) \tag{2.13}$$

These are the Friedmann equations.

For a flat Universe $k = 0$ and eq. (2.12) yields

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} \tag{2.14}$$

Recognising that the LHS is just the square of the Hubble constant we get for the present day

$$\rho_0 = \frac{3H_0^2}{8\pi G} \equiv \rho_{crit} \tag{2.15}$$

ρ_{crit} is the 'critical' density for which the Universe is exactly flat and therefore the density that is needed to just close the Universe. The Friedmann equations can now easily be extended to include Einstein's addition of a cosmological constant Λ to the stress-energy tensor, as well as other contributions to the matter density. All these can be written in dimensionless form, for example from eq. (2.12)

$$\frac{kc^2}{H_0^2 a(t)^2} = [\Omega_m(t) + \Omega_r(t) + \Omega_\Lambda(t)] - 1 \tag{2.16}$$

where $\Omega_{m,r} = \rho_{m,r}/\rho_{crit}$ account for matter and radiation energy density and $\Omega_\Lambda = \Lambda/3H_0^2$. This implies that for a flat Universe we simply have $\Omega_m + \Omega_r + \Omega_\Lambda = 1$. Hereafter, unless an explicit time dependence is mentioned, the Ω parameters refer to present day quantities. The scale factor dependence of the Hubble constant is then

$$\frac{H(t)}{H_0} = \left[\frac{\Omega_m}{a^3} + \Omega_\Lambda + \frac{\Omega_r}{a^4} - \frac{kc^2}{H_0 a^2}\right]^{1/2} \tag{2.17}$$

As can be seen the matter and photon number densities simply scale as a^{-3}. However, in the case of radiation we have an additional factor a^{-1} arising from the redshift of individual photons because their wavelength gets stretched by the scale factor, too.

The latter is an example of the cosmological redshift, which can be derived by considering the RW line element eq. (2.5) for photons. Since these travel at the speed of light, $\mathrm{d}s = 0$, and so for a light signal travelling along the radial direction

$$c\ \mathrm{d}t = \pm\frac{a\ \mathrm{d}r}{(1 - kr^2)^{1/2}} \tag{2.18}$$

where we have also used the scale factor a.

Now consider a star following the Hubble flow. If the star at position r_1, say, emits a photon at time $t = t_1$, which is being received at earth at the present time t_0 then the integral along the travel path is

$$\int_{t_1}^{t_0} \frac{c\ \mathrm{d}t}{a(t)} = \int_0^{r_1} \frac{\mathrm{d}r}{(1 - kr^2)^{1/2}} \tag{2.19}$$

This uses the negative RHS of eq. (2.18) since the signal is travelling towards us at $r = 0$. For a subsequent signal emitted at $t_1 + \Delta t_1$ and received at $t_0 + \Delta t_0$

$$\int_{t_1+\Delta t_1}^{t_0+\Delta t_0} \frac{c\ \mathrm{d}t}{a(t)} = \int_0^{r_1} \frac{\mathrm{d}r}{(1 - kr^2)^{1/2}} \tag{2.20}$$

If $a(t)$ changes on a timescale much larger than $\Delta t_{0,1}$ then we can linearly expand eq. (2.20) to first order, subtract eq. (2.20) from it and rearrange to obtain

$$\frac{\Delta t_0}{\Delta t_1} = \frac{1}{a} \equiv 1 + z \tag{2.21}$$

This defines the cosmological redshift z. We can now re-express the evolution of cosmological parameters in terms of z. For the matter density parameter, for instance, we get

$$\Omega_m(z) = \frac{\Omega_m[1+z]^3}{\Omega_m[1+z]^3 + \Omega_\Lambda + (1-\Omega_m-\Omega_\Lambda)[1+z]^2} \tag{2.22}$$

where we have omitted Ω_r, as the contribution from radiation quickly becomes negligible after the epoch of matter radiation equality.

An important relation that we will use repeatedly is the relation between redshift on the one hand and cosmic time as well as comoving distance on the other. If two events are separated by an infinitesimal redshift interval $\mathrm{d}z$ apart at redshift z, this corresponds to a differential time interval

$$\begin{aligned} \frac{\mathrm{d}t(z)}{\mathrm{d}z} &= -H(z)^{-1}(1+z)^{-1}\ \mathrm{d}z \qquad (2.23) \\ &= -\left(\frac{1}{H_0}\right)(1+z)^{-1}[\Omega_m(1+z)^3 + \Omega_\Lambda + (1-\Omega_m-\Omega_\Lambda)(1+z)^2]^{-1/2} \end{aligned}$$

where we have also restated the time dependence of the Hubble constant eq(2.17) in terms of the redshift, z. For a flat Universe, $\Omega = 1$ and

$$\frac{\mathrm{d}t(z)}{\mathrm{d}z} = \left(\frac{1}{H_0}\right)(1+z)^{-1}[\Omega_m(1+z)^3 + \Omega_\Lambda]^{-1/2} \tag{2.24}$$

Similarly the comoving distance is related to the redshift as

$$\frac{\mathrm{d}D(z)}{\mathrm{d}z} = c\,(1+z)\frac{\mathrm{d}t}{\mathrm{d}z} \tag{2.25}$$

The dynamics of the universal expansion as outlined above is the basis for our work described in later chapters.

2.1.2 Origin and Description of Density Fluctuations

In the previous section we have introduced the cosmological parameters governing the dynamics of a uniformly expanding homogeneous, isotropic Universe.

However, while the global dynamics of our Universe appears to be well described in this way, we know that at least small inhomogeneities in the matter energy distribution must have been present if we are to explain the present ‘clumpiness’ of the Universe that is manifest in the stars, galaxies and clusters that we observe. This is if we assume that these have grown mainly by gravitational instability (c.f. section 2.2) from small density perturbations in the past. Observations of the *cosmic microwave background* (CMB) radiation that are essentially a snapshot of the matter distribution in the early Universe long before structures formed, indicate that these small density perturbations were indeed present.

Origin of density perturbations

There are several theories for the origin of those small density fluctuations that are superimposed on an otherwise homogeneous matter distribution. They are mostly variations or combinations of two generic models referred to as *topological defects* and *inflation* models.

In the topological defects model the very early Universe is supposed to have undergone a phase transition. Density perturbations here correspond to the defects at locations in space where phase transitions, triggered at and propagating away from often widely spaced 'nucleation' points, meet.

Inflation models maintain that at a very early stage , about 10^{-34} seconds after the Big Bang , the Universe underwent a phase of inflation during which it expanded exponentially and thereby stretched random quantum fluctuations to macroscopic perturbations on cosmological scales. These constitute the random initial perturbations in the primordial matter density field. A generic prediction of many although not all variants of inflationary models is that these fluctuations are Gaussian distributed and so reflect directly the Gaussian nature of the quantum fluctuations before inflation. This is essentially confirmed by observations of the CMB radiation and we shall work with the assumption that the field is Gaussian throughout.

The latter is currently favoured both on observational and theoretical grounds and hereafter we will exclusively focus on this as the source of the initial perturbations. [2]

---

[2] Both defects and inflation models can be embedded in a higher dimensional context, such as the *Brane World* scenario which has received considerable interest recently. Originally motivated by the quest for the unification of gravity with the other three fundamental forces, brane world scenarios are also of particular interest in cosmology as they potentially offer a radically new view of the nature of what constitutes the cosmological constant ('dark energy').

Description of density perturbations

We will now briefly introduce the basic statistical tools to describe the fluctuations. Since the fluctuations are small it is suitable to represent them as a fractional deviation δ from the homogeneous matter density background that pervades the Universe

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \overline{\rho}}{\overline{\rho}} \tag{2.26}$$

where we have used comoving spatial coordinates and omitted the explicit cosmic time dependence in all quantities involved. In particular we will be looking at the problem in terms of the amplitude of fluctuations present on any given range of scales. This information is contained in the coefficients $\delta_{\mathbf{k}}$ of the Fourier decomposition of $\delta(\mathbf{x})$

$$\delta_{\mathbf{k}} = \int \mathrm{d}^3x \; \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} \tag{2.27}$$

The Fourier pair $\delta_{\mathbf{x}}$ and $\delta(\mathbf{k})$ describe the components of the spectrum of density fluctuations in real space and k space respectively. They are the basic ingredients of any description of the density field as a whole. One of the most important ones is the two-point correlation function, also referred to as the *auto-correlation function* as it correlates two objects from the same field

$$\xi(\mathbf{r}) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle \tag{2.28}$$

We will subsequently refer to this simply as the 'correlation function'. What we referred to as 'objects' can be density perturbations, galaxies etc.; in principle any type of object whose collective spatial distribution can be described by a random statistical field. Physically $\xi(\mathbf{r})$ is a measure of the conditional probability of finding an object in a volume element δV_1, given that there is one in another volume element δV_2 at distance $\mathbf{r}_{12}$

$$\delta^2 P = n_{obj}^2 \left[1 + \xi(r_{12})\right] \delta V_1 \delta V_2 \tag{2.29}$$

where n_{obj} is the average number density of objects. The k space equivalent of this is the *power spectrum* $P(k)$ of density fluctuations

$$P(k) \equiv \langle|\delta_k|^2\rangle \tag{2.30}$$

$$= \int \mathrm{d}^3 r\ \xi(\mathrm{r}) e^{i\mathbf{k}\cdot\mathbf{r}} \tag{2.31}$$

where the second equality only holds if the density field is gaussian. This in turn is related to the correlation function as

$$\xi(r) = \frac{1}{(2\pi)^3} \int P(k) \frac{\sin kr}{kr} 4\pi\ k^2 \mathrm{d}k \tag{2.32}$$

Generally we assume that the cosmological density field is isotropic and in all the integrals we can therefore simply replace $\mathrm{d}^3 k$ with $= 4\pi k^2\ \mathrm{d}k$.

To relate the field and derived quantities with observations we typically 'smooth' the field using a smoothing (or 'window') function W and then determine the variance of the smoothed field

$$\sigma^2(r) = \frac{1}{2\pi^2} \int P(k) W^2(kr)\ k^2\ \ \mathrm{d}k \tag{2.33}$$

where W is typically chosen to be a gaussian or top-hat function in real space. Together with their Fourier counterparts,$\hat{W}$, these are given by

$$W = \frac{1}{(2\pi)^{3/2} R_G^3} \exp\left[-\frac{r^2}{2R_G^2}\right] \ \Rightarrow\ \hat{W} = \exp\left[-\frac{k^2 R_G^2}{2}\right] \tag{2.34}$$

for a gaussian filter on scale R_G and

$$W = \frac{3}{4\pi R_T^3}\ \ (r < R_T)\ \Rightarrow\ \hat{W} = \frac{3[\sin y - y\cos y]}{y^3} \tag{2.35}$$

for a top-hat filter on scale R_T, where $y \equiv kR_T$.

For the variance of the unsmoothed density field we simply set $\hat{W} = 1$.

We typically use comoving scales, although any scale can also be expressed equivalently as a physical mass scale. This is simply the mass of the matter density field enclosed in a sphere with radius equal to the respective length scale

$$M(r) = \frac{4}{3}\pi r^3 \Omega_m \rho_{crit} \tag{2.36}$$

where r is comoving/physical if ρ_{crit} is.

2.1.3 Cold Dark Matter models

A main constituent of the model we will be working with is the existence of dark matter. This has been proposed to explain the discrepancy between the amount of directly observed matter in the Universe and what appears to be required. A host of observations, ranging from rotation curves of galaxies to the CMB and abundances of galaxies clusters indicate that, judging by its gravitational effect, a lot more mass is necessary. The simplest explanation is that the dark matter is just ordinary 'baryonic' matter too dark to be observed. This is true to an extent. However, constraints from Big Bang nucleosynthesis (BBN) place upper limits on the total amount of baryonic matter in the Universe, which still falls far short of what is required. One straightforward conclusion would then be that there must be some other non-baryonic form of matter. This distinction is made explicit by the introduction of another dimensionless density parameter for the baryonic mass in the Universe, Ω_b, with $\Omega_m - \Omega_b$ accounting for the non-baryonic dark matter. Since most measurements of the baryonic matter content depend directly on the actual value of the Hubble constant, Ω_b is typically given in units of the square of the dimensionless Hubble parameter h^{-2} With this distinction a clear cut definition is then that dark matter constitutes matter which is not only practically but in principle unobservable other than by its gravitational effect. That is to say it does not interact with or couple to baryonic matter or photons such that we would be able to see it.

Primordial BHs as dark matter

The nature of dark matter is still completely uncertain, although over the years a number of candidates have been excluded. Black holes have been suggested as a dark matter candidate as they are also unobservable other than by their gravitational effect on their surroundings. The emergence of Hawking radiation at the event horizon is far too weak to be used as an observational constraint. Through this effect any primordial BHs must have a mass of at least 10^{-19} $M_{\odot}$ (Peacock 1999) or else they will have evaporated all their mass through Hawking radiation. At the same time dark matter cannot be exclusively made up of less than 10^2 $M_{\odot}$or more than 10^4 $M_{\odot}$stellar mass BHs. Their resulting dynamical effect on the structure of galaxies has not been observed (Carr, Bond, & Arnett 1984;Hut & Rees 1992;Rix & Lake 1993;Klessen & Burkert 1996). Constraints on dark matter BHs in the range $10^2 < M_{\bullet} < 10^4 M_{\odot}$may be possible through a combination of methods (Gould 1992) and the *GAIA* astrometric satellite emission planned for launch in 2010. It appears possible that BHs in the range $\sim 0.1 - 1$ $M_{\odot}$could make up at least a sizable part of DM although their accretion in the early Universe would arguably leave an ionisation signature that also does not agree with observations (Miller 2000). Many of these arguments also hold for similar DM candidates like quark and boson stars.

A more radical approach is based on the realisation that many observations implying the need for DM are actually referring to large distances. Instead of invoking DM it may equally well be possible that the effect deviates from the inverse square law at these large distances. An example of this is *Modified Newtonian Dynamics* (MOND) (Milgrom 1983). Other more unusual explanations for DM include the idea that DM arises from scalar fields similar to that constituting the cosmological constant Λ or more generally the 'dark energy' component of the Universe, and that in fact the two are related (Padmanabhan & Choudhury 2002).

Dark matter: hot, warm and cold

However, throughout this work we will work with another and arguably more conservative scenario, and assume that DM is made up throughout of some type of elementary particle which does not couple to ordinary baryonic matter. This category can be subdivided again according to the particle energy at the time of matter - radiation equality. DM of this kind is referred to as *hot dark matter* (HDM) if the HDM particles were relativistic at that time. Unfortunately their associated free streaming length implies that density perturbations on corresponding scales are wiped out. As a result presently observed small galaxies would have had to be created through 'top-down fragmentation' of very large galaxies, which is problematic (e.g. Hut & White 1984) [3]

---

[3] The epoch of matter-radiation equality is important in this context as after this time the growth of density perturbations is governed by how the matter behaves. To the extent that DM particles are in motion after the time

If the DM particles are moving at non-relativistic speeds at equality they are said to constitute *Cold Dark Matter* (CDM). In this case density perturbations on small scales are preserved and can grow in time. For typical cosmological matter power spectra it turns out that perturbations collapse to form bound objects beginning at these small scales and progressing to increasingly larger scales. This is referred to as hierarchical structure formation in contrast to the top down fragmentation scenario for HDM. The class of CDM models in which $\Omega_m = 1$ is typically referred to as *standard* cold dark matter (SCDM) models.

Predictably there is also the idea of *Warm Dark Matter* (WDM) which combines features of both HDM and CDM. In particular it essentially leaves the hierarchical formation scenario in place but erases perturbations on very small scales just as is needed to reconcile the number of satellite galaxies obtained from WDM simulations with observations. However, the detection of reionisation in the CMB with the *Wilkinson Microwave Anisotropy Probe* now appears to rule out the class of WDM models (Spergel et al. 2003)

So far in HDM, WDM and CDM the particle nature has remained unspecified, although HDM and WDM put certain limits on the mass of any potential candidate. Finally it has also been suggested that the nature of DM could be probed by observations of any decay radiation and particles arising from DM if it was weakly self-interacting. If this was the case the effect would be expected to be largest where DM densities are highest such as in the centre of the DM halos enveloping galaxies.

The fiducial model: ΛCDM

Throughout this work we will be working with the CDM model although we may occasionally comment on some of the others. In particular we will use a CDM model which appears to be in best agreement with a wide range of observations including data from the CMB, galaxy surveys of large scale structure (Tegmark, Zaldarriaga & Hamilton 2001; Percival et al. 2002) and crucially, observations of supernovae at high redshifts (Riess et al. 1998; Perlmutter et al. 1999). The latter has so far provided the strongest case for the presence of a 'dark energy' component as

---

of equality they will erase features in the matter distribution smaller than the free streaming length associated with the DM motion. However, only for dynamically hot dark matter, the free streaming length becomes significant and covers scales up to and including that corresponding to small galaxies.

Table 2.1: Parameters of ΛCDM model used throughout this work. σ_8 is the variance of the matter density field smoothed on scales of 8 h^{-1} Mpc.

Ω_m	$\Omega_b\, h^2$	Ω_Λ	h	σ_8	n_{prim}
0.3	0.02	0.7	0.7	0.9	1.0

parametrised by the cosmological constant and the corresponding density parameter Ω_Λ. The model that emerges from these observations and that we use is shown in table 2.1.3

The power spectrum of density perturbations corresponding to this model is created initially by inflation which produces a primordial spectrum that follows a power law characterised by an index n_{prim}, $\delta^2_{\mathbf{k}} \propto k^{n_{prim}}$. Radiation pressure and other damping effects subsequently suppress the growth of perturbations on small scales, particularly during the radiation dominated epoch. This introduces the characteristic bend in the power spectrum after matter radiation equality. The latter effect is encapsulated by the *transfer function*, T_{tf} and the resulting power spectrum, $P(k)$ normalised to the present day is then

$$P(k) = C_{norm}\delta_{\mathbf{k}}\; T^2_{tf}(k) \tag{2.37}$$

where C_{norm} is chosen such that $\sigma_8 = 0.9$ (using eq. 2.33). For all our computations we used the CDM transfer function given by Eisenstein & Hu 1999 .

$$T_{tf}(k) = \left[1 + \frac{q^2(k)\,(14.2 + 731/(1 + 62.5q(k)))}{\ln(2e + 1.8q(k))}\right]^{-1} \tag{2.38}$$

where

$$q(k) = \frac{k}{\Omega_m\; h_0\; \mathrm{Exp}(-\Omega_b - \Omega_b/\Omega_m)} \tag{2.39}$$

and k is given in units of h Mpc^{-1}.

2.2 Gravitational Growth of density Perturbations

2.2.1 Adiabatic vs iso-curvature perturbations

Fundamentally there are two ways in which the primordial matter-radiation field can be perturbed. If the field is perturbed adiabatically the matter and photon number densities change by the same factor. However, since their respective energy densities scale differently with a change in volume the overall energy density is also changed.

The opposite is the case for iso-curvature perturbations. Here the combined matter + radiation density remains homogeneous, which implies that the spatial curvature does, too. Density perturbations in any one component would therefore have to cancel each other $\rho_m \ \delta_m = -\rho_r \ \delta_r$. In the context of inflation it is more difficult to produce these as they cannot be generated by the simplest one-field inflation models (Mukhanov, Feldman & Brandenberger 1992).

The type of perturbation modes has implications for the subsequent evolution as the Universe expands. Throughout we will always assume that we are dealing with adiabatic perturbations.

2.2.2 Linear Regime

The subsequent linear growth of density perturbations occurs through gravity. The problem is complicated by the fact that this happens in a uniformly expanding cosmological background. This process can be followed by considering the evolution of density perturbations in an ideal pressure-less fluid of particles. In comoving coordinates this can be described by the continuity and Euler equations as follows (Peebles 1980)

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot [(1+\delta)\mathbf{u}] = 0 \tag{2.40}$$

$$\frac{\partial \mathbf{u}}{\partial t} + H\mathbf{u} + \frac{1}{a}(\mathbf{u} \cdot \nabla)\mathbf{u} = -\frac{1}{a}\nabla\phi \tag{2.41}$$

where $\mathbf{u}$ is the peculiar velocity, obtained by subtracting the Hubble expansion from the physical velocity of a given particle

$$\mathbf{u} = \ \mathrm{d}\mathbf{r}/\ \mathrm{d}t - H\mathbf{r} \tag{2.42}$$

If the density perturbations are small $\delta \ll 1$ we can combine equations 2.40 and 2.41 and expand to first order to obtain

$$\frac{\partial^2 \delta}{\partial t^2} + 2H\frac{\partial \delta}{\partial t} = 4\pi G\bar{\rho}\delta \tag{2.43}$$

Of the two solutions to this equation, one corresponds to a growing mode that subsequently dominates the evolution of density perturbations. This mode can be described by a linear growth factor $D_{grow}(t)$ which takes the general form

$$D_{grow}(t) \propto \frac{\left[\Omega_\Lambda a^3 + \Omega_m + (1 - \Omega_\Lambda - \Omega_m)a\right]^{1/2}}{a^{3/2}} \int^a \frac{a^{3/2}\, \mathrm{d}a}{\left[\Omega_\Lambda a^3 + \Omega_m + (1 - \Omega_\Lambda - \Omega_m)a\right]^{3/2}} \tag{2.44}$$

2.2.3 Non-Linear Growth and Collapse of Objects

As the density contrast of the perturbations becomes of order unity the linear treatment breaks down. In the non-linear regime the further growth and eventual collapse of density perturbations can only be followed analytically if they display particular symmetries. In the simplest case we consider a perturbation that corresponds to a constant matter overdensity within a sphere. This spherical overdensity will initially follow linear growth and increase in diameter until it reaches a point of maximum expansion after which it collapses to a point. It can be shown that this 'turnround' occurs when the enclosed overdensity has reached a critical value δ_c (Peebles 1980; Lacey & Cole 1993)

$$\delta_c(z) = \frac{1.686}{D_{grow}(z)} \tag{2.45}$$

for a SCDM cosmology, where we have normalised D_{grow} such that $D_{grow}(z=0) = 1$. This only varies slightly for other CDM models and can be well approximated numerically. For $\Omega_m < 1$ and $\Omega_\Lambda = 0$ we simply multiply eq 2.45 by $\Omega_m(z)^{0.0185}$ and for $\Omega_m + \Omega_\Lambda = 1$ we multiply by $\Omega_m(z)^{0.0055}$ (Navarro, Frenk & White 1997).

Although the 'top-hat' collapse is a good approximation, the required degree symmetry will in practice never be attained. The smallest deviations from spherical symmetry lead instead to the virialisation of the collapsing perturbation. Although we said earlier that we are dealing with non-interacting dark matter, the DM particle energies can be scrambled up during virialisation through violent relaxation (Lynden-Bell 1967), that is the effect of the rapidly changing gravitational potential into which they collapse.

In its final state the perturbation then forms a bound extended object - or *halo* inside which the average overdensity relative to the average background density is $\Delta_c = 18\pi^2 \approx 178$ (see e.g. Lacey & Cole 1993) for a SCDM model. For other CDM models with $\Omega_m + \Omega_\Lambda = 1$, Δ_c can be approximated by (Bryan & Norman 1998)

$$\Delta_c(z) = 18\pi^2 + 82(\Omega_m(z) - 1) - 39(\Omega_m(z) - 1)^2 \tag{2.46}$$

For these halos we can now determine a number of quantities, such as their virial radius, circular velocity at the virial radius, virial temperature and binding energy all in dependence of redshift z (Barkana & Loeb 2001).

$$\begin{aligned}
r_{vir} &= 0.784 \left(\frac{M}{10^8 h^{-1} M_\odot}\right)^{1/3} \left[\frac{\Omega_m}{\Omega_m(z)} \frac{\Delta_c}{18\pi^2}\right] \left(\frac{1+z}{10}\right)^{-1} h^{-1}\text{kpc} && (2.47)\\
V_{circ} &= \left(\frac{G\ M}{r_{vir}}\right) = 23.4 \left(\frac{M}{10^8\ h^{-1}\ M_\odot}\right)^{1/3} \left[\frac{\Omega_m}{\Omega_m(z)} \frac{\Delta_c}{18\pi^2}\right]^{1/6} \left(\frac{1+z}{10}\right)^{1/2} \text{km s}^{-1} && (2.48)\\
T_{vir} &= \frac{\mu_{mol} m_p V_c^2}{2k_B} = 1.98 \times 10^4 \left(\frac{\mu_{mol}}{0.6}\right) \left(\frac{M}{10^8 h^{-1} M_\odot}\right)^{2/3} \left[\frac{\Omega_m}{\Omega_m(z)} \frac{\Delta_c}{18\pi^2}\right]^{1/3} \left(\frac{1+z}{10}\right) \text{K} && (2.49)\\
E_{bind} &= \frac{1}{2} \frac{G\ M^2}{r_{vir}} = 5.45 \times 10^5 3 \left(\frac{M}{10^8 h^{-1} M_\odot}\right)^{5/3} \left[\frac{\Omega_m}{\Omega_m(z)} \frac{\Delta_c}{18\pi^2}\right]^{1/3} \left(\frac{1+z}{10}\right) h^{-1} \text{erg} && (2.50)
\end{aligned}$$

where μ_{mol} is the mean molecular weight, m_p the proton mass and k_B is Boltzmann's constant. The virial radius corresponds to the size of the halo when the average density of material inside is equal to the virial overdensity,Δ_c, mentioned above, while the virial temperature corresponds to the virial kinetic energy, which is just 1/2 of the potential energy of material inside r_{vir}. We already noted that in typical CDM models perturbations on small scales collapse before those on larger scales. In this picture of hierarchical structure formation we thus have small halos forming at earlier times than large ones. We expect real halos collapsing in this cosmological context to have a more complicated structure than a simple constant density sphere. N-body simulations of halo collapse in an expanding CDM Universe reveal that collapsing halos can be described by a sphere although with a radial density profile. Early simulations (Dubinski & Carlberg 1991) produced halos that were well fit by a profile of the form $\rho(r) \propto r^{-1}(r_s + r)^{-3}$, where r_s is some scale radius. Better resolved simulations (Navarro, Frenk & White 1997) indicated a profile that rises towards the centre of the halo as $\rho(r) \propto r^{-1}(r_s + r)^{-2}$. The most recent simulations with still higher resolution (Moore et al. 1999; Klypin et al. 2001) converge to a profile that varies as $\rho(r) \propto r^{-1.5}(r_s + r)^{-1.5}$.

Generally, throughout this work we will be less concerned with the particular structure of halos and will continue to work with equations 2.47 - 2.50 for the case of a top-hat, constant density sphere. They still remain reasonable approximations even for the halos from the simulations, since the mass and average overdensity inside the halo virial radius is the same.

2.2.4 Abundance of Non-Linear Objects

In addition to the properties of individual halos it is also desirable to determine the cosmic abundance of halos at a given time. An estimate of this is fairly straightforward with our assumption that the initial matter density field from which halos collapse is gaussian. First we can determine the fraction of mass, f_m, in the Universe that is contained in halos of mass M at a given redshift z. To do this we consider the initial matter density field and identify particles contained in the peaks of this field above a certain height as those that get subsequently incorporated in halos collapsing from these peaks. The height of the peak on a particular length (or mass) scale, is simply characterised by the amplitude of the overdensity inside, δ. A halo of mass M is therefore considered as having collapsed from a peak of comoving size $\sim (\frac{4}{3}\pi\rho)^{-1/3}$ and height greater than δ_c. That means the fraction of mass in halos of a given mass is equal to the mass fraction contained in peaks with $\delta > \delta_c$ at a given length (mass) scale

$$f_m(> M, z) = 1 - \text{ erf}\left[\frac{\delta_c(z)}{\sqrt{2}\sigma(M)}\right] \tag{2.51}$$

where the scale dependence enters through the variance $\sigma(M)$.

The same ansatz was pushed further by Press & Schechter (1974) to determine the abundance and spectrum of halo masses in the Universe. For the differential mass function of halos they obtain

$$\frac{\mathrm{d}n_h}{\mathrm{d}\ \ln M}\,\mathrm{d}\ \ln M = \left(\frac{2}{\pi}\right)^{1/2} \frac{\bar{\rho}}{M}\frac{\delta_c}{\sigma(M)} \left|\frac{\mathrm{d}\ \ln\sigma(M)}{\mathrm{d}\ \ln M}\right| \ \mathrm{Exp}\left[-\frac{\delta_c^2}{2\ \sigma(M)^2}\right]\ \mathrm{d}\ \ln M \tag{2.52}$$

In further course we will make repeated use of this last equation.

For completeness we also mention a frequently used extension of the Press-Schechter treatment. Fundamentally this *Extended Press-Schechter* (EPS) formalism allows to determine analytically the probability that a halo with a mass M_1 at time t gains an amount of mass ΔM some time Δt later to produce a more massive halo with mass M_2. As Δt becomes infinitesimally small any mass increase ΔM must arise from the merging of a single mass clump with the halo. In this limit one then obtains the merger rate (probability per unit time) per unit mass of the clump that is being merged with

$$\begin{aligned}\frac{\mathrm{d}^2 p}{\mathrm{d}\ \ln\Delta M\ \ \mathrm{d}t}\ \mathrm{d}\ \ln\Delta M\ \ \mathrm{d}t &= \left(\frac{2}{\pi}\right)^{1/2}\frac{1}{t}\left|\frac{\mathrm{d}\ \ln\delta_c}{\mathrm{d}\ \ln t}\right|\left(\frac{\Delta M}{M_2}\right)\times\left|\frac{\mathrm{d}\ \ln\sigma_2}{\mathrm{d}\ \ln M_2}\right|\frac{\delta_c(t)}{\sigma_2}\\ &\times\frac{1}{(1-\sigma_2^2/\sigma_1^2)^{3/2}}\ \mathrm{Exp}\left[-\frac{\delta_c(t)^2}{2}\left(\frac{1}{\sigma_2^2}-\frac{1}{\sigma_1^2}\right)\right]\ \mathrm{d}\ \ln M\ \ \mathrm{d}t\end{aligned} \quad (2.53)$$

where subscripts '1' and '2' refer to the halo before and after merging. For individual halos equation 2.53 can also be used to determine halo formation and survival times.

Since the underlying probability distribution of fluctuations in the matter density field is random gaussian, the probability distribution of the merging clump masses ΔM is, too. If we consider a sequence of mergers, we then obtain a random 'trajectory' of mass increases for any given halo. Using this one can then generate random ('Monte-Carlo') halo merger histories. This forms the basis of our semi-analytical approach to halo merging and dynamical evolution in chapter 5.

2.3 Evolution of the Baryonic Component

It is exclusively gravity that determines the linear growth of perturbations and the collapse of halos. Non-gravitational forces only affect the baryonic component and even then only become important inside collapsed halos where the baryon density has grown sufficiently high.

2.3.1 Jeans Mass and baryonic collapse

As the baryon density becomes large enough, pressure forces can impede or halt gravitational collapse of the baryonic component. A crucial quantity in this context is the *Jeans length*, λ_J. It defines the diameter of a sphere of matter inside which the sound travel time equals the gravitational free-fall time

$$\lambda_J = c_s\sqrt{\pi}G\ \hat{\rho} \quad (2.54)$$

where c_s is the speed of sound and $\hat{\rho}$ represents the constant density of matter inside the sphere. The amount of mass enclosed in the sphere is called the *Jeans mass*, M_J

$$M_J = \frac{4\pi}{3}\left(\frac{\lambda_J}{2}\right)^3 \hat{\rho} \tag{2.55}$$

For a gas at temperature T and with baryon number density n_b this is (Peacock 1999)

$$M_J = 2 \times 10^5 \left(\frac{T}{\mathrm{K}}\right)^{3/2} \left(\frac{n_b}{\mathrm{m}^{-3}}\right)^{-1/2} M_\odot \tag{2.56}$$

An object with the Jeans mass will become gravitationally bound; whether it can actually collapse is another question. As the gas of baryons collapses it heats up. If this happens adiabatically the temperature increases with the gas density and so does the Jeans mass. As a result the collapse is halted unless the gas can radiate away energy and cool down sufficiently.

In reality this issue is further complicated in two different ways. First, if we consider the linear regime, the Jeans mass is modified by the universal expansion, in which case we get

$$M_J = 6 \times 10^3 \left(\frac{\Omega_m\, h^2}{0.15}\right)^{-1/2} \left(\frac{\Omega_b\, h^2}{0.02}\right)^{-3/5} \left(\frac{1+z}{10}\right)^{3/2} M_\odot \tag{2.57}$$

which in this case is the total dark + baryonic mass. Here Ω_b is the cosmic density parameter for baryons. More importantly, however, a baryonic cloud does not collapse by itself, but will typically do so inside the potential well of a dark matter halo that we assume has already virialised. Assuming baryons settle into the DM well when their mean overdensity exceeds $\Delta_{c,b} = 100$ (c.f. $\Delta_c = 178$ for the respective dark matter overdensity) and neglecting cooling for the moment, the minimum halo mass required is

$$M_J = 5 \times 10^3 \left(\frac{\Omega_m\, h^2}{0.15}\right)^{-1/2} \left(\frac{\Omega_b\, h^2}{0.02}\right)^{-3/5} \left(\frac{1+z}{10}\right)^{3/2} M_\odot \tag{2.58}$$

where the connection between the baryonic and dark matter components arises from the fact that the baryons share the DM halos virial temperature. This connection enters through $\Delta_{c,b}$ which is a function of the halo virial temperature (Barkana & Loeb 2001).

2.3.2 Baryonic Cooling and the Formation of Galaxies

In the common picture of hierarchical structure formation the baryons in a gravitationally bound halo object do no just linger across the size of the halo, but settle to the centre to form galaxies. For this to happen, baryons have to be able to dissipate the heat energy they acquire when falling to the centre. Cooling is facilitated through Thomson scattering, or radiatively through ion-electron collisions (bremsstrahlung) if the baryons are fully ionised. Otherwise cooling proceeds radiatively mainly through electron recombination and de-excitation of partially ionised atoms. All cooling processes can be summed up in an overall cooling rate Λ_{cool} which is a function of temperature but also depends on the density of the cooling gas and chemical abundances. We can then define a characteristic time scale on which a gas at a given temperature, T, cools

$$t_{cool} = \frac{\frac{3}{2} n_b \, k_B \, T}{n_b^2 \, \Lambda_{cool}(T)} = 6.6 \times 10^9 \left(\frac{T}{10^6 \mathrm{K}}\right) \left(\frac{n_b}{10^{-3} \mathrm{cm}^{-3}}\right)^{-1} \left(\frac{\Lambda_{cool}}{10^{-24} \mathrm{erg\ cm}^3\ \mathrm{s}^{-1}}\right)^{-1} \tag{2.59}$$

where n_b is the baryon number density. If this time scale is much longer than the Hubble time, $t_{hubble} = 1/H$ - which is a measure of the halo's collapse time - the baryonic component will not settle to the centre of the halo and galaxies fail to form. If $t_{cool} < t_{hubble}$, cooling can proceed efficiently and baryons can settle to create galaxies.

The present day Universe has been metal enriched through supernova feedback of processed elements from stars. The presence of metals significantly increases the number of reaction channels through which cooling can occur. As a result cooling generally proceeds on a much shorter time scale. However, before the first stars form the chemical abundance of the baryons is essentially primordial, i.e. consisting primarily of hydrogen, helium, small amounts of lithium and insignificant traces of higher elements. In this case it turns out that there is a minimum temperature above which the only significant cooling channel (molecular hydrogen) becomes effective. A minimum virial temperature in turn implies a minimum halo mass inside which baryonic clouds can collapse and form the first stars. This process of first star formation and the subsequent creation of BH remnants constitutes one of the pillars of our model and we will return to this aspect in chapter 4.

2.4 Non-standard cosmologies - an example

We have already made reference to so-called 'non standard' cosmologies. In particular we mentioned the MOND model. Although we will work with a common ΛCDM model throughout the following chapters, we will here briefly consider another non-standard model.

2.4.1 A primordial feature in the matter power spectrum

We consider the possibility that the primordial spectrum of density perturbations from inflation does not follow a straight power law as assumed above but instead has a feature on some scale.

Observationally this possibility is certainly not excluded. A number of observations, such as cluster redshift surveys (Einasto et al. 1997) and galaxy surveys (Broadhurst et al. 1990;Hamilton & Tegmark 2000;Gaztañaga & Baugh 1998) reveal possible traces of features in the matter power spectrum. In addition, a lower than expected second peak in the CMB spectrum as measured by BOOMERANG (Lange et al. 2001) appears to be incompatible with the simplest models of cold dark matter and a scale invariant primordial spectrum.

While the results, particularly from the surveys are still not beyond statistical doubt, a genuine feature(s) in the primordial spectrum cannot be ruled out. Several mechanisms have been proposed that could generate features in the primordial spectrum during the epoch of inflation. They commonly involve an extension of the simplest one field inflation model, e.g. by coupling the inflaton to a massive particle (Chung et al. 1999) or considering two-field inflation (Lesgourgues, Polarski & Starobinsky 1998). There are also more exotic ideas (see e.g. Martin, Riazuelo & Sakellariadou 2002). Another idea proposed is to Taylor expand the primordial power spectrum to include higher order terms that account for a running spectral index (Lidsey et al. ;Hannestad, Hansen & Villante). The latter is able to introduce a very broad negative or positive bend into the power spectrum.

While possibly providing a motivation for observed features on a specific scale (i.e. scales of about 100 Mpc), the inflationary mechanisms proposed appear, in principle, capable of producing features on other scales as well. Here we examine the effect of primordial features on smaller scales $k \sim 0.4 - 0.8$ $h\mathrm{Mpc}^{-1}$, where, due to (the onset of) non-linear evolution and the problem of biasing, a connection with the primordial spectrum is much harder to establish. The effects of

Table 2.2: Specifications of the log-normal features introduced into an otherwise unmodified, COBE normalised standard CDM (SCDM) model (Model 1a). The corresponding mass scale $M = \rho_{\text{crit}}\Omega \cdot \frac{4\pi}{3}(2\pi/k_0)^3$ is given in the last column.

label	$2\pi/k_0$	A	σ_{mod}	sign	mass scale
Model 1a	SCDM, Ω_m=1, h=0.5, σ_8=1.18				
Model 1b	SCDM, Ω_m=1, h=0.5, σ_8=0.52				
Model 2	8 Mpc/h	3.00	0.10	+	$6 \cdot 10^{14} M_\odot/h$
Model 3	8 Mpc/h	0.80	0.12	-	$6 \cdot 10^{14} M_\odot/h$
Model 4	8 Mpc/h	0.95	0.48	-	$6 \cdot 10^{14} M_\odot/h$
Model 5	16 Mpc/h	0.96	0.24	-	$5 \cdot 10^{15} M_\odot/h$
Model 6	16 Mpc/h	0.96	0.48	-	$5 \cdot 10^{15} M_\odot/h$
Open CDM	Ω_m=0.5, h = 0.7, σ_8=0.96				

such bumpy power spectra on the cosmic microwave background has been investigated previously (Griffiths, Silk & Zaroubi 2001).

2.4.2 Modified power spectra and their evolution

For this investigation we used COBE normalised standard CDM (SCDM) power spectra and modified them by introducing a Gaussian feature in log space given by

$$P_{\text{mod}}(k) = P(k) \cdot \left(1 \pm A \exp\left[-0.4\left(\frac{\log k - \log k_0}{\sigma_{\text{mod}}}\right)^2\right]\right) \tag{2.60}$$

where P(k) is the unmodified spectrum.

These features are parametrised according to their width σ_{mod}, height A, and the location k_0 of the bump (+) and dip (-), respectively. Their parameters are also summarised in Table 2.4.2 and the modified spectra shown in figure 2.1 with all linearly extrapolated to redshift $z = 0.0$.

We have shown in N-body simulations that the modified power spectrum evolve in such a way that by the present day they cannot be distinguished anymore (Knebe, Islam & Silk 2001, hereafter KIS01). The modified spectra converge to that of the unperturbed one at late times. In any case, the *evolution* of the models *has* to be different, and one needs to think of other ways to investigate the influence of such features in the present day universe or to detect them in the evolution of P(k) and related quantities. As far as the spectrum itself is concerned, current galaxy clustering surveys offer little prospect of detecting these features in the evolution of P(k). Observations of the Lyα forest, however, appear to be a promising tool for constraining P(k) at

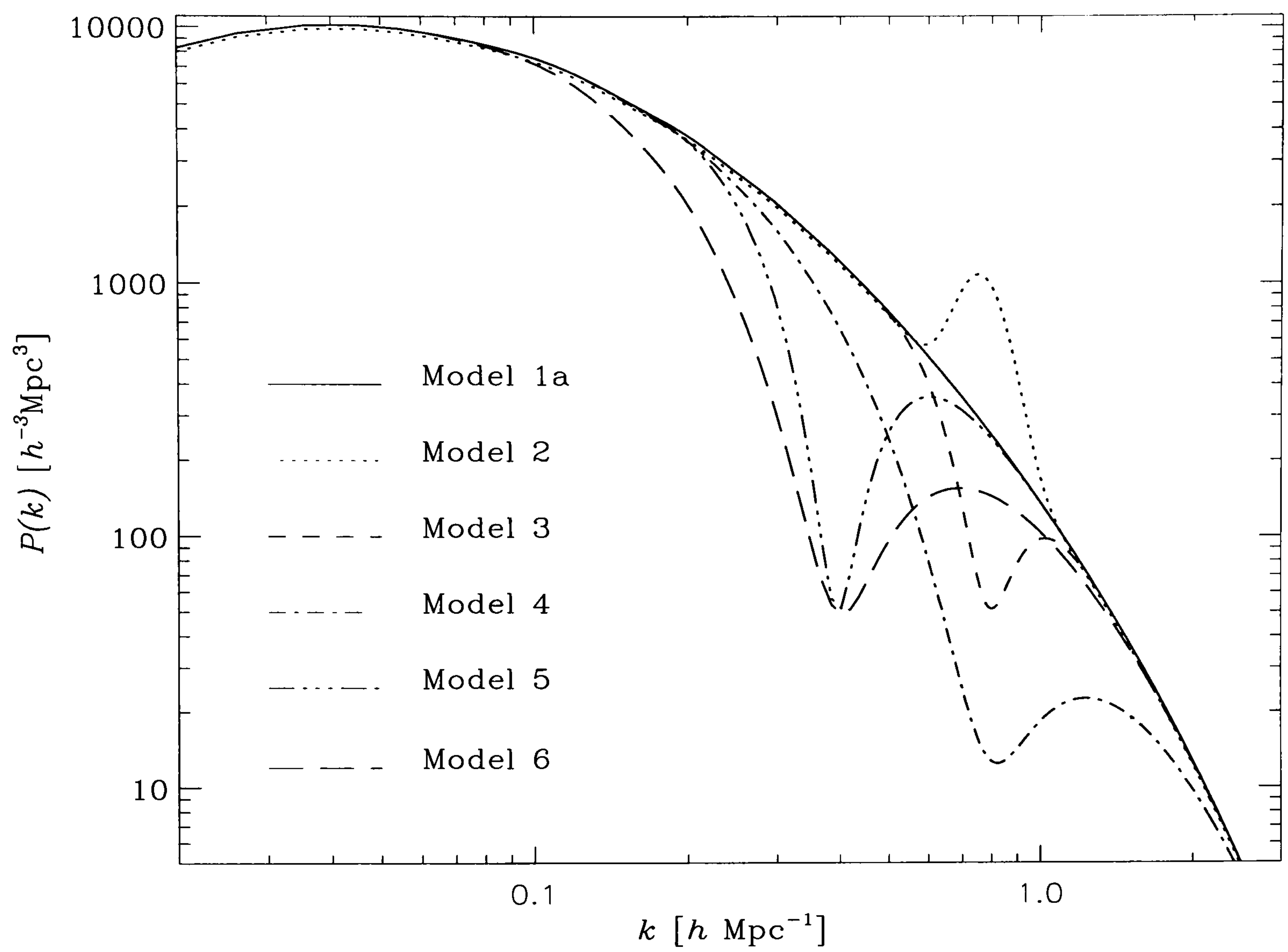


Figure 2.1: CDM input spectra at redshift $z = 0$.

high redshift (see e.g. Croft et al. 2002), particularly on the scales under investigation here.

The different evolution is also documented in the evolution of cluster abundance. Clusters of galaxies correspond approximately to a comoving size scale of 8 h^{-1} Mpc, which is considered a scale that is just about to go non-linear at present. The variance of the matter density field on these scales is given by $\sigma(8\ h^{-1}\text{Mpc}) - \sigma_8$ which is related to the matter density parameter as $D_{grow}(z)\ \sigma_8\Omega_m^{0.5}(z) \simeq 0.5$ This is crucial, because for unmodified CDM models the cluster abundance evolution can then be used to determine the density parameter of the Universe (see e.g. Bahcall, Fan & Cen 1997 and references therein). We have seen above that SCDM models with features at small scales may display an evolution of their cluster abundance that is similar to that in the OCDM model to within some constant offset in amplitude.

2.4.3 Primordial features and the evolution of clusters

The question remains as to how far we can push this resemblance between SCDM + feature and OCDM models and to what extent this also holds for cluster masses larger than the ones we were able to investigate numerically. To examine this we performed Press-Schechter (Press & Schechter 1974) calculations to investigate a larger set of models. However, since we are considering the evolution of cluster abundances, this also requires a relation for the redshift dependence of the critical overdensity δ_c, which is left unspecified in the original PS approach. We determined approximate relations for the unmodified SCDM and OCDM models by matching their cluster abundance evolution to that of the corresponding N-body simulations:

$$\delta_c^{SCDM}(a) = 1.2 + 0.6a, \delta_c^{OCDM}(a) = 1.45 + 0.3a \tag{2.61}$$

where $a = 1/(1+z)$ is the cosmic scale factor at redshift z.

For the OCDM model the δ_c relation almost exactly reproduces the abundance evolution. The corresponding relation for the SCDM model was rather chosen to fit both unmodified SCDM and featured SCDM models reasonably well, matching the N-body evolution to within 10 percent for the unmodified SCDM model, typically to within 20 percent for models 2 to 5 and 30 percent for model 6. This is shown in figures 2.2 and 2.3. An accurate fit to the unmodified SCDM model only would have required $\delta_c(a) = 1.25 + 0.5a$.

However, the agreement between N-body and PS data should be viewed with care, particularly when large sharp features are included; at the location of the features, we essentially introduce a more rapidly changing spectral index. This might contribute to a larger discrepancy between N-body and PS data. To our knowledge agreement between PS and numerical simulations has so far only been established for power spectra with constant (power law) or slowly varying (e.g. SCDM) spectral indices - it is not obvious why this should also hold when a sharply varying index is being introduced, particularly on scales that only now turn non-linear. In the following, however, we will assume that this is not a problem and use the δ_c^{SCDM} relation from equation (2.61) as a best estimate for PS predictions of the abundance evolution for other small scale features in the SCDM model. We also assume that the δ_c relation for OCDM describes the abundance evolution

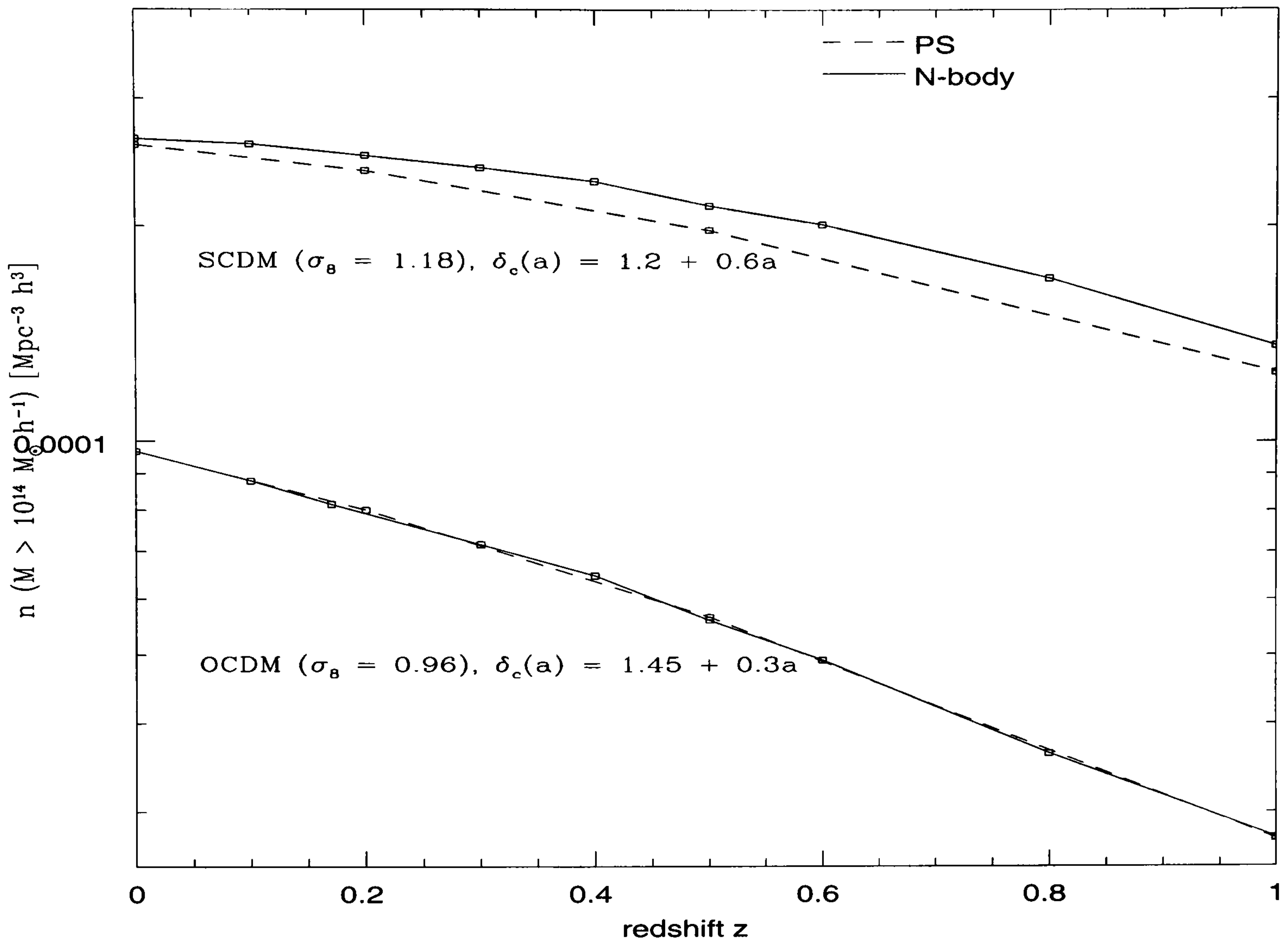


Figure 2.2: Comparison of evolution of clusters of mass $M > 10^{14} M_{\odot} h^{-1}$ in simulations and PS using the relations for δ_c from the simulations.

in higher mass cuts as well as it does for $M > 10^{14}$ h^{-1} M$_{\odot}$in our simulations.

As seen above, one might expect that adding (subtracting) power only at specific scales leads to a larger (smaller) abundance of objects with mass corresponding to these scales. This is not obvious for scales that are already in the non-linear regime, as we have a coupling across scales. Objects not only get newly created but now have also been partly incorporated into larger objects. As a result the effect of a feature on a specific scale spreads out, not only in the evolution of the power spectrum as we have seen above, but also in the abundance of objects of corresponding mass. However, for broad features we may assume that the latter is less important, since a whole range of scales and corresponding masses are affected. To check this, we placed a broad Gaussian dip in the SCDM model ($A = -0.55$, $\sigma_{\mathrm{mod}} = 2.5$) at $10h^{-1}$ Mpc, which is expected to affect a similarly broad range of masses centred on $M \sim 10^{14} - 10^{15} M_{\odot} h^{-1}$. A broad feature of this

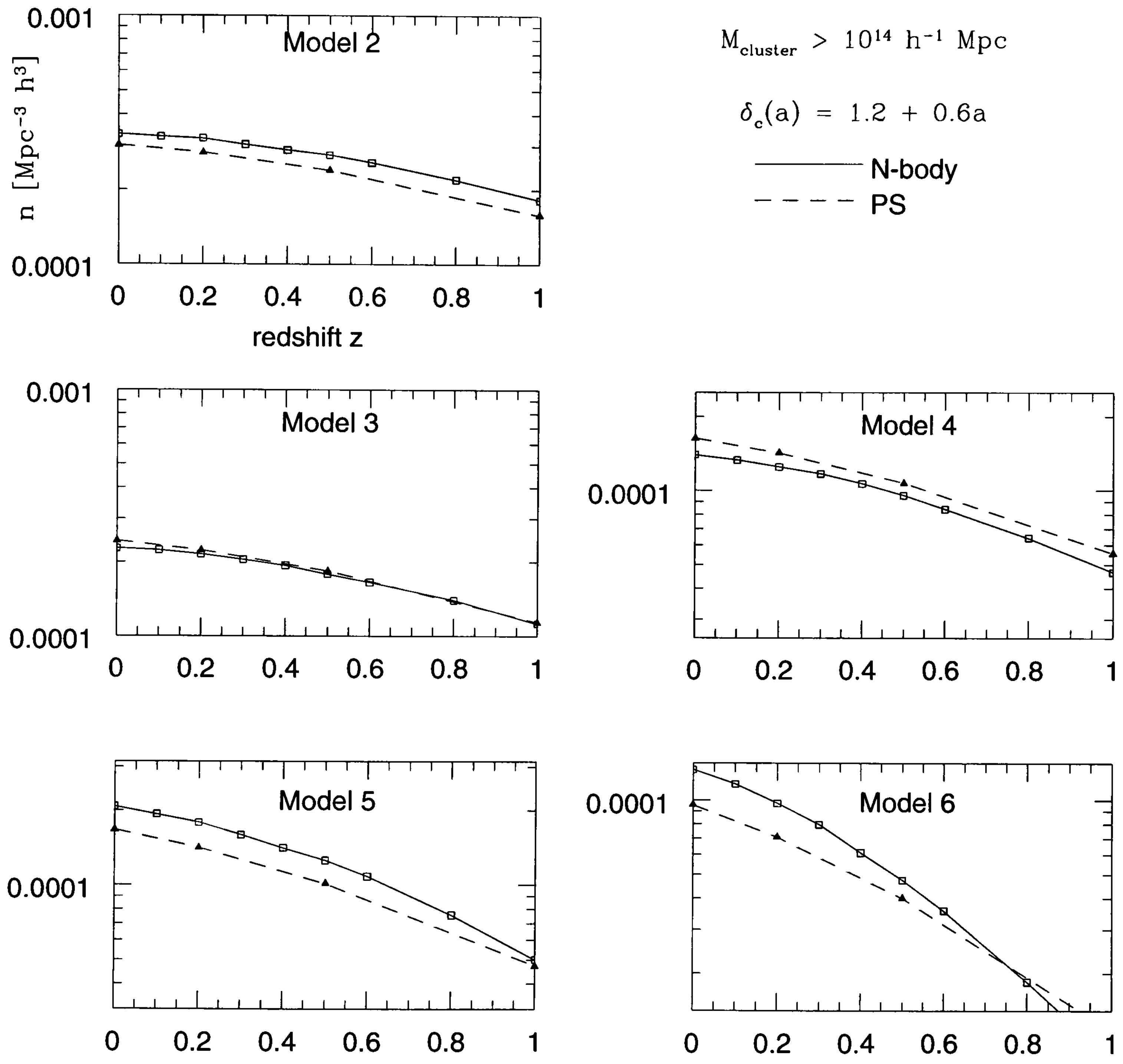


Figure 2.3: Comparison of abundance evolution for clusters with mass $M > 10^{14} M_{\odot} h^{-1}$ in simulations and PS. All PS calculations use $\delta_c(a) = 1.2 + 0.6a$ (eq (2.61)).

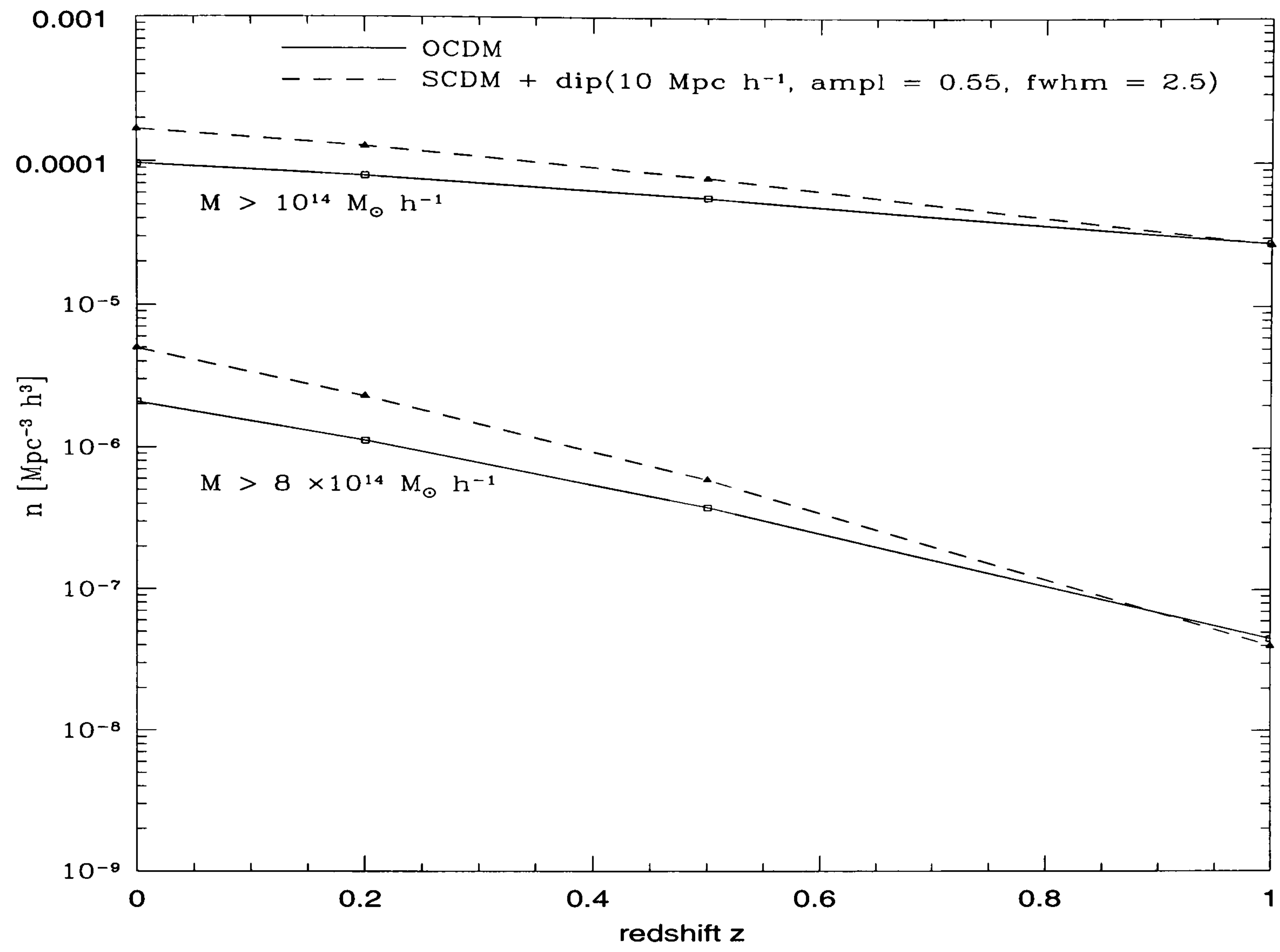


Figure 2.4: PS calculations of abundance evolution for OCDM and SCDM with a broad dip at 10 Mpc h^{-1}. For the mass cuts the two agree to within less than a factor of 2 and 3 respectively across the whole redshift range.

kind is naturally generated by e.g. including a running spectral index term when expanding the primordial perturbation spectrum (Hannestad, Hansen & Villante 2000).

Comparing Fig. 2.2 with Fig. 2.4 confirms that the abundance of masses in these ranges ($M > 10^{14}$, $M > 8 \cdot 10^{14} h^{-1}$ $\mathrm{M}_\odot$, the mass range covered by e.g. Bahcall & Fan 1998) is significantly suppressed. Moreover, the abundance evolution for the mass cuts agrees in logarithmic slope *and* magnitude with that in the OCDM model to within a factor of less than two and three across all redshifts up to z = 1 for the lower and upper mass cuts respectively.

Given that observations of the most distant massive clusters currently only constrain the abundance to within one or two orders of magnitude (see e.g. Bahcall & Fan 1998), a SCDM model with broad features towards small scales, which mimicks (in shape and absolute magnitude) the abundance evolution expected from OCDM models, can therefore not be ruled out.

Chapter 3

Massive black holes and galaxies

For a long time it was thought that the above picture of structure formation describes an essentially complete framework within which the formation of galaxies and stars takes place. Although some major pieces are still missing - notably a fundamental theory of star formation - this framework has gone a long way to explaining the present day distribution and structure of galaxies.

There had been some work on the possible presence of black holes in galaxies and galactic centres, motivated by the observation of active galactic nuclei (AGN) and QSOs since these can only really be explained by very massive BHs operating as their central engine. Since the mid 1990s a number of observations of galactic centres have been made that strongly suggest – especially for the case of our own Milky-Way galaxy – the presence of 'super massive black holes' with masses in the range between $10^6 - 10^9$ $M_\odot$. In addition it was found that the SMBH mass is correlated with certain properties of the host galaxy bulge far beyond the SMBH's range of direct dynamical influence, implying that the past evolution of the galaxy and SMBH must have been somehow fundamentally related. In the wake of these observations it was realised that this important connection had so far been neglected and a lot more work has since been carried out on the origin of these SMBHs.

In section 3.1 we summarise the observations of both the abundance of SMBHs as well as various correlations between the mass of SMBHs and properties of their galactic hosts' bulges. In section 3.2 we review the theoretical efforts to explain the observations and try to position our model among them. Finally in section 3.3 we focus on the significance and origin of intermediate mass black holes. These will subsequently play a crucial role in our model as the link between stellar mass BHs and SMBHs.

3.1 SMBHs in galactic centres - observations

A number of studies since the mid 1990s have used data obtained from HST as well as ground based observations to investigate the kinematics of stars and gas near the centres of a number of early-type galaxies (c.f. Kormendy & Richstone 1995 for a review, Magorrian et al. 1998, hereafter Mag98). Of these the most extensively surveyed one is the centre of the Milky-Way. From the kinematics of stars and gas in these central regions the presence of massive dark objects (MDOs) was inferred. In these studies, however, the resolution was too poor to establish whether MDOs are constituted by SMBHs, clusters of brown dwarfs or other. It was argued that for a few galaxies, including our own, the inferred density of the MDO is so high, that any cluster of stellar mass BHs or brown dwarfs would not have survived (due to evaporation and collision processes) to the present day to observe them (Maoz 1998). The case for a SMBH is therefore strong for these galaxies and especially for the central MDO in the Milky-Way whose mass is now known to be between 2.6×10^6 and 3.3×10^6 $M_\odot$ (Genzel et al. 2000). HST observations today (see Kormendy & Gebhardt 2001, hereafter KG01, for a review of recent HST observations) have a much better resolution ($\sim 1''$) than ground based observations and have essentially ruled out any plausible astrophysical explanation for the detected MDOs other than SMBHs for many more galaxies although for some the resolution is still not high enough to rule out clusters. Clusters made up of elementary particles or low mass BHs (≤ 0.04 $M_\odot$), would survive for long enough to make them an alternative to SMBHs, even for galaxies with the highest inferred matter densities in MDOs (Maoz 1998). However, there are other serious problems connected with these alternatives, especially concerning possible formation scenarios.

If a SMBH is present there should be a region of influence within which it dominates the gravitational potential. With the improved HST observations, it has finally become possible to resolve this region. In the case of the Milky-Way central MDO the proper motion of stars on Keplerian orbits has now been observed, which is the strongest indication yet, that the Milky-Way MDO really is a SMBH (Schödel et al. 2002). SMBHs have now been detected in this way for a number of galaxies and they appear to be a common phenomenon, which also has implications for

AGNs. On the basis of the inferred small size of the emission region combined with the energetics, AGN had long been suspected to be powered by accreting SMBHs. Although necessary in this case, SMBHs were long regarded as a rather exotic possibility, given that no other observations would indicate their presence. With the HST observations, however, SMBHs now turned from the only viable to the natural candidate for the engines of AGNs.

In addition the masses of SMBHs in AGNs have now also successfully been determined using a technique called 'Reverberation mapping' of the broad emission-line region [1].

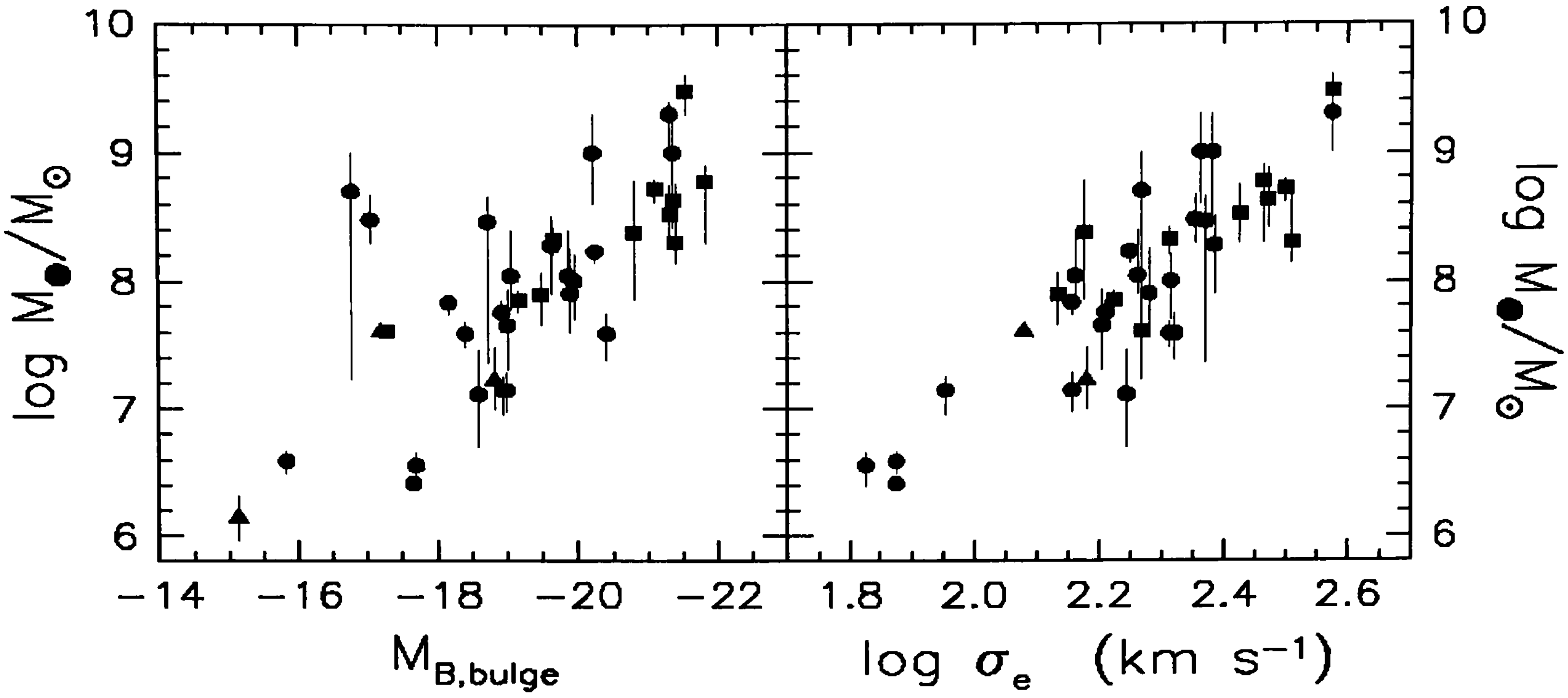
As mentioned, HST and ground based spectroscopy of stellar kinematics as well as reverberation mapping provide strong evidence for the presence of SMBHs due to the lack of plausible alternatives. However, only observations of the relativistic effects in the direct vicinity of the SMBH would constitute a 'smoking gun'. Recent CHANDRA and XMM detections of the iron Kα line from gas in the centres of AGNs may begin to provide just that (Turner et al. (2002) and references therein). In particular the observations allow to resolve the shape of the Kα line which is thought to depend sensitively on general relativistic effects, i.e. the strong gravity in the direct vicinity of the SMBHs.

3.1.1 The SMBH – bulge connection

Not long after the first detections of SMBHs at galactic centres, a number of correlations were discovered between the inferred SMBH mass and properties of the galactic bulges that host them. The first correlation to be established was one between the SMBH mass, M_{smbh}, and bulge mass, M_{bulge}. Mag98 infer an approximately linear relationship $M_{smbh} \sim 0.006 M_{bulge}$, while Ho (1999) and Kormendy (2000) find $M_{smbh} \sim 0.002 M_{bulge}$, Although this may have to be revised to a non-linear relation on the basis of more up to date HST data (KG01, Laor 2001). This correlation is determined from the observable bulge luminosity, L_{bulge} using the fundamental plane relation

---

[1] This is the measurement of the time lag between the light curves of the continuum - as emitted by the compact source and a broad line in the extended Broad Line Region (BLR) around the source. Interpreting the time lag as the light travel time, one can deduce the radius of the BLR and, assuming the motion of matter in the BLR is virialised, its velocity spread from the line width. Both can then be used to infer the mass of the central compact source, i.e. the SMBH.



t

Figure 3.1: SMBH mass versus the absolute blue magnitude of the bulge component of the host galaxy (left) and the luminosity-weighted mean velocity dispersion inside the effective radius of the bulge. In both panels, the different symbols indicate measurements of M_{smbh} based on stellar dynamics (filled circles), gas dynamics (squares) and maser disk dynamics (triangles). Figure taken from Kormendy & Gebhardt 2001.

$M_{bulge}/L_{bulge} \sim L_{bulge}^{0.2}$. According to the most recent data (KG01)

$$M_{smbh} = 0.78 \times 10^8 \mathrm{M}_\odot \left(\frac{L_{B;bulge}}{10^{10} L_{B_\odot}}\right)^{1:08} \tag{3.1}$$

However, the scatter in either the $M_{smbh} - M_{bulge}$ or $M_{smbh} - L_{bulge}$ relations is significant ($\sim \pm 0.5$ rms in the $\log(M_{smbh}/M_{bulge})$ for a given M_{smbh}, Laor 2001).

More recently, a significantly tighter correlation has been determined between M_{smbh} and the bulge velocity dispersion, σ_e at some effective distance r_e from the bulge centre. The most recent estimates lie at (Ferrarese 2002)

$$M_{smbh} = (1.66 \pm 0.32) \times 10^8 (\sigma_c/200)^{4.58 \pm 0.52} \mathrm{M}_\odot \tag{3.2}$$

and (Tremaine et al. 2002)

$$M_{smbh} = (1.35 \pm 0.43) \times 10^8 (\sigma_c/200)^{4.02 \pm 0.32} \mathrm{M}_\odot \tag{3.3}$$

The reason for the discrepancy between the two is not clear; it has been suggested that it originates in the use of different statistics, different definitions of the central velocity dispersion (Tremaine et al. 2002) or the use of discrepant values for the velocity dispersion of the Milky-Way (Merritt & Ferrarese 2001b, hereafter MF01b).

The latest type of correlation found is one between M_{smbh} and the light profile of the bulge as parametrised by a shape index n (Graham et al. 2001). Another new analysis (Ferrarese 2002) of published bulge velocity dispersions and disk circular velocities, v_c, beyond the optical radius has determined a tight correlation between the two. If v_c is a faithful tracer of the underlying total mass distribution, this implies a correlation between the mass of the dark matter component and σ_c and thus M_{smbh} (c.f. equation 3.2).

Since all these correlations extend well beyond the direct dynamical influence [2] of the SMBH it seems likely that there is a close link between the formation of both SMBH and host galaxy.

3.1.2 SMBH demographics

With the increasing number of detections of SMBHs in local quiescent galaxies and AGNs as well as their inferred presence in high red shift QSOs it seems that SMBHs are generic constituents of many if not most galaxies. This allows the question of SMBH demographics to be addressed.

One of the earliest estimates of the space density of SMBHs was given by Soltan (1982) based on the luminosity function of high redshift QSOs. If the luminosity function is entirely the result of accretion onto nuclear SMBHs and assuming a radiative accretion efficiency of 10 per cent, the space density of SMBHs is found to be $\rho_\bullet \sim 2 \times 10^5$ M$_\odot$ Mpc^{-3}. A recent estimate by Yu & Tremaine (2002) is based on velocity-dispersion data from the Sloan Digital Sky Survey (SDSS) and the QSO luminosity function from the 2dF survey. From the two data sets they determine $\rho_\bullet = (2.5 \pm 0.4) \times 10^5\ h_{0.65}^2$ M$_\odot$ Mpc^{-3} and $\rho_\bullet = 2.1 \times 10^5 (C_B/11.8)[0.1(1-\eta)/\eta]$ M$_\odot$ Mpc^{-3} respectively, where $h_{0.65}$ is the dimensionless Hubble constant in units of 0.65, C_B is the B band bolometric correction and η the radiative efficiency of accretion, i.e. the fraction of the rest mass energy of the accreted material that is released as radiation.

Combining the $M_{smbh} - M_{bulge}$ relation mentioned above with an estimate of the space density of galaxies (Fukugita, Hogan & Peebles 1998) – which can be computed using the Press Schechter (1974) mass function – Merritt & Ferrarese (2001a), hereafter MF01a, find $\rho_\bullet \sim 4.9 \times 10^5$

---

[2] That is to say the region within which the bulge could respond to the presence of the SMBH within a dynamical timescale or so.

$M_\odot$ Mpc^{-3}, while Mag98 based on their observations determine $\rho_\bullet \sim 2 \times 10^6$ $M_\odot$ Mpc^{-3}.

3.2 SMBHs in galactic centres - theoretical approaches

The assumption of a close link between galaxy and SMBH formation is the common starting point for most approaches that have so far been put forward to account for the correlations. KG01 identify three generic models:

I SMBHs growing to their present mass before the host galaxy assembled and so affecting the subsequent formation of the galaxy,

II seed BHs growing to SMBH size along with the host,

III SMBHs forming after the host has been assembled.

Regarding the actual mechanism by which the SMBHs grow to their present size, the models may be further subdivided according to

a SMBHs growing primarily by accretion of gas and stars,

b SMBHs growing though the merger of less massive precursor BHs.

An example of category **I** is the model proposed by Silk & Rees (1998), which assumes that SMBHs with mass $\geq 10^6$ $M_\odot$form as a result of the coherent collapse of the first baryonic clouds (which if anything places this approach more into the **a** subcategory) and thus precede the epoch of galaxy formation. The feedback from accreting SMBHs of this kind at the centres of protogalaxies may then subsequently alter star formation and gas-dynamics in a way so as to limit the gas available for further accretion and establish a relationship $M_\bullet \propto \sigma^5$, which would be narrowly consistent with the relation established by Ferrarese (2002). An example of category **I b** is the idea by Dokuchaev & Eroshenko (2002) that primordial BHs of mass $\sim 10^5$ $M_\odot$at the centres of growing protogalaxies undergo multiple mergers to build up the presently observed correlations. Another model from the same category is the one by Menou, Haiman & Narayan (2001) that assumes the presence of SMBH precursors in only a small fraction of galaxies at high redshifts and then follows their evolution as they merge along with their hosts to build up the present day SMBH mass density. Here the mass of the precursor BHs is left unspecified since it could in principle be inferred from the merger gravitational wave signatures picked up by the LISA gravitational wave observatory.

A number of category **II a** type models suggest that SMBHs have gained most of their mass during the QSO epoch, when they were accreting gas at the centres of young host galactic nuclei. Feedback from the accretion process would then set up a $M_\bullet - M_{bulge}$ or $M_\bullet - \sigma$ (Haehnelt, Natarajan & Rees 1998; Cavaliere & Vittorini 2002) relation consistent with observations. A similar idea is that by Fabian (1999) , who proposes that an SMBH grows together with its host bulge. Most of the accreting SMBH's radiation is absorbed by cold gas at the centre, and it is possible that the winds from the accreting SMBH can eject the gas, thereby terminating further star formation and thus the growth of a stellar bulge. In this context no specific statement is made as to where the lighter precursor MBHs come from, but their formation can in principle precede the formation of the galaxy (c.f. Silk & Rees 1998). In this sense the model incorporates both, elements of category **I** and **II**.

Another model proposed by Adams, Graff & Richstone (2000), although part of the same category, is different from the ones above in that the mass of the accreting BH at the centre of a protogalaxy is actually determined by the initial dynamical properties of the protogalaxy (modelled as an isothermal sphere) and therefore not as a result of the BH's feedback on the bulge gas and stars. For a similar model where the SMBH and the halo as well as the bulge form at the same time through matter infall, see MacMillan & Henriksen (2003) .

Although in these models gas accretion may be the primary mechanism by which the SMBHs grow in mass, the merging of galaxies and the subsequent merging of central SMBHs to form more massive SMBHs is still always possible. There are, however, models that explicitly require both mechanisms to explain the correlations. The model by Kauffmann & Haehnelt (Haehnelt & Kauffmann 2000; Kauffmann & Haehnelt 2000) for instance, assumes that at early times SMBH formation through gas accretion is triggered by major mergers of halos/galaxies in the context of hierarchical structure formation and thus explains the onset of QSO activity. At late times galaxies are gas-poor and further SMBH mass growth proceeds by merging of lighter SMBHs only.

Concerning category **III**, SMBH formation following the formation of galaxies in general appears difficult, although in the context of hierarchical structure formation it could plausibly occur alongside the other options. For example, a galaxy with no previous SMBH could acquire one when merging with another galaxy or a galactic nucleus. Gas accretion induced in major mergers could subsequently lead to further significant growth of the acquired SMBH. This implies that in general category **III** applies mainly to galaxies whose mass was essentially in place well before the end of the QSO epoch, i.e. when significant amounts of gas for SMBH growth were still available.

So far all models mentioned, consider and make predictions for both correlations on an equal footing or if anything treated the $M_\bullet$ – M_{bulge} relation as the primary one. The considerably smaller scatter in the $M_\bullet$ – σ_c relation, however, implies that this may in fact be the more fundamental correlation. In this context Zhao, Haehnelt & Rees (2002) suggest that this relation may have been set up more recently by the central SMBH capturing stars and dark matter from an isothermal cusp at the galactic centre after the mass of both had essentially already been in place.

In either the merger and accretion models the explanations and assumptions regarding the origin of the seed BHs that grow into SMBHs remain somewhat unsatisfactory. No predictions are made that could be tested to discriminate between either of these assumptions. The seemingly most obvious candidate for the seed BHs - stellar mass BHs - is problematic for at least two reasons.

One is that the required mass increase by a factor $\sim 10^8 - 10^{10}$ [3] to get from stellar mass BHs to the most massive SMBHs is not easily achieved in either the merger or accretion scenario alone. In the accretion case, even if any particular stellar mass BH has been continuously accreting at the Eddington limit, it may still not reach the masses inferred for some of the most massive SMBHs, during the QSO epoch which lasted about 10^9 years (as inferred from the FWHM of the redshift evolution of the QSO space density; see Richstone et al. 1998). Assuming a BH accretes at the Eddington rate, we have

$$\eta \frac{\mathrm{d}Mc^2}{\mathrm{d}t} = L_{edd} \tag{3.4}$$

---

[3] E.g. Liang & Liu (2003) who find SMBH masses of up to $10^{10.2}$.

where the radiative efficiency $\eta \sim 0.1$ for standard thin disk accretion (see chapter 6). L_{edd} is the *Eddington luminosity*, which is defined as

$$L_{edd} \equiv \frac{4\pi G M m_p c}{\sigma_t} \approx 1.26 \times 10^{38} \left(\frac{M}{\mathrm{M}_\odot}\right) \mathrm{erg\ s}^{-1} \tag{3.5}$$

where σ_t is the Thomson scattering cross-section.

This gives rise to the Salpeter time scale, the time in which which a BH approximately doubles its mass

$$t_{salp} \sim \frac{\eta\ M\ c^2}{L_{edd}} = \frac{\eta\ \sigma_t c}{4\pi G m_p} \sim \eta\ 4.5 \times 10^8 \mathrm{yr} \tag{3.6}$$

which is independent of the BH mass. During the QSO epoch a BH can therefore increase its mass by only a factor f

$$f = \mathrm{Exp}\left[\ln 2 \frac{t_{qso}}{t_{salp}}\right] \approx (4 \times 10^9)^{1/\eta_{0.1}} \tag{3.7}$$

where $\eta_{0,1} = \eta/0.1$. Since the duty cycle (the optically bright phase) of any one QSO is much shorter than this [4], the implication is that only a small part of the most massive SMBHs today has actually been accreted during the optically bright QSO phase. This does not contradict the statement that most of the present day mass density in SMBHs is due to the gas accreted during the optically bright QSO phase [5], but some BH merging will be required to build up the extremely massive end (i.e. the SMBH end) of the BH mass function.

Conversely the build up of present day SMBHs from consecutive mergers of stellar mass BHs alone at the galactic centre is also problematic. The most important mechanism by which any two BHs are brought together to form a binary and eventually coalesce is by dynamical friction. The dynamical friction force experienced by a BH against the galactic stellar/gas background is proportional to the background density and the square of the BH mass. Therefore, at early times, when only stellar mass BHs are present, BH merging can only proceed efficiently (i.e. well within a Hubble time) in exceptionally dense environments, of which there must be a large

---

[4] This is derived using the observational $M_{smbh} - M_{bulge}$ relation, the luminosity function of spheroids and the evolution of the QSO luminosity function. Using this, one obtains a QSO duty cycle of about 0.01 (Haehnelt, Natarajan & Rees 1998; Ciotti, Haiman & Ostriker 2001), i.e. a QSO lifetime that is about 1/100 of the QSO epoch.

[5] Strictly speaking only gas/star accretion can actually increase the net mass density in BHs; the merging of BH, to form heavier BHs only changes the shape of the overall BH mass function but does not affect its normalisation or if anything reduces the net BH mass density slightly due to mass loss from gravitational radiation.

number to allow for the required number of mergers. However, in this case, for the BH growth dynamical interactions between individual BHs are probably more important ('runaway' growth of BHs; see section 3.3.2) than the dynamical friction induced formation and subsequent merging of individual BH binaries. While the very central regions of early (proto)galaxies may provide for such a dense environment after a major merger, there are simply not enough of these. BH mergers that are just triggered by major halo/galaxy mergers alone are not sufficient, either for building up SMBHs. We will return to this issue in the next section.

While gas accretion is more efficient in increasing the mass of BHs at early times (e.g. during the QSO phase), merging is more efficient at late times when SMBH precursors of intermediate mass ($\geq 10^2$ $M_\odot$) may have already been formed through accretion. This complementarity between the two invites combining (c.f. the Kauffman & Haehnelt model) them to achieve the required overall mass increase (c.f. above) by factor $10^8 - 10^{10}$.

It then seems altogether more plausible that the seed BHs were from the outset more massive than ordinary stellar mass BHs.

3.2.1 SMBH binaries at galactic centres

If SMBH growth is achieved to a large extent through the merging of lighter precursors, then we would expect the presence of SMBH binaries in galactic centres. How prevalent these are would depend on the lifetime of the binary before final coalescence. We here review briefly the individual stages of the merger process as outlined by Begelman, Blandford & Rees (1980) . They identify eight distinct stages

1. After merging of their respective host galaxies, two SMBHs and their associated dense stellar clusters sink to the centre of the galaxy merger remnant through dynamical friction on a galactic dynamical time scale.
2. By dynamical friction the SMBHs move to the centre of the now merged stellar distribution.
3. The SMBHs become bound, but their orbital velocity is still lower than the velocity dispersion of the surrounding stellar distribution.
4. As the binary separation continues to decrease the orbital period becomes shorter and the

total energy exchange with stars becomes less as kinetic energy from the binary is only efficiently transferred to stars encountering the binary on a timescale less than the binary orbital period.

5. Further stellar dynamical evolution - and resulting reduction in binary separation - is controlled by the diffusion of stars into more radial orbits that fall within the binary's 'loss cone'[6].
6. As the binary orbital velocity approaches the stellar velocity dispersion, the binary becomes 'hard'. Further evolution becomes very slow as every interacting star now carries away a minimum fraction of the binary's binding energy.
7. As the binary gets significantly tightly bound gravitational wave emission can reduce the binary separation further on a corresponding time scale.
8. As the binary components get closer evolution accelerates and rapidly leads to final coalescence with associated strong emission of electromagnetic radiation and gravitational waves.

3.3 Intermediate mass black holes

In other words the seed BHs may be constituted by intermediate mass BHs (IMBHs) of at least 10^2 $M_\odot$, i.e. about an order of magnitude more massive than the heaviest stellar mass BHs. In fact, it would seem natural to expect the presence of IMBHs in the wider mass range between $10^2 - 10^5$ $M_\odot$, connecting the stellar mass and super massive extremes, regardless of the particular BH growth mechanism.

3.3.1 The significance of IMBHs

In both the accretion and merging BH scenarios IMBHs should exist as an intermediate stage in the growth process, but have so far not been observed, except for a possible detection that has only recently been made. Gerssen et al. (2002) have performed HST/STIS observations on a number of stars in the central regions of dense globular clusters, indicating the presence of one

---

[6] These are stars with orbital pericentres less than the SMBH's radius of influence

or more central dark objects with a total mass of $\sim 10^3$ $M_\odot$ (van der Marel et al. 2002; Gerssen et al. 2002). The fact that IMBHs have so far not been observed otherwise indicates that the IMBH stage is possibly very short or that IMBHs are located in regions where BH specific detections are dominated by their stellar-mass or super-massive counter parts. Detection in the solar neighbourhood, for instance, is highly unlikely, since the number of IMBHs would be much lower than that of stellar mass BHs. Even if IMBHs were to be present in the solar neighbourhood, detection through gravitational lensing would not be an option as their mass implies a far too long lensing time scale. On the other hand, in galactic centres, their mass, that is significantly lower than that of any central SMBH, would make it difficult to resolve their individual gravitational influence on surrounding stars [7], while any radiative emission from accreting matter would be drowned out by the emission from the (SMBH at the) galactic centre.

Outside the centre of the galaxy, pointed observations of globular clusters may provide the only way of detecting the dynamical influence of IMBHs on their surroundings, but the observations alluded to above indicate the need for improved resolution in order to confirm the existence of a single IMBH at the globular cluster centre. Alternatively emission due to accretion onto IMBHs throughout the galaxy except for the centre may also be a way to observe them. We return to this in chapter 6, where we consider whether detections of ultra-luminous off-centre X-ray sources in a number of galaxies could be accounted for in this way.

3.3.2 Formation scenarios

Above we have argued that the seed BHs for either growth process (hierarchical merging or gas accretion during the QSO phase) should be IMBHs of at least 10^2 $M_\odot$. This means that the 'seed IMBHs' themselves must have been created by a different process beforehand. One such mechanism proposed by several authors is the formation of IMBHs through run-away merging of stellar mass BHs at the centre of dense stellar clusters (Ebisuzaki et al. 2001, Mouri & Taniguchi 2002, Miller & Hamilton 2002, Portegies Zwart & McMillan 2002, hereafter ZM02) . This idea is not new, and it had in fact originally been proposed that this process may be the direct path to the formation of SMBHs at the galactic centre (Shapiro & Teukolsky 1985). For example,

---

[7] Note that this has only just been achieved with HST for SMBHs.

ZM02 find that dense star clusters with an initial relaxation time of less than about 25 Myrs are dominated by stellar collisions. The relaxation time scale is given by:

$$\begin{aligned} t_{rlx} &= \left(\frac{R_c^3}{GM_c}\right)^{1/2} \frac{N_c}{8 \ln \Lambda_c} \\ &\approx 1.9 \times 10^6 \left(\frac{R_c}{\mathrm{pc}}\right)^{3/2} \left(\frac{M_c}{\mathrm{M}_\odot}\right)^{1/2} \left(\frac{< M_\star >}{\mathrm{M}_\odot}\right)^{-1} (\ln\Lambda_c)^{-1}\mathrm{yr} \end{aligned} \tag{3.8}$$

where R_c and M_c are the half-mass radius and the mass of the stellar cluster, $< M_\star >$ is the average mass of the stars and $N_c \equiv M_c / < M_\star >$ is the number of stars in the cluster. $\ln\Lambda_c$ is the Coulomb logarithm and typically of order $\ln\Lambda_c \approx \ln(0.1 N_c) \sim 10$.

In this scenario the more massive stars sink to the cluster centre by dynamical friction and the most massive star in the cluster will experience the first and dominate subsequent collisions leading to a runaway merging process that is terminated when the cluster gets tidally disrupted, e.g. in a host galaxy, or experiences mass loss due to stellar evolution. The final mass of the merged object could be as large as 10 percent of the total mass of the cluster. For instance for a 10^5 $\mathrm{M}_\odot$star cluster with $< M_\star > \sim 0.65$ $\mathrm{M}_\odot$(Scalo 1986) and $R_c \sim 0.4$ pc we obtain $t_{rlx} \approx 23.3$ Myr.

Although in this model IMBHs also grow through mergers of stars or stellar mass BHs, we nevertheless consider this process different from the hierarchical merging process that leads to SMBHs, because it occurs under entirely different circumstances. This is indicated especially by the time scale of oder 10^6 years that is much less compared to the corresponding time scale in hierarchical merging which would be closer to the Hubble time.

However, our enquiry, is based on a different mechanism to create IMBHs as described in the next chapter.

Chapter 4

Massive Black Holes from Population III remnants

In the previous chapter we have argued in favour of massive seed BHs that are produced by processes other than growth through mergers or accretion onto one or a few initially stellar mass BHs at the (proto) galactic centre. In this context 'other' means primarily the fact that seeds will have formed in an entirely different context beforehand, although ultimately they will still be created from the same primary objects, i.e. stellar mass BHs.

An alternative and more immediate route to seed MBHs would be their creation as a result of the direct collapse of very massive stars of similar or higher mass. The present day abundance of stars does not seem to accommodate this possibility, as the most massive stars have masses of at most $\lesssim 10^2$ $M_\odot$.

In chapter 2 we noted that the fragmentation scale that determines the size of baryonic objects inside dark matter halos depends in the first instance on the relation between the dynamical and cooling time scales. At high redshifts $30 \gtrsim z \gtrsim 15$, cooling time scales are expected to have been substantially larger due to the lack of metals in the IGM. Consequently the fragmentation scale is a lot larger, too, and it is conceivable that stars of considerably larger mass may have formed under these conditions. Since stars are responsible for the metal enrichment of the IGM, the implication is that these stars must constitute the first generation of stars to form.

In this chapter we present a number of analytical arguments to investigate star formation under these conditions and explore to what extent these stars could evolve into MBHs. We start by giving analytical estimates of the formation epoch and sizes of the dark matter halos and baryonic clouds within which these stars form in section 4.1. In section 4.2 we describe the numerical procedure we began to use to investigate the formation of dense baryonic cores which form inside the clouds. These cores are the birthplace of the first stars. In section 4.3 we consider how the stars evolve, how feedback affects the star's host cloud as well as neighbouring clouds and address the question of how the stars turn into MBHs. We conclude with a summary in section 4.4.

4.1 Baryonic collapse at z $\gtrsim$ 15 – analytical estimates

We have seen in chapter 2 that the criterion for a collection of baryons to form a gravitationally bound configuration is that their total mass is larger than the Jeans mass. For baryonic clouds at the present day this is also the main criterion for them to collapse and cool, since cooling proceeds efficiently once they have become bound. This is both because the metal enrichment of the IGM provides more cooling channels but also because baryonic clouds are initially heated to the virial temperature of the dark matter halo within which they collapse. Due to the more massive dark matter halos, at the present day this temperature is typically much larger than at high redshifts, which implies that the larger number of cooling channels available (that are effective at this and lower temperatures) can actually be accessed.

4.1.1 Cooling of primordial molecular clouds

The situation is very different at redshifts greater than about z $\sim$ 20 (which we will subsequently refer to as *high redshifts*). Baryonic clouds inside DM halos forming at these redshifts are constituted by an essentially primordial abundance of elements, since metals will only be injected by the first stars which have yet to form. Under these circumstances cooling can only occur via a restricted range of interactions mainly involving collisional cooling of molecular hydrogen as well as line cooling from various atomic species (Li, D, ^{3}He, ^{4}He). In addition, since structure forms hierarchically, dark matter halos collapsing at high z will be rather small and have a low associated virial temperature, which means that not all of the already restricted range of cooling channels may actually be used. Both effects result in a much longer cooling time scale for baryonic collapse than today. For the smallest DM halos with the lowest virial temperatures ($< 10^4$ K), molecular hydrogen provides the only effective cooling channel. In this case cooling proceeds via collisions of H_2 (with other species, typically neutral hydrogen since it is the most abundant one) and the cooling time (c.f. equation 2.59) is given by

$$t_{cool} = \frac{\frac{3}{2} n_H k_B T}{n_{H_2} n_H \Lambda_{cool}(T)} = \frac{3 k_B T}{2 n_H \Lambda_{cool}(T)} \frac{1}{f_{H_2}} \tag{4.1}$$

where f_{H_2} is the H_2 fraction relative to neutral hydrogen and Λ_{cool} is the temperature dependent cooling rate. Λ_{cool} also depends on the baryonic density although this effect is negligible for the relatively low densities we are considering here ($n_H \sim 10^{-4} \mathrm{cm}^{-3}$) and which are typical for objects collapsing at high redshifts.

Because cooling proceeds so slowly, it is not a sufficient collapse criterion anymore that a collection of baryons is more massive than the Jeans mass; it must in fact also be massive enough for cooling to be accomplished in less than a Hubble time. The latter is referred to as the cooling mass and is determined by the condition that $t_{cool}(\rho, T_{vir}) = t_{hubble}(\rho)$, where $\rho(z)$ is the physical matter density. Since T_{vir} depends on the mass and radius of the halo within which baryons condense, this can then be solved for $M_{cool} = M(T, z)$. Today M_{cool} determines the maximum mass for galaxies to be of order 10^{13} $\mathrm{M}_\odot$, which is why for masses larger than this we observe clusters of galaxies (i.e. fragments) rather than single cluster sized galaxies (Silk 1977; Rees & Ostriker 1977). The cooling rate for primordial species as a function of temperature is shown in figure 4.1.

In hierarchical structure formation $T = T_{vir}$ is dependent on the mass and radius of a collapsing dark matter halo, which in turn depend on the redshift (c.f. equation 2.47). In this case M_{cool} therefore becomes a function of z only. $M_{cool}(z)$ is plotted as the blue solid curve in figure 4.2, where we have used the cooling rate from Fuller & Couchman (2000).

The redshift evolution of M_{cool} based on H_2 cooling shown in figure 4.2 has been plotted to the present day, but we know that at some point between the high redshift regime and now, H_2 cooling will cease to be relevant as the typical virial temperatures of halos collapsing at later times are beyond the range where H_2 cooling acts efficiently and in addition other cooling channels open

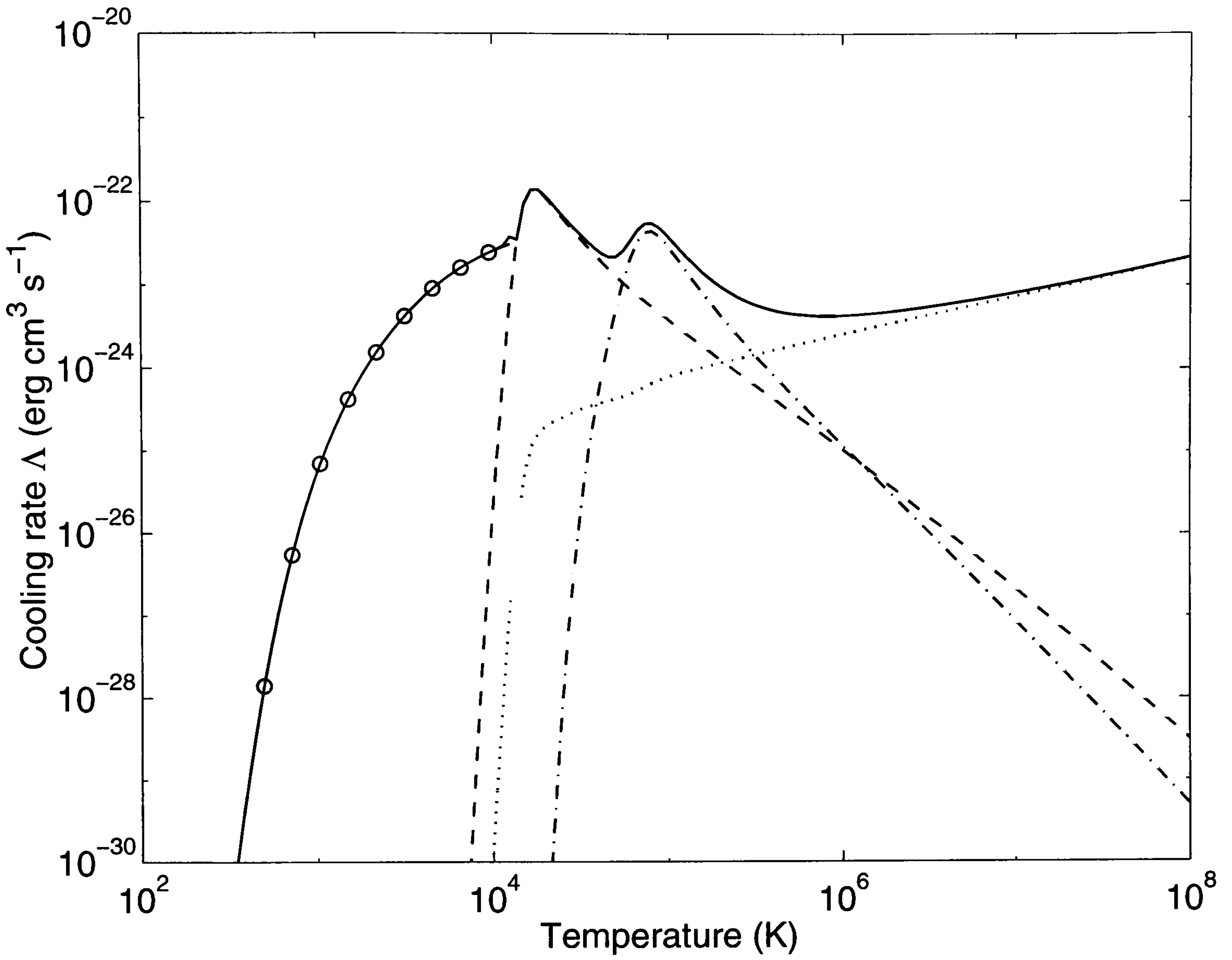


Figure 4.1: "Cooling rates for Bremsstrahlung (dotted line), H (dashed line), and He (dash-dotted line) line cooling, and (circles) cooling. The e^-, H^+, H_2, He^+, and He^{++} abundances were computed assuming collisional equilibrium, and the H_2 fractional abundance was 3×10^{-4}, which is typical for early objects". Taken from Fuller & Couchman (2001).

up due to metal enrichment.

We will subsequently refer to the DM halo mass corresponding to M_{cool} at redshifts $\gtrsim$ 15 as the *Minimum Mass for Baryonic Collapse*, M_{MBC} and to the halos themselves as *MBC halos.* Depending on the cosmology this will typically be of the order of 10^5 – 10^6 $M_\odot$.

Since the cooling time scale is comparable to the Hubble time and thus also to the collapse time scale (which is essentially the same as the Hubble time to within a factor of order unity), the implication is that no significant fragmentation should occur inside the cooling clouds. We would expect the clouds to fragment into at most a few sub-clumps. These sub-clumps are likely to host the dense baryonic cores within which the first stars will eventually form.

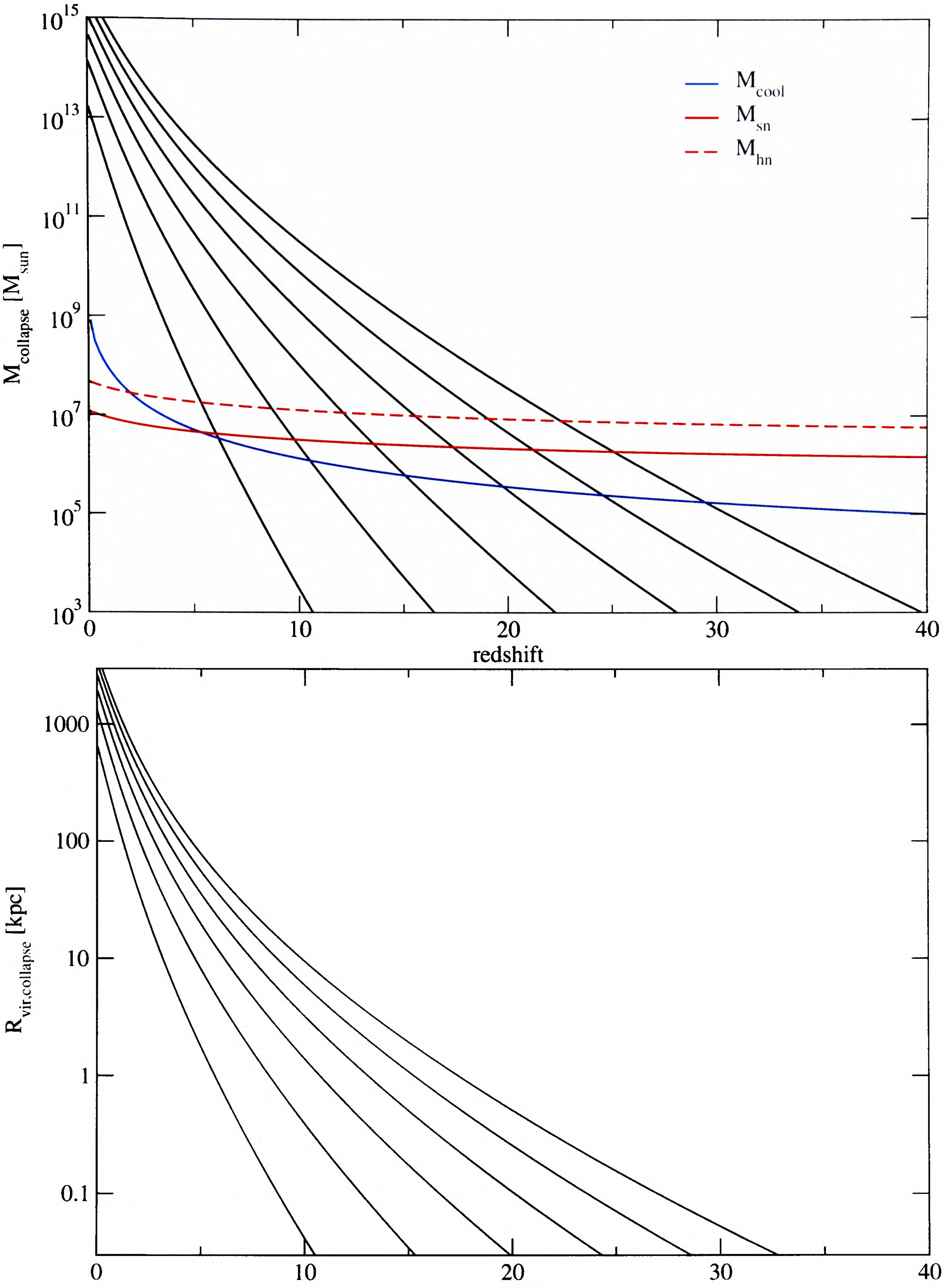


Figure 4.2: The top panel shows virial masses for halos that collapse from 1, 1.5, 2, 2.5, 3, 3.5σ peaks together with the redshift dependent cooling (MBC) mass. The two red lines represent halo masses with binding energies equal to the kinetic energies of super- and hypernovae. The bottom panel shows the virial radii corresponding to the halo masses for the respective collapse redshifts (in same order from top to bottom).

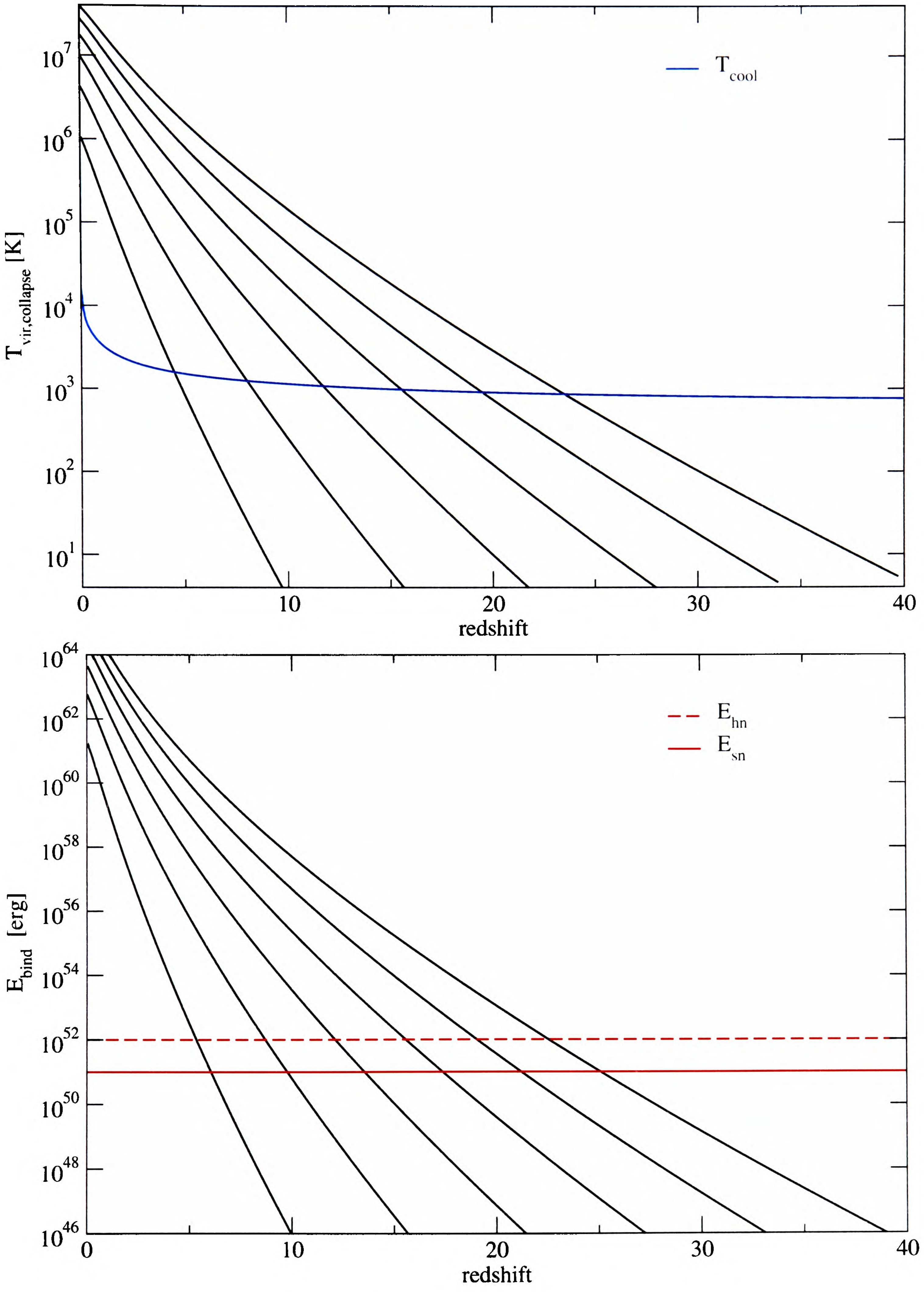


Figure 4.3: Same as figure 4.2 but for halo virial temperature and binding energy.

4.1.2 Collapse of 1, 2, 3 σ peaks

In the introduction to this chapter we mentioned that the baryonic clouds, made up of a primordial abundance of elements, would be where the first stars could form. The nature and evolution of these stars will be discussed below but we can already say that they are special in the sense that they are forming in a zero metallicity environment. Stars of this kind are referred to as *population III stars* - as opposed to populations I and II for stars today.

So far we have only mentioned that these baryonic clouds, and therefore the first stars, form if the cloud mass (and correspondingly that of their associated DM halo) is above some minimum mass, where H_2 cooling can proceed efficiently, but we have not actually specified when this would happen. Under the assumption that the initial matter density field is gaussian, there will always be some peaks that correspond to masses high enough to host baryonic clouds inside which stars could form. This means, that at any time there has always been **some** probability that 'first' stars would form [1]. To illustrate this we have calculated the masses of halos corresponding to 1 – 3.5 σ peaks in the initial gaussian matter density field as a function of collapse redshift. This is shown in the top panel of figure 4.2 together with the evolution of the H_2 cooling mass. The bottom panel shows the corresponding halo virial radii from eq. 2.47. In figure 4.3 we have further plotted the virial temperatures as well as the binding energies for such halos using equations 2.49 and 2.50. We see that at a redshift of about 30, dark matter halos corresponding to 3.5 σ peaks are actually massive enough for H_2 cooling to be efficient and thus for population III stars to form. To specify the *epoch* of first star formation, the question is then at what time first generation stars were numerous enough to affect the IGM (i.e. inject metals) such that subsequent star formation would essentially take place under present day conditions. This will depend on the height of the collapsing peaks within which first generation stars form. In particular we want to know what the minimum number density of population III star forming regions would have to be in order to raise the metallicity Z_M, in present day galaxies to a value of at least $Z_M \sim 10^{-5}$, roughly the minimum level required for present day star formation (Bromm et al. 2001). To estimate this

---

[1] But at extremely high redshifts, $z \gtrsim 100$, the baryonic density will become so high that the H_2 cooling rate for the low density regime that we have used will no longer be valid and H_2 cooling could proceed significantly more efficiently leading to a drastically reduced cooling mass. However, this is ultimately limited by the CMB, since cooling can at most progress to a minimum temperature equal to that of the CMB at the time, $T_{cmb} \sim (1+z)$ 2.7 K.

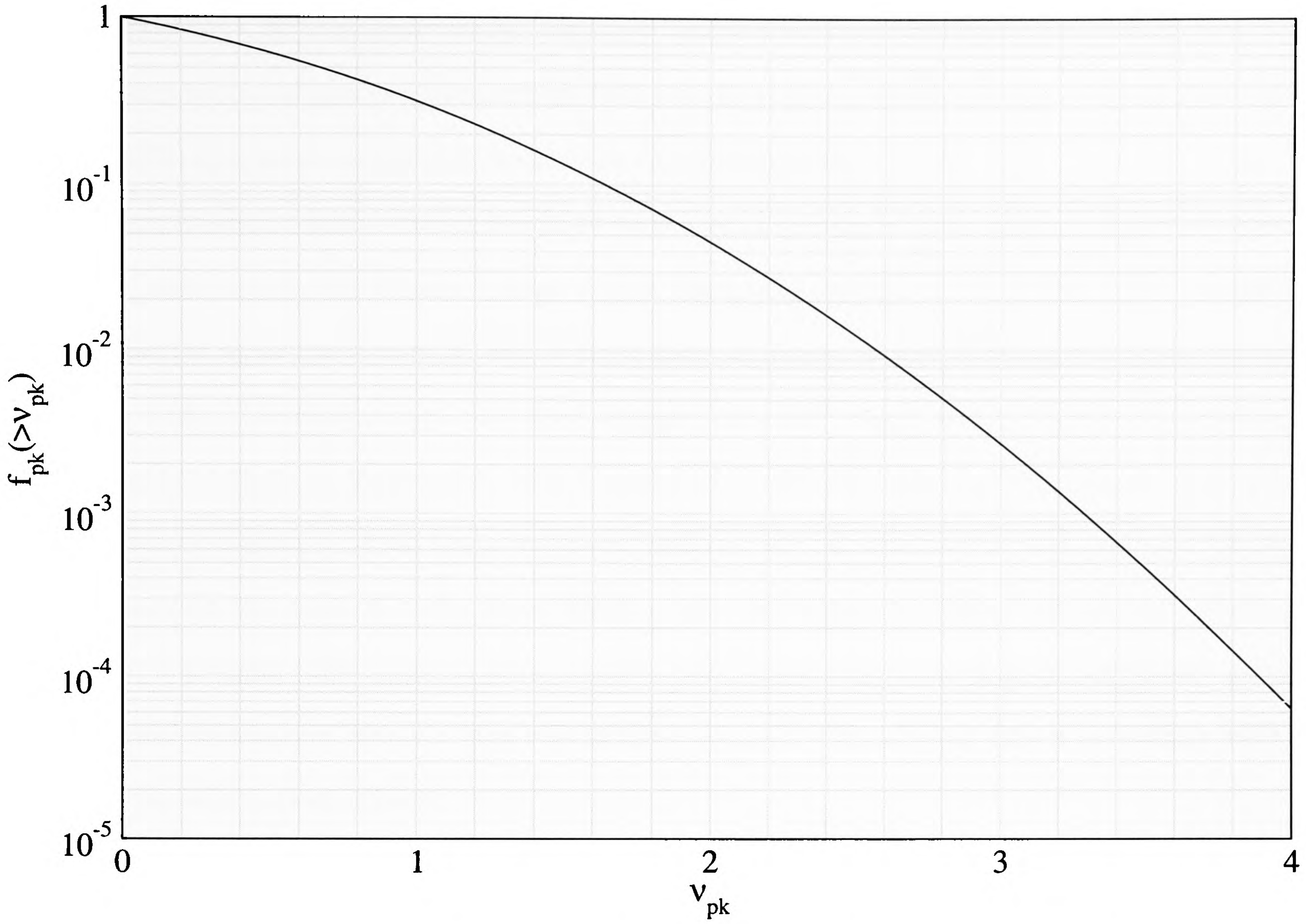


Figure 4.4: Fraction of mass contained in density peaks of height ν_{pk} and above.

we first need to know the fraction of mass $f_{pk}(>\nu_{pk})$ in the initial density field that is contained in peaks of height ν_{pk} or above. For a gaussian field this is

$$f_{pk}(>\nu_{pk}) = 1 - \mathrm{Erf}\left[\frac{\nu_{pk}}{\sqrt{2}}\right] \tag{4.2}$$

If the fraction of metals from stars in halos collapsing from ν_{pk} σ peaks at redshift z is

$$f_Z(\nu_{pk}, z) = \frac{M_{sn}}{M_b(\nu_{pk}, z)} \tag{4.3}$$

where M_{sn} is the mass of metals expelled by super novae and M_b the mass of baryons, then the total metallicity due to these stars is

$$Z = f(\nu_{pk}) f_Z(\nu_{pk}, z) \tag{4.4}$$

For $Z = 10^{-5}$ this could be solved for ν_{pk}, however, $f_Z(\nu_{pk}, z)$ is not known *a priori* as it requires knowledge of the IMF of population III stars and the mass of their associated baryonic clouds. The redshift and height of peaks within which first star formation occurs can thus not be estimated directly, although some analytical constraints can be imposed.

For supernovae to metal-enrich the IGM their ejecta will have to be able to escape their host halo. In other words, the kinetic energy of the supernova ejecta, E_{sn} must be high enough to enable them to leave the potential well of their halo, as characterised by the halo binding energy E_{bind} (eq 2.50). Supernovae typically have energies of $\sim 10^{51}$ ergs, although massive population III stars may explode as *hypernovae* with energies of $\sim 10^{52}$ ergs (see e.g. Nakamura et al. 2001 and references therein). This leads to upper limits of 10^6 $M_\odot$(for supernovae) and 10^7 $M_\odot$(for hypernovae) on the mass of halos from which ejecta can escape as shown by the red curves in figure 4.2. Potential wells of halos more massive than these limits can still get deformed during mergers and become so shallow that ejecta may escape. Considering this a secondary effect, however, we will neglect it here.

Together with the cooling mass criterion this already restricts the mass range of the first star forming halos to roughly $10^5 - 10^6$ $M_\odot$for supernovae ($10^5 - 10^7$ $M_\odot$for hypernovae) and thus the mass of the baryonic component to a corresponding fraction f_b of this. Finally, we still need an estimate of M_{sn}. Pre-empting some of the arguments in the next section we assume that one pop III star forms with a mass of order $\sim 10^2$ $M_\odot$in every baryonic cloud where first star formation occurs. If at the end of its life time that star undergoes a super - or hypernova and expels a large fraction of its mass as metals such that $M_{sn} \sim 10^2$ $M_\odot$, then $f_Z(z) \sim 10^{-2} - 10^{-4}(10^{-3})$. Putting this into equation 4.4 we obtain $2.6(1.6) < \nu_{pk} < 3.3$, and the formation redshift $18(16) < z < 27$. This is no more than an approximate result. $2 - 3.5\sigma$ peaks are highly clustered and lie in regions with general large scale overdensities. This implies that ejecta really have to overcome a potential barrier that might be much larger than just that of their host halo, in order to leave the whole overdense region altogether. While ejecta might thus enrich the (local) IGM between the $2 - 3.5\sigma$ peak halos, they may not be able to enrich the whole of the Universe uniformly, as we had essentially assumed above.

In arriving at the above estimate the supernova constraint was crucial. However, this estimate is consistent with that from semi-analytical studies that compute the full network of chemical rate equations in the context of hierarchical structure formation models (Tegmark et al. 1997 , Hutchings et al. 2002), although they obtain a much more accurate estimate of the minimum mass of the first baryonic clouds. For example Tegmark et al. find that in standard CDM cosmologies the first baryonic clouds form at $z = 27$ and have masses of about 10^5 $M_\odot$, while Hutchings et al. using a different cooling function find that for both SCDM and ΛCDM models form at $z \approx 20$ and have masses of 10^6 $M_\odot$. Similar results are obtained in studies that incorporate the computation of the rate equation network into hydrodynamical simulations (Abel et al. 1998, Abel, Bryan & Norman 2000, Fuller & Couchman 2000, Bromm, Coppi & Larson 2002).

4.1.3 Qualitative features of the primordial stellar IMF

Hydrodynamical simulations can follow the evolution of collapsing baryonic clouds much further than possible by analytical estimates; in principle so far until radiative feedback from any forming proto stellar objects becomes important. A determination of the IMF of the first stars is therefore not possible although certain trends can be deduced from the mass function of baryonic cores that are forming.

Here we just give some analytical arguments to qualitatively describe to what extent the IMF of the first stars would be expected to be different from that at the present day. This is mainly due to the different way in which cooling proceeds in primordial baryonic clouds. A consequence of the rather long cooling time in the primordial clouds is that subsequent break-up into smaller units leads to relatively large fragments. By the same token seed masses within these fragments can in principle accrete large amounts of matter from the cloud without further fragmentation occurring (Omukai & Palla 2001;Bromm, Coppi & Larson 2002). Only radiation pressure from the proto-star on the infalling layers of material could halt accretion and so limit the mass of the star. However, in the absence of dust the infalling matter has too low an opacity for radiation pressure to be significant (Ripamonti et al. 2002). In these stars the role of winds that could otherwise lead to significant mass loss in population I stars, is also negligible. As a result this will likely lead to the creation of very massive stars, potentially as heavy as 10^4 $M_\odot$.

While massive primordial stars seem plausible, it has so far not been possible to establish the primordial equivalent of an initial mass function (IMF), except for arguments in favour of qualitative features (Nakamura et al. 2001). Even the case for only one potentially very massive star forming is not ruled out. Once switched on, its radiative feedback into the host cloud could be strong enough to prevent any subsequent stars from forming within its lifetime. The total ionising flux will depend on the primordial IMF and could potentially be a significant, if not decisive source for early reionisation of the Universe.

4.2 Baryonic collapse at $z \gtrsim 15$ - hydrodynamical simulations

We attempted to run hydrodynamical simulations of baryonic collapse at high redshifts. The aim was in particular to determine or obtain as good an indication as possible of where and how many population III stars form, and what their masses would be. The objective was not to determine the redshift of when first star formation occurs; in fact, since the simulated volume of the Universe is limited, we have to move down to redshifts of $z \sim 15$ rather than 25 to sufficiently increase the probability of finding a collapsing halo with a mass comparable to that of MBC halos at higher redshifts. At $z \sim 15$ the collapsing halos of mass $\sim 10^5$ $M_\odot$correspond to 1.5σ peaks (c.f. figure 4.2) and their comoving number density is of order ~ 10 or so larger.

4.2.1 Initial conditions and parameters

For the simulations, the hydrodynamical code *Enzo* (Bryan, Abel & Norman 2000) was used, which contains a N-body and an Euler solver for dark matter and gas respectively and combines this with a computation of the rate equations for primordial gas physics. Most importantly it features adaptive mesh refinement and so allows an unprecedented dynamic range for the detailed simulation of the densest parts of collapsing baryonic regions. For technical details of the code the reader is referred to the paper by Bryan, Abel & Norman (2000) .

We started a set of simulation runs for our fiducial ΛCDM cosmological model ($\Omega_m = 0.3, \Omega_\Lambda = 0.7, \Omega_b = 0.02\ h^{-2}, \sigma_8 = 0.9, h = 0.7$) in a box of comoving side length $L = 0.116$ h^{-1} Mpc starting at a redshift of $z = 99$, i.e. long before the evolution of any structure on or near the scales we are interested in, turned non-linear.

4.2.2 First results

Only a preliminary set of runs could be completed. Subsequent simulations that were planned to be run to a larger number of adaptive refinement levels and thus to much higher resolution could not be completed. This was due to problems that are probably connected to the particular hardware configuration used. We have not been able to resolve these difficulties. We did, however, find that only one highly dense core region formed, which is in accordance with results from other simulations (Bromm, Coppi & Larson 2002).

4.3 Massive population III stars and black holes

4.3.1 Formation of the first stars

As better computational resources become available, the hydrodynamical code used could in principle follow the collapse of baryonic cores to the highest required resolution. However, to compute the formation of the first stars we are still missing one crucial ingredient and that is the inclusion of radiative transfer. This would lead inevitably to feedback effects from the radiation of the (forming) star. These affect the velocities and mass flows at small scales and ultimately determine how massive a star is, when it switches on, its mass accretion rate, temperature, etc. Apart from 1-D simulations (Haardt et al. 2002) there are to date no fully 3-D simulations of proto-stellar collapse and resulting star formation that also account for the role of radiative feedback.

In the following we take our results from the simulations as our starting point and assume that every baryonic core created will lead to the formation of one population III star. We will continue to use analytical arguments to put limits on the time scale and mass of the stars forming.

We first want to determine the dominant time scale. One is the collapse time scale and may be taken as the time between the formation of the baryonic cores and that of the sink particles. Once formed we want to compare this time scale with the star's life time on the main sequence. For population III stars of mass $M_\star \sim 200$ $\mathrm{M}_\odot$this is of the order of $\sim 2 \times 10^6$ years (El Eid, Fricke & Ober 1983; Marigo et al. 2001). At least for stars in this mass range the lifetime is therefore the dominant time scale.

An indication of the upper limit on the mass of the star would be provided by the mass accretion rate of the sink particles in our simulations. The formation of these stars has also been investigated analytically although it is unclear if any specific upper limits to their mass can be determined (Omukai & Palla 2001; Omukai & Inutsuka 2002).

4.3.2 Radiative feedback from the first stars and its effect on the host cloud

In both the analytical estimates and the simulations we have neglected the effect of the radiative feedback of the first stars on their own host clouds and halos as well as neighbouring star forming regions. The main effect of the radiation from the first stars would be to ionise neutral hydrogen. H_2 could thus cease to be available as the main coolant for other stars that are about to form in the vicinity.

To estimate the maximum extent of this effect we consider a single population III star of mass 300 $M_{\odot}$at the centre of a halo of mass $M_{MBC} = 3 \times 10^5$ $M_{\odot}$collapsing at $z \sim 25$. We further assume that the baryonic mass of the halo $M_b = (\Omega_b/\Omega_m)M_{MBC}$ is distributed uniformly in a sphere extending out to a radius r_b that is equal to the virial radius of the halo. Then the number density of neutral hydrogen in this sphere is

$$\bar{n}_H = \frac{0.76}{m_p} \frac{M_b}{\frac{4}{3}\pi r_{vir}^3} \approx 0.33\mathrm{cm}^{-3} \tag{4.5}$$

where the factor 0.76 accounts for the hydrogen mass fraction of the total baryon mass. Clearly the true baryon density will be higher, since the baryons will have cooled and thus contracted inside the halo before forming the first star. However, the above allows to calculate an upper limit beyond which ionisation due to the star at the centre should not occur. In the absence of recombinations this maximum radius is given by (Barkana & Loeb 2001)

$$r_{max,ion} = \left(\frac{3}{4\,\pi}\,\frac{N_\gamma}{\bar{n}_H}\right)^{1/3} \tag{4.6}$$

where N_γ is the number of ionising photons, i.e. those with an energy of $h\nu > 13.6$ eV. To calculate this we look at the rate per unit stellar mass at which ionising photons are emitted

$$\dot{N}_\gamma = \frac{1}{M_\star} \int_{13.6eV}^{\infty} \frac{\mathrm{d}L_\nu}{\mathrm{d}\nu} \frac{\mathrm{d}\nu}{h\nu} \tag{4.7}$$

To zeroth order the spectrum of massive population III stars is described by a black body with effective temperature $T_{eff} \sim 10^5$ K (Bromm, Kudritzki & Loeb 2001). For stellar masses above about 300 $M_{\odot}$this spectrum is almost independent of mass. In particular, for a 300 $M_{\odot}$pop III star, $L = 6.03 \times 10^6 L_{\odot}$ and $T_{eff} = 1.1 \times 10^5$ K so that one obtains $\dot{N}_{\gamma} \approx 1.4 \times 10^{48} M_{\odot}^{-1} \mathrm{s}^{-1}$ for stars above 300 $M_{\odot}$. This corresponds to about 40 % of all photons emitted [2]. Assuming the star's luminosity remains constant over its lifetime, $t_{\star}$, the total number of ionising photons emitted is then

$$N_{\gamma} \approx 1.32 \times 10^{63} \left(\frac{t_{\star}}{10^5\ yr}\right)^{1/3} \left(\frac{M_{\star}}{300\ M_{\odot}}\right)^{1/3} \tag{4.8}$$

Putting this into eq. 4.6 we find for a 300 $M_{\odot}$star with a lifetime $t_{\star} = 10^5$yr that $r_{max,ion} \approx 8$ kpc in comoving units, corresponding to a physical distance of about 0.3 kpc, which is about 3 times larger than the halo virial radius. I.e. had the baryonic distribution extended out to this radius (with the same constant density) all hydrogen atoms in this volume would have become ionised. If instead we consider a single star with a mass of $M_{\star} \sim 3 \times 10^4$ $M_{\odot}$(i.e. of the order of the total baryonic mass contained in the halo), $r_{max,ion}$ is larger by a factor $100^{1/3}$ and becomes $r_{max,ion} \approx 1.5$ kpc (37 kpc comoving). As we will see below the latter scale already corresponds to a volume that could contain a non-negligible number of neighbouring halos in which star formation could thus be affected.

The inclusion of recombination of ionised hydrogen atoms (H II) changes this conclusion significantly. The *Strömgren* sphere is defined as the volume within which ionisation and recombination rates are equal

$$\alpha_B \bar{n}_H^2 V_{Str} = \frac{dN_{\gamma}}{dt} \tag{4.9}$$

where α_B is the recombination coefficient, which is $\alpha_B \approx 2.6 \times 10^{13} \mathrm{cm}^3 \mathrm{s}^{-1}$ for hydrogen at a temperature of $T = 10^4$ K (Barkana & Loeb 2001). In the absence of recombination the maximum volume for ionisation is proportional to the inverse of $\bar{n}_H$. Including recombination, however, V_{Str} now scales as $\bar{n}_H^{-2}$ provided the star's luminosity and therefore the emission rate of ionising photons stays constant, i.e. $\dot{N}_{\gamma} = N_{\gamma}/t_{\star}$. Substituting $\bar{n}_H \approx 0.33 \mathrm{kpc}^{-3}$ (c.f. eq. 4.5) into eq. 4.9 one gets a proper volume of $V_{Str} \approx 0.5 \mathrm{kpc}^3$ corresponding to a physical radius of

---

[2] In contrast the number of photons required to photo dissociate H_2 (11.2 - 13.6 eV) is about 7.5 % of all photons emitted.

$r_{Str} \approx 0.5$ kpc. In other words, would the star actually live long enough - and therefore not be constrained by $r_{max,ion} \approx 0.3$ lpc (c.f. above)- the ionised volume would eventually reach out to 0.5 kpc physical distance from the centre. However, if we imagine that the baryons are uniformly distributed out to only 1/2 the halo virial radius, the Strömgren sphere constraint becomes more important $r_{max,ion}$ reduces to 0.16 kpc, but r_{Str} becomes 0.13 kpc etc. Since both $N_\gamma \propto M_\star$ and $\dot{N}_\gamma \propto L \propto M_\star$ raising the stellar mass increases r_{Str} by the same factor as $r_{max,ion}$.

In the above cases the ionised volumes were still larger than the **actual** volume that contained all the baryons. We can finally calculate the size of a Strömgren sphere corresponding to this volume by requiring that $V_{Str} = (4\pi/3)r_b^3$. Then, $n_H = N_H/V_{Str}$ where $N_H = 0.76\ M_b/m_p$ is the total number of hydrogen atoms in the halo. Using this in eq. 4.9 one obtains

$$V_{Str} = \frac{\alpha_B\ N_H^2}{\dot{N}_\gamma} \tag{4.10}$$

In our case this leads to $r_{Str} = r_b \approx 0.02$ kpc, about 1/5 of the halo virial radius.

The bottom line is then that if all baryons are uniformly distributed in a sphere with a radius of about no more than 1/5 r_{vir}, ionisation of hydrogen is expected to be limited to the extent of the baryons inside the halo.

4.3.3 Supernova feedback from the first stars and its effect on the host cloud

In contrast the effect of supernovae on their host cloud is straightforward. As far as MBC halos are concerned supernovae will expel virtually all gas. Any remaining gas will soon be chemically enriched. If any gas should have remained in the form of other dense gaseous cores that were about to produce a population III star, these will quickly acquire metallicity levels that prevent any further pop III star formation. As a result low mass star formation as in the present day Universe may have already occurred in the MBC halos at redshifts larger than 20.

4.3.4 Effect of feedback on neighbouring clouds

As the sites of first star formation correspond to high ($\nu_{pk} \gtrsim 2.5\sigma$) peaks in the initial density field, they will be highly clustered, particularly around overdensities on large scales that are associated with long wavelength modes of the matter power spectrum. Since the density fluctuations are gaussian on all scales and therefore both on the scales of the 3σ peaks as on the scales of the long wavelength overdensities, the result is that the correlation function $\xi_{pk}(r)$ of the peaks is just a linear amplification of the correlation function of the underlying matter density field, $\xi(r)$ (Bardeen et al. 1986)

$$\xi_{pk}(r) \simeq \frac{\nu_{pk}^2}{\sigma(M_{pk})^2}\,\xi(r) \tag{4.11}$$

where $\nu_{pk} = \delta/\sigma(M_{pk})$ and $\sigma(M_{pk})$ is the variance of the density field smoothed on scales corresponding to the mass contained in the peaks, which in our case is the mass of the MBC halos M_{MBC}. Eq. 4.11 holds for scales $r > 3r_{pk}$ where r_{pk} is the scale of the peak (e.g. $r_{pk} \sim 10$ kpc for $M_{MBC} \sim 3 \times 10^5$ $M_\odot$collapsing from 3σ peaks at $z \sim 25$)

$\xi(r)$ in turn can be determined directly from the matter power spectrum $P(k)$ (c.f. equation 2.32). $\xi(r)$ and $\xi_{pk}(r)$ for various peak heights are plotted in figure 4.5. We find that for 2.5, 3 and 3.5 σ peaks, ξ_{pk} is boosted over ξ by a constant factor 10.8, 23.4 and 41.9 respectively. These amplification factors are obtained using a top-hat filter to smooth $\sigma(M_{MBC})^2$ in eq. 4.11. The use of different filters should only lead to minor corrections, since ξ_{pk} is fundamentally determined by the underlying gaussian density field, not by the characteristics of any particular filter applied to it. For different redshifts the linear amplification factors will remain the same although the overall normalisation of the correlation functions will change in accordance with the linear growth factor.

The strong clustering of high peaks gives rise to the possibility that some peaks may be located closely enough, some distance r_{fb} from each other, say, such that ionising radiation from a star forming in one peak may be strong enough to suppress any star formation in the other peak. r_{fb} can in principle be determined from the Strömgren sphere and will depend on both the luminosity and the lifetime of the star.

If one star forms ordinarily per peak (as implied by the simulations), we can now ask what the mean number of peaks would be that lie within a distance r_{fb} of another peak. To answer this we first look at the differential probability, δP of finding one peak inside a volume element δV some distance r away from another one (Peebles 1980)

$$\delta P = n_{pk}\ \delta V[1 + \xi_{pk}(r)] \tag{4.12}$$

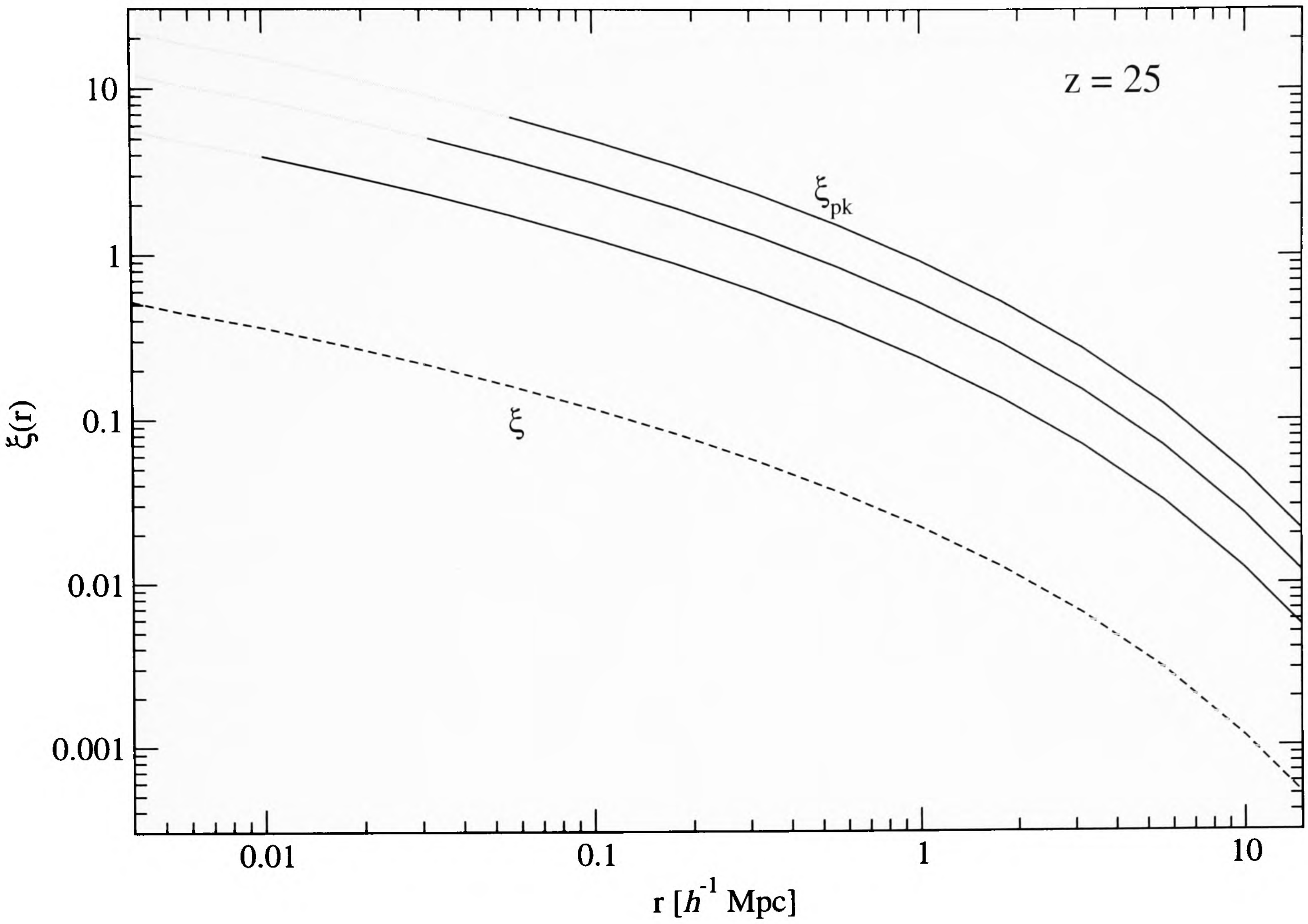


Figure 4.5: Correlation functions for comoving distances are shown for the matter density field (dashed) as well as for 3.5, 3, 2.5 σ peaks collapsing at $z = 25$ (solid lines, from top). The peak correlation functions ξ_{pk} are amplified over the matter correlation function ξ by a constant factor as described in the text. The grey portions of the ξ_{pk} curves indicate where the linear amplification approximation eq. 4.11 breaks down as the correlation scale approaches the size of the peaks themselves.

where n_{pk} is the mean number density of peaks. The exact derivation of the number density of peaks of a given height and scale is quite involved (Bardeen et al. 1986), but in the high peak limit converges to

$$n_{pk}(\nu_{pk}) \simeq \frac{1}{(2\ \pi)^2}\left(\frac{\sigma_1^2}{3\ \sigma_0^2}\right)^{3/2}(\nu_{pk}^2 - 1)\exp[-\nu_{pk}^2/2] \tag{4.13}$$

where σ_0^2 and σ_1^2 are moments of the matter density field smoothed on the scale r_{pk}

$$\sigma_i(r_{pk}) = \frac{1}{2\pi^2}\int P(k)W(kr_{pk})\ k^{2i}\ \mathrm{d}^3k \tag{4.14}$$

Eq. 4.12 can be integrated to obtain the mean number of neighbouring peaks within some distance r of one peak

$$< N_{pk}(r) >= \frac{4}{3}\pi\ r^3\ n_{pk} + 4\pi\ n_{pk}\int_0^r \xi_{pk}(r')r'^2\mathrm{d}r' \tag{4.15}$$

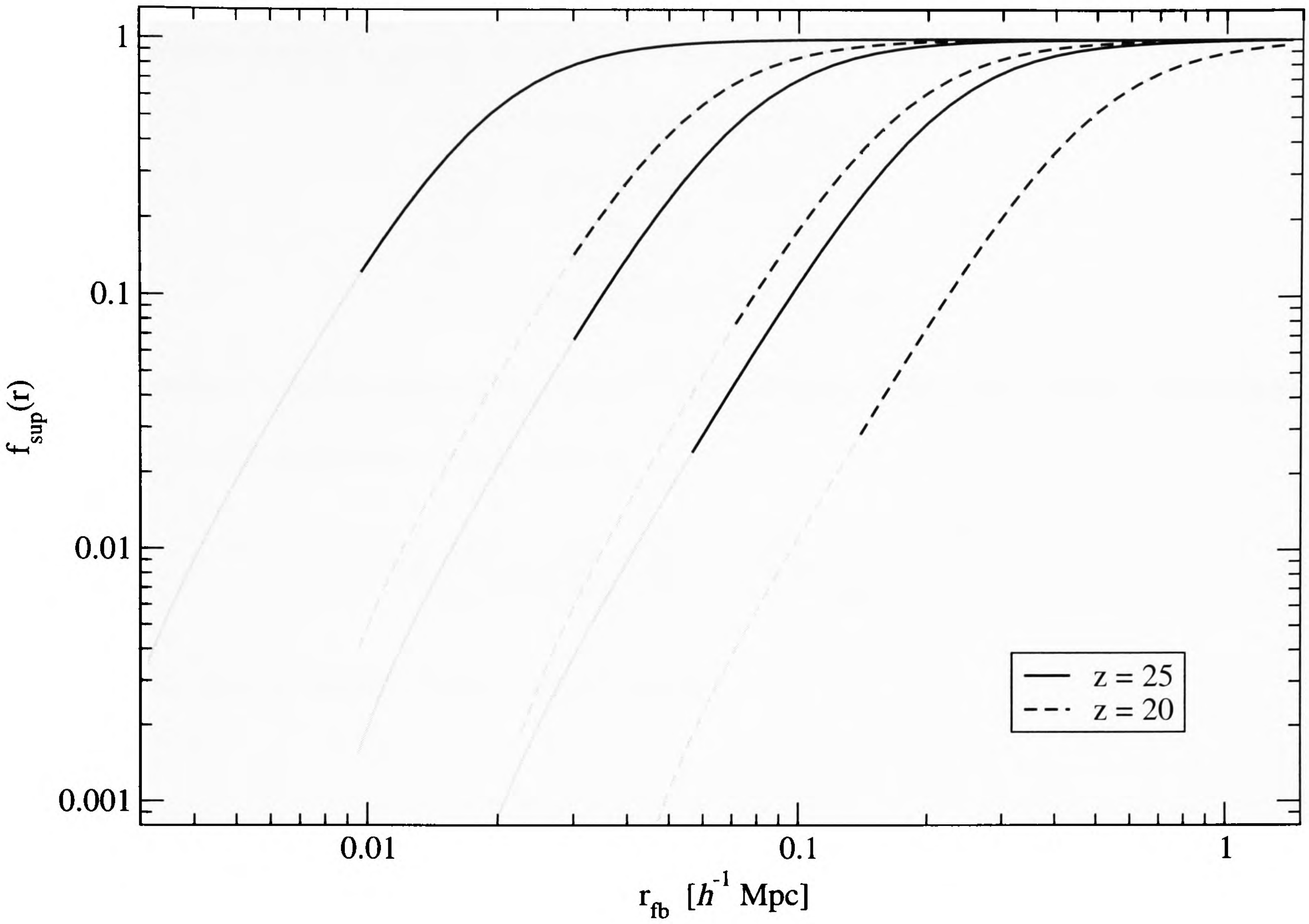


Figure 4.6: Average fraction of peaks that lie within a comoving distance r_{fp} of another one, shown for 2.5, 3, 3.5 σ peaks at redshifts $z = 25$ (solid, left to right) and $z = 20$ (dashed). Curves are plotted beginning at r_{fb} corresponding to the peak size r_{pk}. As in figure 4.5 the grey part denotes the region where the ξ_{pk} and thus f_{sup} becomes inaccurate.

This means that in a fraction of peaks

$$f_{sup}(r) \simeq \frac{< N_{pk}(r) >}{1+ < N_{pk}(r) >} \tag{4.16}$$

star formation is suppressed if on average the feedback from one star forming peak propagates out to some radius $r = r_{fb}$. This is shown in figure 4.6. For instance, in our fiducial case of 3σ peaks collapsing at $z = 25$, if one star forming peak emits radiation that ionises H_2 out to about $r_{fb} \sim 80\ h^{-1}$ kpc(about 8 times the peak size) star formation in every second peak will be suppressed.

For outflows from supernova explosions, r_{fb} is the radius of the expanding supernova shock front. The propagation of these fronts is a complex problem (Ostriker & McKee 1988). To estimate the effect on the IGM surrounding the MBC halos, we use a very simple analytical argument.

Table 4.1: Number density, $n_{pk}(\nu_{pk})$, (in h^3 Mpc^{-3}) of peaks *collapsing* at redshifts $z = 25$ and $= 20$.

	$z = 25$	$z = 20$
2.5σ	2.72×10^4	1.12×10^3
3.0σ	431	38.5
3.5σ	20.3	1.65

For an idealised outflow expanding into a homogeneous IGM, Voit (1996) determines the comoving radius of a supernova shock front as

$$r_{sn} = 280 \left(\frac{0.022}{\Omega_b\, h^2} \frac{E_0}{10^{56}\mathrm{erg}} \right)^{1/5} \left(\frac{10}{1+z_1} \frac{\hat{t}}{10^{10}\mathrm{yr}} \right)^{2/5} \;\mathrm{kpc} \tag{4.17}$$

where E_0 is the kinetic energy of the supernova and

$$\hat{t} = \int_{t_1}^{t} (1+z')^2 \frac{\mathrm{d}t}{\mathrm{d}z'}\, \mathrm{d}z' \tag{4.18}$$

accounts for the cosmological expansion while the shock front moves through the IGM and $t_1 < t$ denotes the time at which the explosion occurs. $\frac{\mathrm{d}t}{\mathrm{d}z}$ is the time - redshift relation as given in eq (2.23). Additional assumptions here are that as the shock front escapes from the host halo the energy loss is negligible as is the energy lost in ionisation. More detailed calculations show that eq. 4.17 somewhat overestimates the radius of the shock front (Barkana & Loeb 2001).

Using this approach we find that for MBC halos collapsing from 3σ peaks at $z = 24.6$ and forming massive stars that explode as hypernovae about 3×10^6 years later, the shock front will have travelled out to a comoving radius of $r_{fb} \approx 15$ kpc by redshift $z = 24$ for a hypernova energy of 10^{52} ergs. For energies of 10^{53} ergs, that radius grows to $r_{fb} \approx 23$ kpc. From the plot in figure 4.6 we estimate that this would lead to the suppression of massive star formation in about a third of all 3σ peaks at that redshift.

4.3.5 When does the epoch of first star formation end?

In principle the above analysis does not depend on the type of mechanism we are looking at as long as it has the potential to suppress star formation. The propagation of ionisation fronts is one such mechanism. The outflow of metal enriched gas from exploding stars is another one. Both impact on the formation of the first massive stars. If the first stars reionise large regions of the Universe, it means that molecular hydrogen also gets photo dissociated. As a result low mass halos with virial temperatures less than about 10^4 K are no longer able to cool efficiently which would prevent any more stars from forming at these early times.

The metal enrichment from exploding stars increases the cooling efficiency, which means that even if baryonic clouds have already formed, fragmentation into dense clumps can proceed more efficiently, too. This implies smaller fragments and thus less massive stars. Metal enrichment also means that the stars have significantly smaller temperatures than their metal-free counterparts.

Observations of the Lyα forest indicate that the IGM was metal enriched to a level of $10^{-3.5} Z_{\odot}$ by $z \sim 5$ (Songaila 2001). However, when exactly the transition occurs is not entirely clear. Above we used the supernovae constraint to arrive at a redshift below 16 - 27 respectively. Mackey, Bromm & Hernquist (2003) estimate a redshift of $z \sim 15 - 20$ also based on chemical enrichment from the first stars. In the following we assume that the epoch of first star formation will end at a redshift of about $z \sim 20$, coinciding with the redshift at which 2.5σ peaks collapse to form MBC halos.

The fact that the epoch of first star formation is not just confined to a specific redshift as assumed in section 4.1 but extends over a significant redshift range is not necessarily a problem. We do find that at $z \sim 25$ about two to three times as many MBC halos collapse than at $z \sim 29$. In light of the previous section this difference is likely to be smaller. Because high peaks tend to be highly clustered, MBC halos collapsing from slightly lower peaks at $z \sim 25$ will still be located very near those that collapse from higher peaks at $z \sim 29$. In this way a significant fraction of the much more numerous (c.f table 4.1) 3.0σ peaks at $z \sim 25$ may have star formation suppressed within them due to the feedback from stars in neighbouring clouds that collapsed from 3.5σ peaks earlier.

Even then the actual number of MBC halos will still be larger. Nevertheless, the total number of MBC halos forming is only underestimated by at most a factor 5 or so. Our results for the abundance of MBH remnants in present day galactic halos, presented in the next chapter, must therefore be adjusted accordingly.

4.3.6 End of stellar evolution and MBH formation

Because the stars are very massive it seems likely that in their final evolutionary stage they will turn into or leave behind massive black holes (MBHs). Exactly how massive these are will depend chiefly on the mass of the star from which they formed. Numerical simulations indicate (Heger et al. 2002) that population III stars in the mass range $30 - 140 M_\odot$ may experience a supernova but the fall-back will convert the remnant into a black hole . Between $30 - 140 M_\odot$the star gets completely disrupted in pair-unstable supernovae with no remnant left behind. Stars with mass above $260 M_\odot$will turn into massive black holes of essentially the same mass. The latter process is nucleosynthetically sterile as no processed material is ejected. Which of these are the dominant modes obviously depends on the primordial IMF.

For the modelling of MBH merging as described in the next chapter we have used seed MBH masses 260 $M_\odot$. This choice is motivated by the fact that these MBHs could originate from stars of essentially the same mass, i.e. the relation between initial stellar and final MBH mass is relatively tightly constrained by the simulations mentioned above. In contrast any BH remnants from stars in the $30 - 140 M_\odot$mass range may loose a large fraction of their mass in the supernova that precedes the formation of the BH.

4.4 Summary

In this chapter we have investigated how the first stars formed. On the basis of recent analytical work and numerical simulations we have used simple analytical arguments to examine when and where first star formation took place.

- We used a simple argument from the energy of supernova outflows to constrain the redshift range for star forming halos to $16 < z < 27$ corresponding to collapse from $2.6 - 3.3\sigma$ peaks.

- For the star forming clouds we have determined an upper limit on the volume within the cloud that a single population III star can ionise. In doing so we have assumed that gas is distributed at constant density throughout the cloud. For the more realistic distribution of dense lumps the spatial extent of the effect of radiative feedback from the stars is likely to be smaller.

- We have determined the correlation function of MBC halos and computed the fraction of MBC halos that lie within a given distance of other MBC halos.
- Using this we have estimated the fraction of MBC halos in which star formation could be suppressed due to radiative or supernova feedback from stars in nearby MBC halos. We find that the effect of supernovae in individual MBC halos can suppress massive star formation in up to a third of all MBC halos.

Chapter 5

Hierarchical merging of galaxies and MBHs

Once formed the MBH remnants of the first stars will subsequently become part of the hierarchical merging process involving their host halos. Since these MBHs are primarily tied to the potential well of their hosts, the large scale distribution of MBHs is expected to follow that of their hosts. Through merging, MBHs and the halos within which they originally reside become part (hereafter *satellites*) of larger halos, and their orbits inside these are mainly determined by the gravitational interaction between the satellite and the centre of the larger halo (hereafter *host*). The subsequent dynamical evolution of the satellite and its MBH is determined primarily by the tidal forces in the host's gravitational potential. Furthermore the satellite orbits are affected by dynamical friction, which depends on the relative velocities of the satellites and the density of matter in the host. Both forces act to drag the MBH and its associated satellite toward the centre of the host. In the process the satellite might get completely 'tidally stripped' such that very near the host centre only the MBHs and possibly a remnant core of the original satellite are left. Ultimately the MBHs could then merge with any other MBH that may already be present in the host central region.

In this chapter we introduce a semi-analytical approach to the modelling of the hierarchical merging process and subsequent dynamical evolution of halos and their associated central MBHs, starting with the MBC halos that contain the MBH remnants of the first stars. In section 5.1 we describe the semi-analytical scheme used to track the hierarchical merging and dynamical evolution of halos and MBHs between the redshift of first star formation and the present day. We also address in more detail two particular dynamical processes that have not been incorporated into the semi-analytical scheme. One is the continuous erosion of density profiles of satellites orbiting on fixed radii within a host halo, which is described in section 5.2. The second is the effect on the density distribution of a host halo due to the inspiral of a MBH to its centre. This is described in section 5.3. We summarise our results from this chapter in section 5.5.

5.1 Semi-analytical treatment of hierarchical merging and satellite dynamics

To model the merging of MBHs and their hosts we have used a semi-analytical code (Taylor 2001; Taylor & Babul 2001; Taylor & Navarro 2002) that combines a Monte-Carlo algorithm to generate halo merger trees with analytical descriptions for the main dynamical processes – dynamical friction, tidal stripping, and tidal heating – that determine the evolution of merged remnants within a galaxy halo.

5.1.1 Monte-Carlo merger histories

The code uses the algorithm of Somerville and Kolatt (1999, hereafter SK) to construct Monte-Carlo merger histories. In doing so we start with a present-day halo of a specific mass and use the algorithm to trace its merger history backwards to a redshift of 30. This allows us to identify all precursors that have merged to finally form the present day halo we were originally looking at. The SK algorithm itself is based on the extended Press-Schechter (EPS) formalism (Bond et al. 1991;Bower 1991;Lacey & Cole 1993) that allows the calculation of conditional probabilities, such as for instance the mass of a halo, given that it had some specified lower mass at a given point in time earlier. More specifically, the SK algorithm creates multiple random realisations of halo merger histories drawn from the conditional probability distribution function. This function had previously already been used to derive mean quantities such as the average mass increase and merger rate for halos.

Due to computational limitations merger trees can only be traced back to halos with a mass of at least $\simeq 3 \times 10^{-5}$ of the total mass. This sets the minimum *mass resolution* of the tree. It implies that only for present-day halos with a mass of $\lesssim 10^{11}$ $M_{\odot}$can we trace back the merger

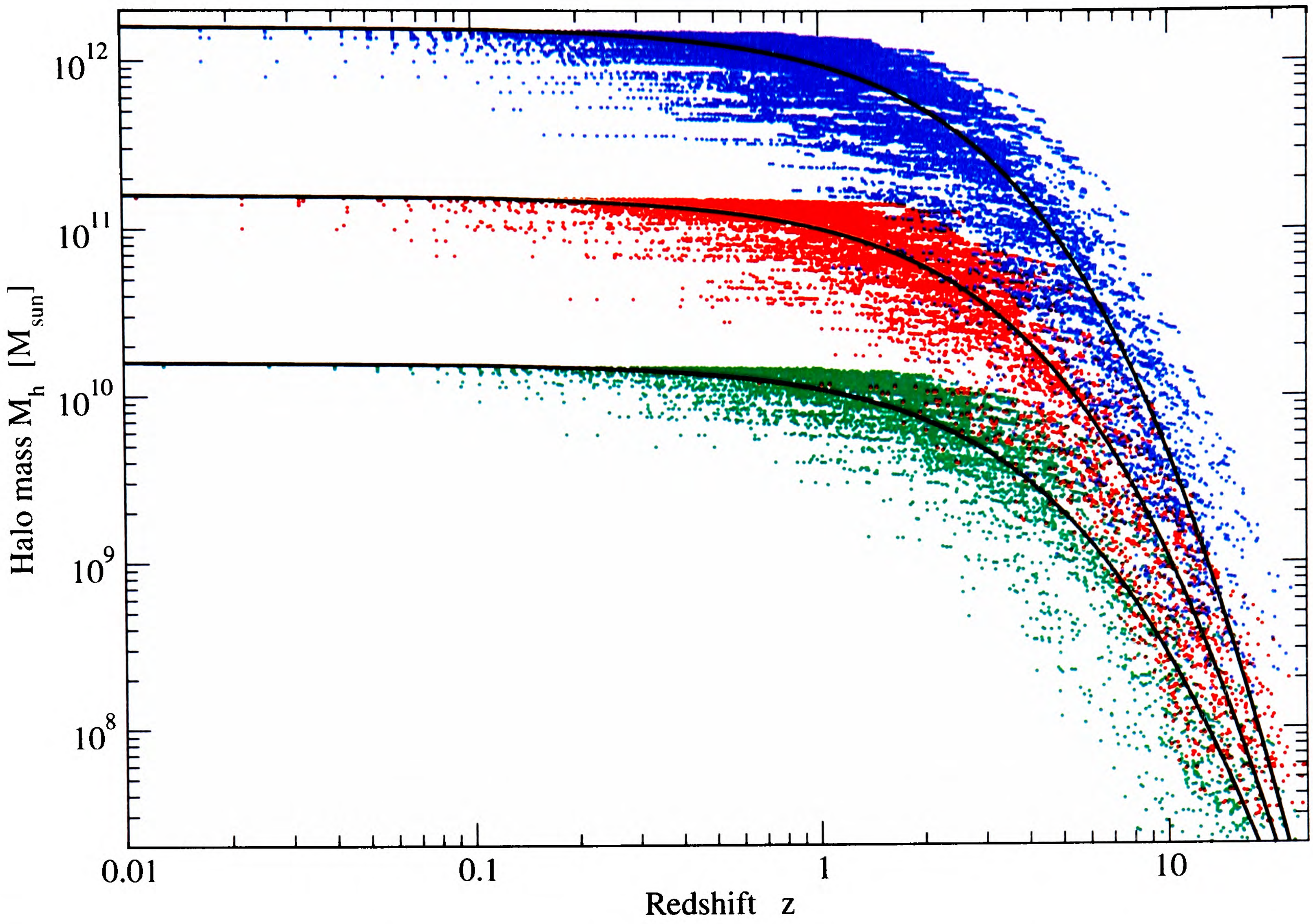


Figure 5.1: Redshift vs Mass distribution of progenitors of halos with final halos masses $1.6 \times (10^{10}, 10^{11}, 10^{12})$ $M_\odot$. The black solid line is a two parameter non-linear best fit depending on the redshift and the final halo mass (see text).

tree far enough to include MBC halo precursors.

For more massive halos the mass resolution is typically too high for MBC halos to be included. Many tree branchings will drop below the mass resolution limit at redshifts lower than 25 already, which means that for these halos we cannot always trace the merger history back to the formation of every remnant MBH. To overcome this problem, whenever a halo more massive than M_{MBC} appears in the merger tree after MBHs started forming at $z = 24$, we calculate the probability of the halo hosting a MBH remnant, based on the frequency of 2.5, 3 and 3.5 σ peaks. The drawback is that for present-day halos more massive than 10^{11} $M_\odot$, systems entering the tree at the mass resolution limit may contain more than one MBH. In this case we are assuming the MBHs have merged to form a single but correspondingly more massive MBH, since we cannot account for the dynamics of MBHs in halos with mass between M_{MBC} and the mass resolution limit (see also section 5.1.2)

In figure 5.1 we show the redshift vs mass distribution of progenitor halos for the $1.6 \times (10^{10}, 10^{11}, 10^{12})$ M$_{\odot}$final halos. The black solid lines represent a non-linear best fit of the form

$$\ln M_h(z) = 12 + \tilde{m} - 0.04\ \tilde{m}(z - 0.01z^2) \tag{5.1}$$

where $\tilde{m} = 11.5 + \ln\left(\frac{M_{h,0}}{1.6\times10^{10}M_{\odot}}\right)$ and $M_{h,0}$is the final halo mass.

5.1.2 Dynamical evolution of satellites and MBHs

Whenever the merger tree indicates that a merger has occurred the code follows the dynamical evolution of MBHs and their associated halos inside the new merged halo. Halos are modelled by a smooth density profile of the form $\rho(r) \propto r^{-1.5}(r_s^{-1.5} + r^{-1.5})$ (Moore et al. 1999), except for the central regions ($r < 0.1r_s$), where we assume the profile has a constant density core. The halos grow in mass corresponding to their merger history, and change in concentration according to the scaling relations given by Eke, Navarro & Steinmetz (2001).

Every halo contains a central galaxy which is modelled schematically. We assume that a third of the gas within the halo cools on a dynamical timescale to form a galactic disk. Every time the halo undergoes a major merger the disk is disrupted and transformed into a spheroid with some overall efficiency. We choose as the disruption criterion that the disk collide with an infalling satellite of mass equal to or greater than its own, and set the efficiency with which disk material is then transferred to the spheroid to 0.25. This choice of parameters is required to limit the formation of spheroids and thus produce a reasonable range of morphologies in isolated present-day $10^{12}M_{\odot}$ systems, as discussed in Taylor & Babul (2002). We do not expect the results for halo black holes to depend strongly on these parameters, although they may have some effect on the properties of the central black holes. Finally, the evolution of halos in side branches of the merger tree is followed more approximately, by assuming that higher-order substructure (that is sub-halos within sub-halos) merges over a few dynamical times, causing its black hole component to merge as well, while unmerged substructure percolates down to a lower level in the tree. This is necessary for the overall calculation to remain computationally viable. Since the number of side branches increases exponentially as we move up the tree (i.e. back in time), the computational effort would inflate accordingly, were we to follow the dynamical evolution of substructure in side branches in similar detail as that in the main branch. Details of this model are discussed in Taylor & Babul (2003); here it serves only as a backdrop for the dynamical calculations of black hole evolution.

The vast majority of halo 'mergers' really are 'accretion' events, meaning that a satellite halo is being accreted by another and typically much larger host halo. In these cases the host halo is assumed to remain in dynamical equilibrium. The effect of dynamical friction, tidal heating and tidal stripping on the orbit of an MBH and its associated satellite can then be estimated analytically, since the background potential provided by the host remains well-defined.

Dynamical friction

The dynamical friction force acting on the satellite system in the host can be approximated by the Chandrasekhar formula (Binney & Tremaine 1987)

$$\frac{d\mathbf{v}_{sat}}{dt} = -16\pi^2 \ln\Lambda_c G^2 m(M_{sat} + m) \frac{\int_0^{v_{sat}} f(v')v'^2 \, dv}{v_{sat}^3} \mathbf{v}_{sat} \tag{5.2}$$

where $\mathbf{v}_{sat}$ is the satellite velocity with respect to the host centre, M_{sat} the satellite mass, m the mass of individual host halo particles that have an overall isotropic phase space number density $f(v)$. $\ln \Lambda_c = \ln(b_{max}/b_{min})$ is the Coulomb logarithm, that is the ratio of maximum and minimum impact parameters, b_{max} and b_{min}, of the host particles with respect to the satellite system.

If the host particles obey a Maxwellian distribution with dispersion σ_h

$$f(v) = \frac{n_0}{(2\pi\sigma_h^2)^{3/2}} \exp(-\frac{1}{2}v^2/\sigma_h^2) \tag{5.3}$$

where n_0 is the number density of particles in the host. Substituting this into equation 5.2 and using $\rho_{host} = mn_0$ we then get

$$F_{df} = -\frac{4\pi \ln\Lambda_c \, G^2 \rho M_{sat}^2}{v_{sat}^2} \left[\mathrm{erf}(X_{df}) - \frac{2X_{df}}{\sqrt{\pi}} e^{-X_{df}^2} \right] \tag{5.4}$$

where $X_{df} = v_{sat}/(\sqrt{2}\sigma_h)$.

In our context the minimum and maximum impact parameters and thus $\ln \Lambda_c$ are not strictly defined, but typically b_{max} can be taken to be equal to some characteristic or half-mass radius of the host system, while b_{min} is of the order of the 'radius of influence' of the satellite, $b_{min} \sim GM_{sat}/v_{sat}^2$ (c.f. chapter 3 for the MBH range of influence).

Equation 5.2 gives the dynamical friction force for a point particle moving through an infinite uniform matter distribution of density ρ. In practice this also works well in our case as long as the satellite system's mass and size are small compared to that of the host. In particular, the satellite system should not be more massive than about 20 percent of the host (Binney & Tremaine 1987) and should orbit neither outside the host nor within its core.

A more general form of the dynamical friction time that incorporates non-circular orbits was formulated by Colpi *et al.* (1999) (Colpi, Mayer & Governato 1999; Taylor & Babul 2003)

$$\tau_{df} = 1.2\ e \frac{(M_{host}/M_{sat})}{\ln(M_{host}/M_{sat})} \epsilon^{0.4} \frac{P_{vir}}{2\ \pi} \tag{5.5}$$

where M_{host} and M_{sat} denote the mass of the host and satellite halos respectively and P_{vir} is the circular orbital period at the virial radius of the host. e is a numerical constant and of order $e \sim 2$ in our case (Taylor & Babul 2003). The circularity parameter $\epsilon = J(E)/J_{circ}(E)$ is the ratio of the angular momentum of the actual satellite orbit and that of a circular orbit with the same energy.

Tidal heating

The finite size of the satellite system does matter in other respects though. The tidal force due to the host potential gradient across the satellite system is responsible for both tidal heating and tidal stripping of the satellite system, depending on whether the tidal force across the satellite varies rapidly or remains essentially static.

A 'rapidly varying' tidal force is defined as an external gravitational shock acting on a satellite system on a timescale t_{shock} that is smaller than the internal orbital time scale $t_{orb,sat}$ of the satellite, the latter being essentially a measure of the satellite crossing time. In our model this can happen in two ways. Satellites on elongated orbits will experience close encounters with the host centre which are typically brief enough to satisfy the above tidal heating criterion. Alternatively as satellites cross the galactic disk a similar gravitational shock is induced.

The shock 'heats' the satellite by an amount ΔE that depends on the velocity change $\Delta V_{int,sat}$ induced in the satellite's internal particle distribution. In particular, during the shock the heat transfered to a satellite mass element at distance r from the satellite centre is (Taylor & Babul 2001)

$$\Delta E(r) = \frac{1}{2}\Delta V^2_{int,sat}(r) = \frac{1}{2}\int_0^t \mathbf{A}_{tid}(r,t')\ \mathrm{d}t' \cdot \int_0^t \mathbf{A}_{tid}(r,t'')\ \mathrm{d}t'' \tag{5.6}$$

where $\mathbf{A_{tid}(r,t)}$ is the tidal acceleration induced by the shock. Assuming the satellite remains virialised and no shell crossings occur in the satellite mass distribution, one can then deduce the resulting change in the satellite's mean density inside some radius r.

$$\Delta\bar{\rho}_r \propto \Delta\left(\frac{M_{sat}(<r)}{r^3}\right) \propto -\frac{\Delta r}{r^4} \propto -\frac{\Delta E(r)}{r^2} \tag{5.7}$$

Here we have used the fact that the total energy of a mass element $E(r) \propto -1/r$, which means that heating by an amount $\Delta E(r)$ leads to a change in radius of $\Delta r \propto \Delta E(r) r^2$

Tidal stripping

In contrast a static or slowly varying tidal force leads to a tidal radius r_t in the satellite where all radial forces due to the host and satellite potentials combined with any centripetal forces cancel out. For instance a satellite on a circular orbit at radius R from the centre and with angular velocity ω, the tidal force due to the host gravitational potential Φ is static and we get for the tidal radius (King 1962)

$$r_\mathrm{t} \simeq \left(\frac{GM_{sat}}{\omega^2 - \mathrm{d}^2\Phi/\,\mathrm{d}R^2}\right)^{1/3} \tag{5.8}$$

This is only approximate for two reasons. Firstly, it only applies to particles on the line connecting the host and satellite centres, but as such imposes a lower limiting radius from the satellite centre above which particles may get stripped away. Secondly, it is only valid for the case that the satellite mass is much less than the host mass and r_t is much smaller than the satellite orbital radius, since we just consider the tidal force due to the first order (i.e linear) variation of the host potential across the extent of the satellite. Another definition for a tidal radius will be discussed in the next section.

For the general satellite orbits equation 5.8 can still be applied, although we use it only to compute instantaneous tidal radii for a number, n_{orb}, of discrete orbital sections each corresponding to a fixed time step Δt, but then only remove a fraction $\Delta t/t_{orb}$ of the mass outside $r_{\rm t}$ at that instant. For a circular orbit this means that all the mass outside $r_{\rm t}$ is lost over exactly one orbit.

The tidal stripping condition equation 5.8 can also be re-expressed as the radius within which the satellite mean density $\bar{\rho}_{sat}$ is larger than the mean density, $\bar{\rho}_{gal}$ of the halo/galaxy within its orbital radius,R, by a factor κ

$$\bar{\rho}_{sat}(< r_{\rm t}) = \kappa\bar{\rho}_{host}(< R) \tag{5.9}$$

and κ is given by

$$\kappa = \frac{\omega^2}{\omega_c^2} - \frac{1}{\omega_c^2}\frac{{\rm d}^2\Phi}{{\rm d}R^2} \tag{5.10}$$

where ω_c is the angular velocity for a circular orbit at radius R from the host centre.

The main overall contribution to dynamical friction comes from the halo, with corrections accounting for the role of the disk. Tidal heating depends mainly on the disk and bulge components, while tidal stripping is affected by the overall potential, which at various radii from the centre is dominated by one or the other of the three components.

5.1.3 Inclusion of MBHs

In most ways the semi-analytic code does not need to be modified beyond the addition of the central MBH to the respective satellites with a corresponding increase in mass. The dynamical properties of the satellite in the host halo will remain essentially unaffected by the presence of the MBHs. Table 5.1 lists the combinations of seed MBH masses and the peak heights we have investigated. The latter affects primarily the fraction of halos that contain seed MBHs at the lower mass resolution limit of our simulations.

'Naked' MBHs

As the satellite is tidally stripped to the point that it gets disrupted, the MBH will remain and come to dominate the dynamical interaction with the host system. We will also refer to these MBH left-overs from disrupted satellites as 'naked MBHs'. While satellites do not need to be

Table 5.1: Masses of seed MBHs and heights of peaks in initial density field within which they formed.

	$M_{\bullet,seed}$	peak height ν_{pk}
A	260 $M_\odot$	3.0
B	1300 $M_\odot$	3.0
C	260 $M_\odot$	2.5
D	260 $M_\odot$	3.5

tracked by the code once they are disrupted, naked MBHs require continued updating of their orbits. This is particularly important if we want to establish the abundance and observational signatures of naked MBHs – whose associated satellites may have been disrupted long ago – in present day galactic halos.

The code was therefore modified to enable it to follow the dynamical evolution of these naked MBHs in the same way as for ordinary satellite systems, with the exception of tidal effects which obviously do not apply to what are essentially point-like particles. Naked MBHs are thus only subject to dynamical friction.

5.2 Tidal stripping and erosion of satellite density profiles

In addition to tidal stripping and heating there is yet another although secondary effect, which is the (adiabatic) back reaction of the particle distribution inside the tidal radius following the satellite tidal truncation.

If a satellite gets truncated from its initial virial radius r_{vir} to some tidal radius $r_{\rm t} < r_{vir}$ the normalisation of the satellite potential inside $r_{\rm t}$ changes by an amount

$$\begin{aligned}\Delta\phi_{sat} = \phi_{vir} - \phi_{\rm t} &= -4\pi G\left[\frac{1}{r_{vir}}\int_0^{r_{vir}} \rho_{sat}(r')r'^2\ {\rm d}r' - \frac{1}{r_{\rm t}}\int_0^{r_{\rm t}} \rho_{sat}(r')r'^2\ {\rm d}r'\right] \\ &= -G\left[\frac{M(< r_{vir})}{r_{vir}} - \frac{M(< r_{\rm t})}{r_{\rm t}}\right] \qquad (5.11)\end{aligned}$$

As a result some fraction of the previously most loosely bound particles inside $r_{\rm t}$ will now become unbound, too, as their total energy becomes positive. This effect has not been included in our semi-analytic treatment of satellite dynamics. However, most of the satellite mass is concentrated in its centre and we expect this effect to be small there. Here we quantify this and in particular look at how it affects the evolution of the satellite density profile interior to r_{t}. To do this we iteratively apply the Jeans equation to the particle distribution function of the satellite.

5.2.1 Assumptions and initial conditions

To avoid any contribution from tidal heating due to rapidly varying tidal forces we assume a satellite halo moving on circular orbits inside the static smooth potential of a larger host halo. For both the satellite and the host system, the matter distributions are spherically symmetric. The radial density profile of the satellites is initially of a form as described by Navarro, Frenk and White (NFW) (Navarro, Frenk & White 1997)

$$\rho(x) = \frac{\rho_0 r_s^3}{x(1+x)^2} \tag{5.12}$$

where ρ_0 is a characteristic density, r_s is the scale radius and $x \equiv r/r_s$ is the dimensionless radius. The host halo is also described by a NFW profile throughout, as it is considered to remain unaffected by the satellite eroding and spiralling towards the host centre. Although we have chosen a NFW profile to model the satellite and host here, we do not expect our results to be markedly different for other density profiles.

The satellite mass profile is given by

$$M_{sat}(< x) = 4\pi r_s^3 \int_0^x \rho(\tilde{x})\tilde{x}^2 \, \mathrm{d}\tilde{x} \tag{5.13}$$

For the case that the satellite density distribution is truncated at some radius r_{t}, the gravitational potential is then

$$\phi(x) = \phi(x_{\mathrm{t}}) - \frac{G}{r_s} \int_x^{x_{\mathrm{t}}} \frac{M_{sat}(< \tilde{x})}{\tilde{x}^2} \, \mathrm{d}\tilde{x} \tag{5.14}$$

where G is the gravitational constant and $\phi(x_{\mathrm{t}}) = -GM_{sat}(< x_{\mathrm{t}})/(r_s x_{\mathrm{t}})$. The extension of equations (5.13) and (5.14) to the corresponding forms for the host halo is straightforward.

We also assume that the satellite halo particles are initially in virial equilibrium and have an isotropic velocity distribution with spread $\sigma(r)^2 \propto U(r)$, where U(r) is the potential energy at a given radius r from the centre of the satellite. Despite the halo dark matter particles being non or weakly interacting, the assumption of virial equilibrium is still accurate, particularly in the more central parts of the halo, where violent relaxation could most efficiently scramble up individual particle energies in the early stages of collapse so as to achieve equilibrium.

So far no explicit timescale has entered the problem, but in the following we identify the individual steps in the iterative sequence to recompute tidal radii, velocity dispersion etc. of the satellite, with individual satellite dynamical timescales.

5.2.2 Satellite properties and their evolution

Computation of profiles

The properties of the satellite we wish to determine are primarily its density and mass profiles $\rho_{sat}(r)$ and $M_{sat}(r)$ which will be computed mainly from two other components, the velocity spread $\sigma(r)^2$ and the tidal radius $r_{\rm t}$. Applying the Jeans equation (Binney & Tremaine 1987), the profile of the radial velocity dispersion $\sigma(r)$ can be obtained by solving

$$\frac{\mathrm{d}\sigma_r^2}{\mathrm{d}x} = -\left[\frac{GM_{sat}(<x)}{r_s x^2} + \sigma_r^2 \left[\frac{\mathrm{d}}{\mathrm{d}x}\ \ln\rho(x) + \frac{2}{x}\beta_\sigma(x)\right]\right] \tag{5.15}$$

where $\beta_\sigma = 1 - \sigma_t^2/(2\sigma_r^2)$ is the dimensionless anisotropy parameter with σ_r and σ_t being the radial and tangential velocity dispersions respectively. For the case of isotropic velocities $\beta_\sigma = 0$ and $\sigma^2(r) = 3\sigma_r^2(r)$.

The dimensionless tidal radius $x_{\rm t}$ is defined as the smaller of two radii. One is the radius within the satellite where the force on a particle due to the satellite potential is equal to that induced by the host tidal field. This is given by

$$x_{\rm t} = X\frac{R_s}{r_s}\left[\frac{M_{host}(<X)}{M_{sat}(<x_{\rm t})}\left(2.0 - \frac{\mathrm{d}\ \ln M_{host}(<X)}{\mathrm{d}\ \ln X}\right)\right]^{-1/3} \tag{5.16}$$

where X is the satellite distance from the host in terms of the host scale radius R_s and $M_{host}(<X)$ is the host mass profile. We will subsequently refer to $x_{\rm t}$ defined in this way as the *equal force* condition. Alternatively the *resonance condition* (Weinberg 1997 and references therein) yields a radius for which the satellite angular velocity matches that of the host.

$$x_{\rm t} = X\frac{R_s}{r_s}\left[\frac{M_{host}(<X)}{M_{sat}(<x_{\rm t})}\right]^{-1/3} \tag{5.17}$$

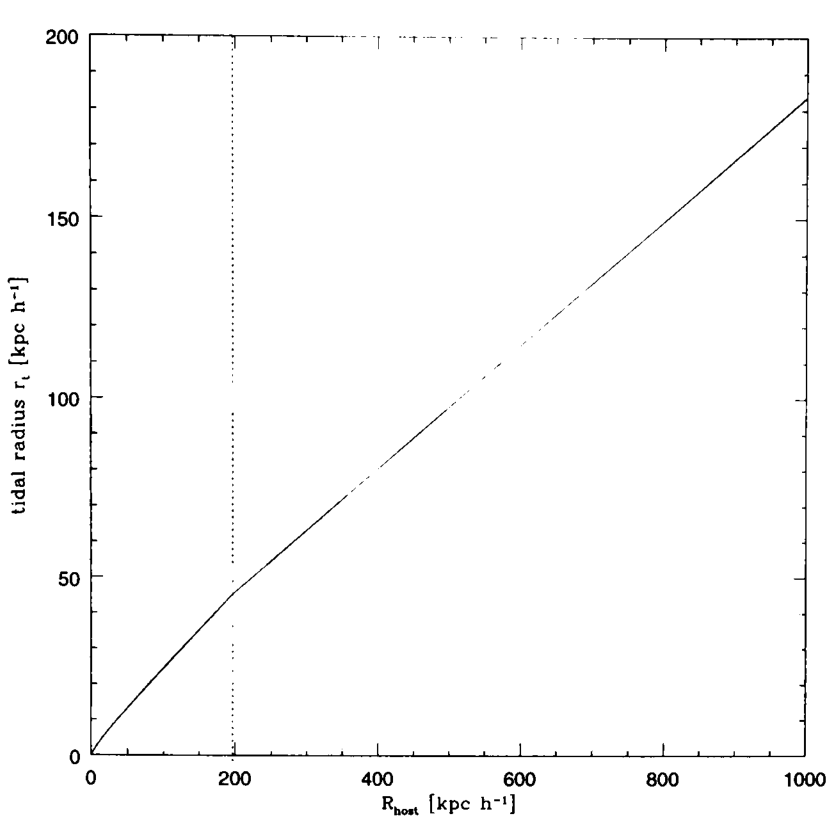


Figure 5.2: Tidal radius r_t of a $10^{12}M_\odot h^{-1}$ satellite (r_{vir} = 163 kpch^{-1}, r_s = 13.4 kpch^{-1}), scaling with distance R from the centre of a $10^{14}M_\odot h^{-1}$ host halo (R_{vir} = 757 kpch^{-1}, r_s = 91 kpch^{-1}. The dashed line indicates the satellite tidal radius below which it is determined by the resonance condition

In this case a satellite particle on an average circular orbit at this radius gains enough kinetic energy to leave the satellite, even if, from the equal force condition, it should remain with the satellite. An example of the scaling between host distance and tidal radius for both satellite and host being of NFW form, is shown in figure 5.2.

The tidal radius encapsulates most directly the effect of the host tidal field on the satellite, with the other quantities mentioned above only changing in response to the truncation of the satellite at increasingly smaller tidal radii. This is reflected in the iterative scheme we use to track the evolution of satellite properties: A new tidal radius $r_{n+1,\mathrm{t}}$ will be computed using the current density profiles and velocity spreads of the satellite $\rho_n(r)$ and $\sigma_n(r)$, where n denotes iteration order.

Once $r_{n+1,\mathrm{t}}$ has been determined we truncate the density profile of the satellite at that point, i.e. we consider all particles at radii larger than $r_{n+1,\mathrm{t}}$ as stripped away by the tidal field within one orbital timescale. In reality there will always be some bound orbits outside as well as unbound ones within r_t at any one time, however, averaged over a whole orbit we take these contributions as roughly cancelling each other and thus instant truncation at $r_{n+1,\mathrm{t}}$ is probably a reasonable assumption, especially in the case of circular satellite orbits as considered here.

Stripping of the satellite at $r_{n+1,\mathrm{t}}$ also implies that the satellite potential $\phi(r)$ within $r_{n+1,\mathrm{t}}$ has become shallower, precisely by an amount equal to that contributed previously by those particles outside $r_{n+1,\mathrm{t}}$ that have now been stripped away. As a result, a fraction of the most loosely bound particles [1]inside $r_{n+1,\mathrm{t}}$ will now escape the satellite, too. If the initial distribution function is $f(v)$, the fraction η_{esc} of particles in the high velocity tail above the escape velocity is then

$$\eta_{esc}(x) = \frac{\int_{v_{esc}}^{v_{max}} \mathrm{d}v v^2 f(v)}{\int_0^{v_{max}} \mathrm{d}v v^2 f(v)} \tag{5.18}$$

For instance, for a Maxwellian velocity distribution

$$f(v) = C_m \exp\left(-\frac{|v|^2}{\tilde{\sigma}^2}\right) \tag{5.19}$$

where $\tilde{\sigma} = \sqrt{2\sigma_r^2}$, C_m is a normalisation constant and $k = 3 - 2\beta_\sigma$. Equation 5.18 then gives

$$\eta_{esc}(x) = 1 - \frac{1}{\mathrm{erf}(\frac{v_{max}}{\tilde{\sigma}})}\left[\mathrm{erf}\left(\frac{v_{esc}}{\tilde{\sigma}}\right) + \frac{2}{\sqrt{\pi}}\sqrt{\frac{1}{\tilde{\sigma}}}\left[\exp\left(-\frac{v_{max}^2}{\tilde{\sigma}^2}\right)v_{max} - \exp\left(-\frac{v_{esc}^2}{\tilde{\sigma}^2}\right)v_{esc}\right]\right] \tag{5.20}$$

Since the halo dark matter is weakly or non-interacting, there are no two body interactions that could act to re-thermalise the distribution once it is truncated. In our iterative scheme this means that v_{max} is just equal to v_{esc} of the preceding iteration, i.e. the point where the distribution got truncated previously

$$\eta_{esc}^{[n+1]}(x) = \frac{\int_{v_{esc}^{[n+1]}}^{v_{esc}^{[n]}} \mathrm{d}v v^2 f^{[n]}(v)}{\int_0^{v_{esc}^{[n]}} \mathrm{d}v v^2 f^{[n]}(v)} \tag{5.21}$$

The form of the radial profile of $v_{esc}(r)$ can be obtained by integrating the equal force condition equation (5.16), which gives

$$\phi(x) - x\frac{r_s}{R_s}\frac{\mathrm{d}\Phi}{\mathrm{d}X} - \Phi(X) = C \tag{5.22}$$

C is the constant of integration and must be zero if we require that the left hand side vanishes at the tidal radius x_{t}. In this case the LHS defines the escape velocity for particles inside the

---

[1] Here and subsequently, 'loosely bound' refers to particles within the tidal radius that have a radial component to their orbits such that they may be carried beyond the tidal radius, where they would be stripped. Such particles must therefore have a kinetic energy $\geq \sqrt{GM(< r)/r}$. This issue will be addressed in the last part of this section.

satellite, v_{esc}

$$v_{esc}(x) = \begin{cases} \sqrt{2\left|\phi(x) - x\frac{r_s}{R_s}\frac{\mathrm{d}\Phi}{\mathrm{d}X} - \Phi(X)\right|} & \text{if } 0 < x < x_{\mathrm{t}}, \\ 0 & \text{if } x \geq x_{\mathrm{t}} \end{cases} \tag{5.23}$$

The $\Phi(X)$ term adds the relative contribution of the host (not host tidal) potential well to v_{esc} at the position of the satellite. This is important as the overall potential from both host and satellite is initially normalised with respect to the centre of the host, whereas we require v_{esc} to be defined with reference to the satellite centre.

Adiabatic expansion

The fact that on tidal truncation particles evaporate from within the tidal radius means that the satellite becomes less tightly bound. In contrast to the evaporation which explicitly requires radial orbits, the subsequently lowered mass inside the tidal radius means that particles on all orbits at larger radii will now become less tightly bound.

In response to this the satellite will expand, thus loosing more particles beyond r_{t} and a re-calculation of r_{t} for the now even more loosely bound satellite will yield an even smaller value of r_{t} and so forth. It appears that once tidal stripping of a fraction of particles within r_{t} is taken into account, the satellite will not remain stable any more, possibly even if it stays on an orbit of constant radius around the host with an associated constant tidal force [2]. However, if the mass lost per iterative step is small compared to the total remaining mass of the satellite, the relative reduction in r_{t} as well as the overall expansion of the satellite over one orbital timescale will be small, too, and thus we may treat the problem as an adiabatic process. Although mass loss and adiabatic expansion will occur simultaneously, in the numerical scheme we implement these in two consecutive steps: The satellite properties are being updated first to account for mass loss, and then recomputed again this time purely accounting for the adiabatic expansion.

Adiabatic invariance of the tangential action implies angular momentum conservation for individual particles, i.e.

---

[2] I.e. if other processes like dynamical friction that act to erode the satellite orbit are ignored.

$$\frac{\mathrm{d}\mathbf{L}_i}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r}_i \times \mathbf{v}_i) = 0 \tag{5.24}$$

Assuming the velocity distribution remains isotropic, we can derive the rms value of the angular momentum

$$\begin{aligned} < L^2(r) > &= \int_0^{v_{esc}} \mathrm{d}v \int_0^{\pi} \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi \, r^2 \, v^2 \, \sin^2\theta f(v) \, v^2 \, \sin\theta \\ &= r^2 \int_0^{\pi} \sin^3\theta \, \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi \int_0^{v_{esc}} v^4 \, f(v) \, \mathrm{d}v \\ &= \frac{8}{3}\pi \, r^2 \int_0^{v_{esc}} v^4 \, f(v) \, \mathrm{d}v = \mathit{constant} \end{aligned} \tag{5.25}$$

where we have used $|L| = r \; v \; \sin \; \theta$. This implies

$$\begin{aligned} < L^2(r) > &\propto r^2\sigma^2(r) \\ &\propto r^2 \; K(r) = \mathit{constant} \end{aligned} \tag{5.26}$$

Since $K(r) \propto \overline{v^2}$ the velocity spread scales inverse proportionally to the adiabatic expansion in radius. However, it is the decrease in mass interior to a given radius in the satellite, $\Delta M(< r)$, that is responsible for the adiabatic expansion of the mass distribution, which can be expressed as a stretching of the radial scale $\Delta r(r)$. Here we chose $\Delta r(r)$ such that

$$\frac{\Delta r}{r} = - \left. \frac{\Delta M_{sat}}{M_{sat}} \right|_r \tag{5.27}$$

As the mass contained in any spherical shell must be conserved, the corresponding stretching of the density profile is

$$\rho_f(r_f) = \left(\frac{r_i}{r_f}\right)^2 \frac{\mathrm{d}r_i}{\mathrm{d}r_f} \rho_i(r_i) \tag{5.28}$$

where $r_f = r_i + \Delta r_i$ and subscripts ‘i’ and ‘f’ denote before and after mass loss..

Circular vs radial particle orbits

It is important to note that both, continued tidal stripping and subsequent adiabatic expansion as considered above rely on the presence of radial or, more generally, non-circular orbits of particles inside the satellite. Particles on circular orbits only respond to the mass distribution inside their orbital radius, which is fixed. In a satellite consisting entirely of particles on circular orbits, particles then remain completely undisturbed by the stripping of material at larger radii. In other words, stripping of a satellite composed of particles on circular orbits would only depend on the external tidal force field, but no back reaction would be provoked in the particle distribution inside the tidal radius. In contrast for radial orbits, particles originally at a radius smaller than the tidal radius, may travel out beyond the tidal radius, where they would be stripped away. This means that the mass distribution inside the tidal radius could change over a dynamical time scale. The extent of this effect depends on how strongly non-circular orbits contribute as well as their eccentricity. It could be argued that, since only particles on non-circular orbits can travel from within to outside the tidal radius, continued tidal stripping leads to the depletion of radial orbits. However, this is not obvious. The changing of the mass distribution within the tidal radius on a time scale of the order of the dynamical time, means that particles previously on circular orbits may attain radial components, too.

5.2.3 Satellite moving on circular orbits at fixed host distance

We have applied the iterative scheme above to a satellite system moving in a host system on exactly circular orbits of constant radii. In this way we have ensured that the satellite is moving in a constant tidal field and we have also ignored dynamical friction which would otherwise result in a decrease in distance from the host centre. Any subsequent change in satellite tidal radius and mass loss is thus presumably caused by the repeated adiabatic expansion and tidal truncation. Because the expansion leads to the satellite being more loosely bound, the tidal radius (in the same static tidal field) will actually become smaller, implying that a run-away mass loss could occur, although this is not obvious *a priori*.

The system we are considering is a $10^{12} M_{\odot} h^{-1}$ NFW satellite halo moving on a range of circular orbits at fixed host distance within a NFW host halo of mass $10^{14} M_{\odot} h^{-1}$. The specific

Table 5.2: Parameters of the host and satellite halos.

	Host	Satellite
Mass $[M_{\odot} h^{-1}]$	10^{14}	10^{12}
R_{vir} [kpc h^{-1}]	757	163
R_s [kpc h^{-1}]	91	13.4

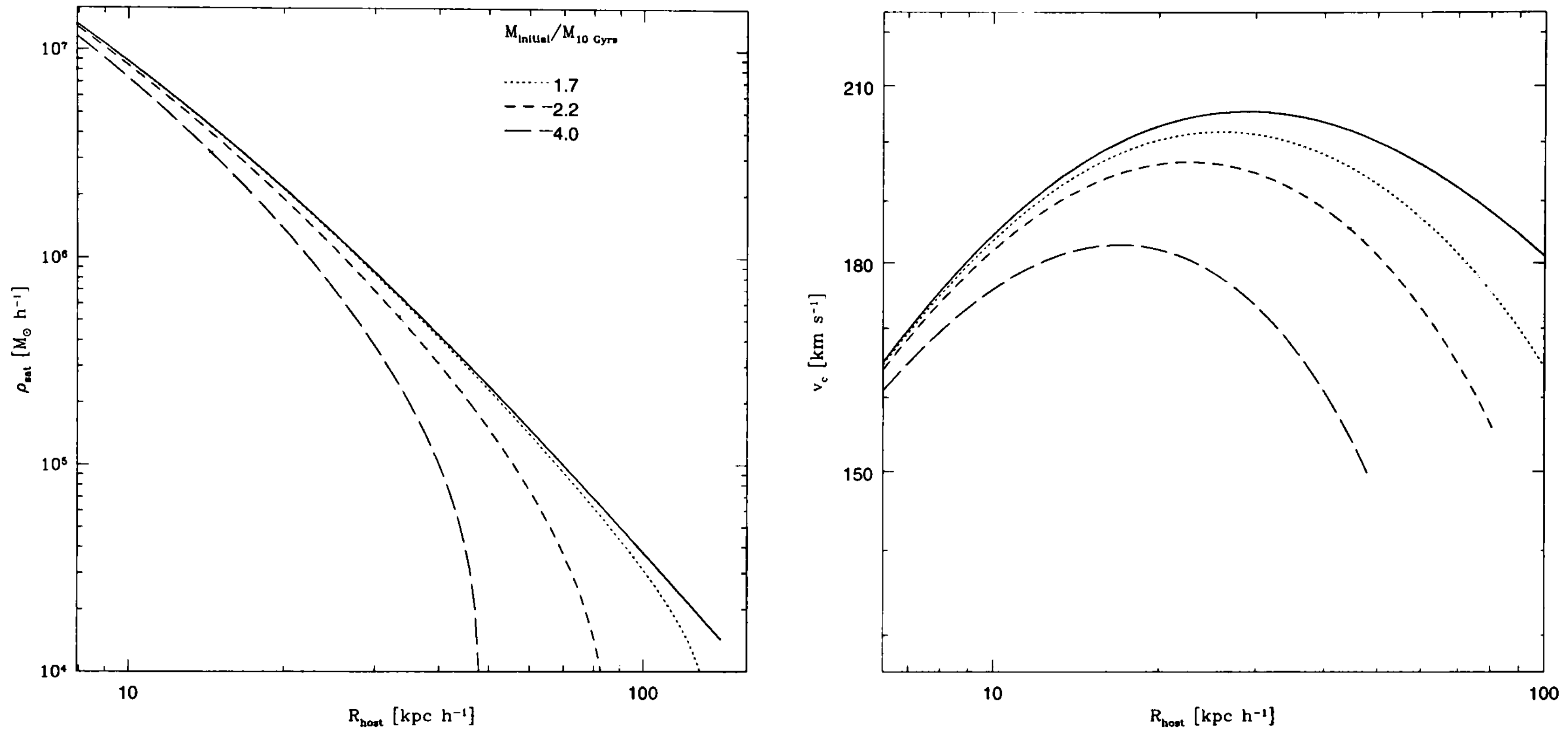


Figure 5.3: Profiles of satellite density and circular velocity v_c, initially (solid line) and after approximately 10 Gyrs for the satellite at 700, 500 and 300 kpc h^{-1} from the host centre (shown by the dotted, dashed and long dashed lines respectively).

properties of the two halos are given in table 5.2.

In the top panels of figure 5.4 we plot the density and circular velocity profiles of the satellite for the original satellite and after 10 Gyrs, orbiting at host distances of 700, 500 and 300 kpc h^{-1} respectively. On these orbits 10 Gyrs corresponds to 3.7, 2.3 and 1.2 orbital periods respectively. The bottom panels in figure 5.4 show the evolution with time of the tidal radius, as well as the satellite mass as a fraction of the initial mass. All of these tend to definite values at late times, implying that a run-away mass loss does not occur at constant distance from the host halo.

We can now compare this with the results of N-body simulations (Klypin et al. 1999; Moore et al. 2001). In particular, Klypin et al. (1999) simulate a satellite - host system similar to the one on which we have based our particular analysis. Although qualitatively correct the simulated

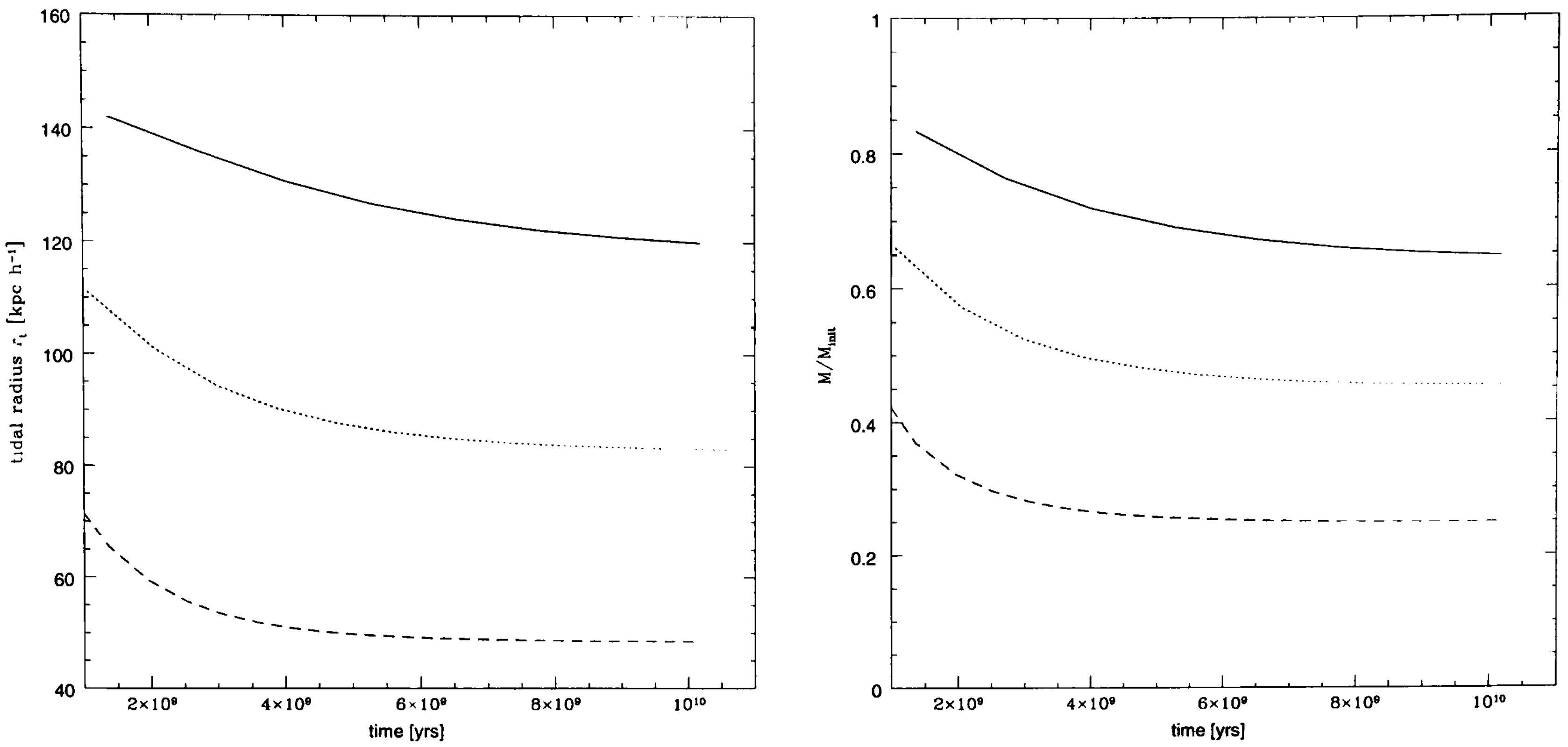


Figure 5.4: Time evolution of satellite tidal radius and bound mass as a fraction of the initial mass. At late times these converge to a constant value.

satellite system appears to loose significantly more mass. Some of this may be attributable to the fact that our satellite has a higher concentration (c_{sat} = 12.2 vs 8.3 in the simulation) and is thus more tightly bound. In addition the host halo used in the simulations is about twice as massive as the one we have used, leading to a larger tidal force.

However, one of the main reasons for the discrepancy between the numerical and our results is likely to be our assumption that the velocity distribution of the satellite particles will always remain (close to a) Maxwellian. Qualitatively our results should also apply to NFW halos with the main result being that adiabatic expansion as a response to tidal stripping of a satellite system can significantly affect the satellite density profile after only a few orbits.

Effect of including central MBHs

The presence of a central MBH in halos is likely to have an impact on tidal mass loss in the central region only within the MBHs range of dynamical influence. Depending on how the MBH in the centre arose, it can either exert a stabilising effect or increase mass loss in this range. If the MBH has grown adiabatically (by baryon accretion or other) at the halo centre, the density profile around the growing MBH develops a steep power law profile or ‘adiabatic spike’ (Bahcall

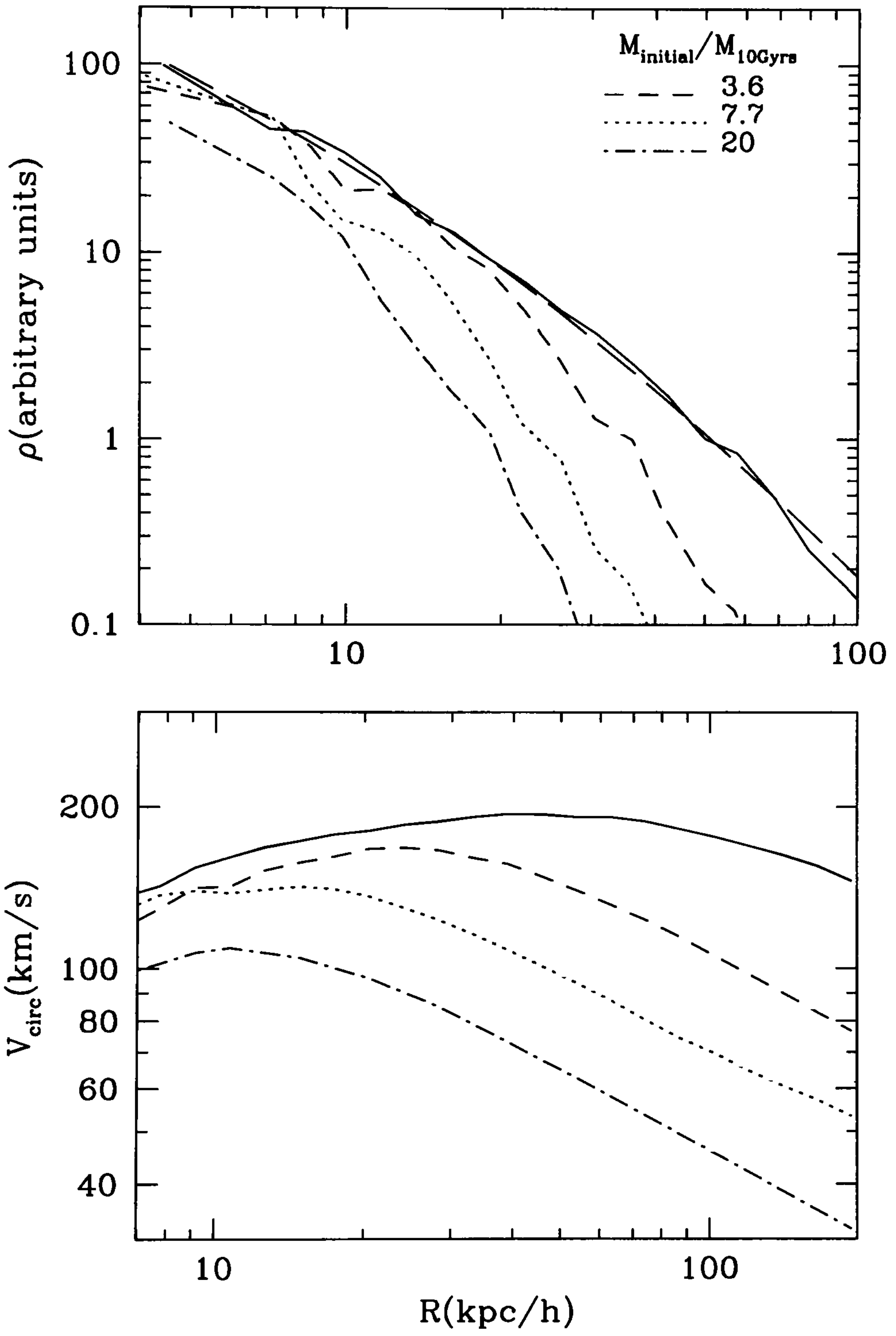


Figure 5.5: N-body simulations (Klypin et al. 1999): Initial satellite density profile (solid line) and after 10 Gyrs on circular orbits at 700, 500, 300 kpc (top to bottom) from the host centre. The long-dashed line plots the NFW profile on which the satellite was modelled.

& Wolf 1976), and particles may well be bound too tightly to be stripped. In this case the core region around the MBH is much more stable to tidal stripping and mass loss.

On the other hand, if the MBH has entered the halo from the outside and spiralled to its centre via dynamical friction, the resulting amount of heat deposited in the central regions of the halo can lead to an expansion of the particle distribution. In this way the central region becomes less tightly bound, possibly leading to more mass loss from these central regions.

This distinction is important for our suggestion in section 6.2.2 that an MBH may be accreting from a baryonic core remnant of the satellite with which the MBH was originally associated. We will argue that most of the remnant MBHs and their associated satellite halos orbiting in present day galactic host halos have not undergone mergers for a long time. It is plausible that through accretion these MBHs have at least recovered any merger-induced decrease in the central density of their satellite, if not built up an adiabatic spike.

We will first, however, investigate how the central density of a halo is affected by the inspiral of a MBH to its centre.

5.3 (S)MBH mergers and the galactic centre

5.3.1 MBH inspiral and the effect on the galactic centre

If MBHs or SMBHs spiral towards a galactic centre, their total binding energy is reduced via dynamical friction and transferred to the ambient medium through which they are moving. The deposition of the MBH energy in the host matter at this location leads to a local increase in velocity dispersion and it seems plausible that this 'heating' of the ambient medium could thus lead to a reduction of the host density at this location.

This effect will presumably be strongest near the centre where the energy of the MBH relative to the host energy becomes increasingly more significant. This is confirmed by N-body simulations of inspiralling MBHs in galaxies (Nakano & Makino 1999a) as well as mergers of galaxies containing central SMBHs (Merritt & Cruz 2001;Milosavljević & Merritt 2001;Milosavljević et al. 2002). It has been suggested that this process is responsible for the lack of observed density cusps in galactic centres (Ravindranath, Ho & Filippenko 2002) that are predicted in high resolution cosmological N-body simulations which do not explicitly model the MBH infall (Navarro, Frenk & White 1997; Moore et al. 1999).

An analytical estimate of this effect is more difficult and will depend crucially on the assumptions made. First and foremost, we require that the whole system remains spherically symmetric and well-defined at all times. In addition, if we decide to view the inspiralling MBH as an extraneous source injecting heat into the host system, energy conservation requires

$$U_0 + T_0 + \Delta H = U_1 + T_1 \tag{5.29}$$

where U and T are the potential and kinetic energy of the host particles before('0') and after ('1') injection of an amount of heat, ΔH, due to the MBH orbital decay. While U_0 and T_0 are both constrained by the initial density profile via the Jeans equation, provided the system is self-gravitating and in equilibrium, we would need at least one more constraint to specify both U_1 and T_1. One is to apply the Jeans equation to the system as a whole after it has reached its final state with the MBH at the centre. An example of this is the approach of Nakano & Makino (1999) which we will discuss below.

Alternatively, we investigated whether the virial condition could be used as a second constraint to solve eq. 5.29. This implies

$$T = -\frac{1}{2}U \tag{5.30}$$

This applies to a system as whole, but for an isothermal sphere is also approximately valid at individual radii, save for those near the outer edge. Equation 5.29 then becomes

$$-T_0(r) + \Delta H = -T_1(r) \tag{5.31}$$

This is already problematic, as the condition is violated by the earlier assumption that the MBH is an 'extraneous' source. If the virial condition is to hold before and after MBH inspiral, then we also have to account for the fact that the MBH does not only deposit heat in the halo but also deepens the halo's gravitational potential. The latter will essentially cancel the effect of any expansion of the halo matter distribution and associated lowering of the density due to MBH heating.

Consider a mass shell in the halo at radius r, with associated mass $\Delta m = 4\ \pi\rho(r)\ r^2\Delta r$. The energy lost and deposited in the halo as the MBH travels from $r + \Delta r \rightarrow r$ is

$$\Delta H = -\left.\frac{GM(<r)M_\bullet}{r}\right|_{r+\Delta r}^{r} \sim \frac{GM(<r)M_\bullet}{r}\frac{\Delta r}{r} \tag{5.32}$$

where $M_\bullet$ is the mass of the MBH. However, the MBH also adds a contribution to the overall potential energy within radius r

$$\Delta T \sim \frac{GM_\bullet}{r}\Delta m \tag{5.33}$$

ΔH and ΔT above are actually of the same order, provided

$$M(r) \sim 4\ \pi\ \rho(r)r^3 \tag{5.34}$$

which is certainly the case for the halo density profiles we are looking at. The reason why the virial condition does not work here, is that the heat energy is deposited locally i.e. at a given radius to raise the host particles kinetic energy, while the potential energy of the MBH is determined by the global mass distribution inside r. In contrast the virial condition assumes a global equilibrium between the kinetic and potential energy.

An alternative approach we tried is to demand that the total mass of host and MBH inside some radius r_e be conserved. To actually determine the change of the density profile due to the effect of the infalling MBH we further assumed that the final density profile was of the same shape as the initial profile, which follows a single power law

$$\rho(r) = \rho_e\left(\frac{r_e}{r}\right)^\gamma \tag{5.35}$$

Keeping the normalisation ρ_e constant, the heating would then lead to a modification in the power law index. If the total mass inside some radius r_e is conserved this yields

$$\int_0^{r_e} 4\pi\rho_0(r)r^2\mathrm{d}r = \int_0^{r_e} 4\pi\rho_1(r)r^2\mathrm{d}r + M_\bullet \tag{5.36}$$

which gives

$$4\pi\ \rho_e\ r_e^3 \frac{1}{3-\gamma_0} = 4\pi \rho_e\ r_e^3 \frac{1}{3-\gamma_1} + M_\bullet \tag{5.37}$$

This will only work as long as the resultant change in slope is small, such that $\delta\gamma = \gamma_1 - \gamma_0 << \gamma_0$ and

$$\frac{\delta\gamma}{(3-\gamma_0)^2} \approx -\frac{M_\bullet}{4\pi\rho_e r_e^3} \tag{5.38}$$

This ansatz also runs into trouble as the result for $\delta\gamma$ depends extremely sensitively on the choice of r_e, which is not obvious. r_e will have to be much larger than any radius within which the enclosed halo mass is so low as to become comparable to that of the MBH. On these scales the MBH is expected to change the halo mass distribution much more dramatically and the assumption of spherical symmetry and $\delta\gamma$ being small break down.

A reasonable estimate is possible, however, for the case of circular MBH orbits with an orbital decay time scale τ_{orb} much larger than an individual orbital period T_{orb}. The energy dissipated by dynamical friction over one orbit is then

$$\Delta H = \alpha_{orb}\ 2\ \pi\ r\ F_{df} \tag{5.39}$$

and is deposited in the halo mass shell of mass δm at that radius. α_{orb} is a dimensionless parameter of order unity and gives the number of orbits over which the MBH slowly migrates inwards so as to cross a halo shell that contains of the order $\sim M_\bullet$ of mass. If α_{orb} is much less than 1, orbits cannot be considered circular anymore as they spiral in too quickly, which typically happens in the highest density regions. Our approach would therefore be expected to break down close to the halo centre.

The migration rate is given by

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{F_{df} r}{M_\bullet v_c} \tag{5.40}$$

where v_c is the circular velocity. For simplicity, consider an isothermal density profile, $\rho = \sigma^2/(2\ \pi\ G\ r^2)$ for which

$$\frac{\mathrm{d}r}{\mathrm{d}t} = -\frac{0.428\ \ln\Lambda_c G M_\bullet}{v_c r} \tag{5.41}$$

During one orbit the MBH then migrates inwards by an amount

$$\Delta r \approx -2\pi r \, \frac{(\,\mathrm{d}r/\,\mathrm{d}t)}{v_c} \tag{5.42}$$

To obtain the mass contained in a shell of width Δr we multiply both sides by $4\pi\ \rho(r)\ r^2$ to get

$$4\ \pi r^2 \rho(r) \Delta r = 0.856\ \ln\Lambda_c \alpha_{orb} M_\bullet \tag{5.43}$$

where we have used $v_c = \sqrt{GM(<r)/r}$ Since we require the LHS to be equal to $M_\bullet$ we can solve for

$$\alpha_{orb} = (0.856\ \ln\Lambda_c)^{-1} \tag{5.44}$$

The extension to other power law density profiles is straightforward, but typically α_{orb} remains of order unity.

For a MBH on a circular orbit, we can use equation 5.4 and replace $M_{sat} = M_\bullet$ and $v_{sat} = v_c$ to get

$$\begin{aligned} \Delta H &= 8\pi^2\ \ln\Lambda_c G \frac{M_\bullet}{M(<r)} \rho(r) r^2 \left[\mathrm{erf}(X) - \frac{2X}{\sqrt{\pi}} \exp^{-X^2} \right] && (5.45) \\ &\propto \frac{M_\bullet^2}{r} && (5.46) \end{aligned}$$

for any density profile that follows a single power law.

As a result the mass shell will expand outwards by an amount

$$\delta r \approx \left(\frac{\mathrm{d}E}{\mathrm{d}r} \right)^{-1} \Delta H \tag{5.47}$$

where

$$\frac{\mathrm{d}E}{\mathrm{d}r} = \frac{\mathrm{d}}{\mathrm{d}r} \left(-\frac{GM(<r)}{r} \right) = -G \left(\frac{\mathrm{d}M(<r)}{\mathrm{d}r} \frac{\delta m}{r} - \frac{M\delta m}{r^2} \right) \tag{5.48}$$

For the displacement of the shell we have used δr to distinguish it from the width of the shell Δr. Assuming no shell crossing occurs, the mass inside any given shell is constant, $\mathrm{d}M(<r)/\,\mathrm{d}r = 0$ and so equation 5.47 becomes

$$\frac{\delta r}{r} = \frac{\Delta H}{|E(r)|} \tag{5.49}$$

The corresponding change in density, $\Delta\rho$ is

$$\Delta\rho = \left.\frac{\Delta m}{4\pi r^2 \Delta r}\right|_{r}^{r+\delta r} \tag{5.50}$$

Since $\Delta r << r$ we eventually get

$$\begin{aligned} \frac{\Delta\rho}{\rho} \approx 2\frac{\delta r}{r} &= 2\frac{\Delta H}{|E(r)|} && (5.51)\\ &\propto M_{\bullet} r^{\gamma-3} && (5.52) \end{aligned}$$

In the central regions of all common halo density profiles $1 \lesssim \gamma \lesssim 2$, which means that an inspiralling MBH results in a density profile with an increasingly flatter slope $\gamma_1(r) < \gamma$ towards the centre. The above approach is only valid for MBHs on circular orbits, but depends neither on any chosen radial scale within which the heating effect is supposed to occur, nor on the assumption that the system remains strictly virialised.

A more general result already alluded to above was obtained by Nakano and Makino (Nakano & Makino 1999a; Nakano & Makino 1999b) in N-body simulations and subsequently justified theoretically. Their ansatz is more general as it starts with the distribution function $f(\varepsilon)$ from which the density profile is inferred as (Binney & Tremaine 1987)

$$\rho(r) = 4\pi \int_0^{\Psi(r)} f(\varepsilon)\sqrt{2[\Psi(r) - \varepsilon]}d\varepsilon \tag{5.53}$$

where ε is the binding energy and $\Psi(r)$ is the depth of the potential well.

According to their approach, the SMBH heating leads primarily to those particles being ejected that are initially most tightly bound, resulting in $f(\varepsilon)$ vanishing at a finite value of the binding energy ε_0, say, and equation 5.53 becomes

$$\rho(r) = 4\pi \int_0^{\varepsilon_0} f(\varepsilon)\sqrt{2[\Psi(r) - \varepsilon]}d\varepsilon \tag{5.54}$$

This can now be expanded for $\Psi(r) > \varepsilon_0$ as

$$\rho(r) = 4\sqrt{2}\pi\sqrt{\Psi(r)} \times \int_0^{\varepsilon} f(\varepsilon)\left(1 - \frac{1}{2}\frac{\varepsilon}{\Psi(r)} + \mathcal{O}\left\{\left[\frac{\varepsilon}{\Psi(r)}\right]^2\right\}\right)d\varepsilon \tag{5.55}$$

However, towards the centre $\Psi(r) \gg \varepsilon$ and so

$$\rho(r) \propto \sqrt{\Psi(r)} \sim \sqrt{\frac{GM_{SMBH}}{r}} \tag{5.56}$$

That means we would expect any inspiralling MBHs to lead to a flattening of the halo density profile with a slope that converges to $\sim -1/2$ towards the centre. This result is in principle valid for any form of $f(\varepsilon)$, with the precise proportionality constant depending on the particular form of the initial distribution function.

The advantage of this approach is that it does not explicitly require MBHs to travel on circular orbits. At the same time, however, it is not obvious how the limiting binding energy ε_0 is determined and at what radial distance the effect becomes significant. The latter is naturally accounted for in our simple scaling solution given above.

5.3.2 An upper limit to MBH merger efficiency

Up to now we have considered any MBH as having merged with the central MBH, when it comes within one per cent (hereafter referred to as the merger region) of the virial radius of the host halo at that time. To adequately follow the evolution of particles on radii smaller than this requires too short and therefore too many time steps, i.e it is computationally too expensive. Even if we could explore this region in a computationally more efficient way, it is questionable whether our simple symmetric halo + bulge + disk model adequately describes the structure of this central region – while the scaling model provides a reasonable description of the outer parts of halos and galaxies, the core structure in observed galaxies typically displays a much wider range of very different profiles, so that a simple model like ours certainly overidealises the structure in these regions. The assumption that MBHs merge with the centre as they cross into this 'merger region' therefore provides only an upper limit on how much the MBH merger process can contribute to the mass of central and halo MBHs. There are various ways in which the actual merger efficiency could be lower than this.

One major source of inefficiency is of course the time it takes for any MBH to spiral into another and typically more massive MBH at the centre of their common host and how likely it is then for the two to merge. One does not necessarily imply the other – at early times halos are smaller, that is, at the first encounter, the two central MBHs within any two halos will start out much closer and so are more likely to spiral to the common centre of the halo merger remnant in a relatively short time. Because there are more low mass halos, this might then give rise to configurations consisting of more than two MBHs and thus the possibility of sling-shot ejections. In other words some fraction of MBHs, although having travelled to the centre quickly, might eventually end up being expelled rather then merging. This has implications for the most massive trees. Halos at the resolution limit in these trees have a mass above M_{min} and therefore might appear in the tree with several seed MBHs which we have thus far assumed have merged to form one MBH (c.f. section 3.2). This may no longer be the case if slingshot ejections occur. Assuming that in this case the lightest MBHs are ejected, however, this should not significantly reduce the mass of the central MBH.

The assumption that MBHs within a kpc or so from the host centre merge efficiently can be used to determine an upper limit on the mass of central SMBHs. Although MBH merging may proceed much less efficiently, we give examples of a range of processes that can lead to rapid merging of MBHs in the galactic context.

If the mass of only the MBHs is considered, their orbital decay time scale in the host can be longer than a Hubble time. However, MBHs typically remain associated with matter from their original satellite, which increases their effective mass by a factor of at least 100 to 1000 and lowers the orbital decay time scale accordingly, allowing even relatively light MBHs ($M_\bullet \geq 10^3$ $M_\odot$) to spiral into the host central region ($\leq$ kpc) within a Hubble time (Begelman, Blandford & Rees 1980;Yu & Tremaine 2002). This is true even if the satellite itself may have actually lost most of its mass ($\geq$ 99 percent) due to tidal stripping inside the host halo and is thus classified as 'naked' in our treatment. This implies that only at high redshifts could seed mass MBHs have travelled to the host centre, since they would have then entered the correspondingly smaller host halo at smaller distances from the centre.

It seems then that dynamical friction can deliver MBHs to the host central regions efficiently where they then form binaries with any MBH already at the centre. The evolution of a MBH binary system in stellar background has been studied extensively (Begelman, Blandford & Rees 1980;Quinlan 1996;Milosavljević & Merritt 2001;Yu & Tremaine 2002) and the 'hardening' stage of binary evolution has been singled out as the 'bottle neck' on the way to the final merger (Milosavljević & Merritt 2001;Yu & Tremaine 2002). With dynamical friction no longer significant and orbital decay due to gravitational wave emission not yet important, one way for the binary MBHs to reduce their orbital radius is by interaction with stars in their vicinity, which can take significantly longer than a Hubble time. Typically, a star in the MBH binary's loss cone (i.e. stellar orbits with pericentres smaller than the binary's radius of influence) gets ejected [3] Depending on whether and how the depleted loss cone can be repopulated, the binary may still merge efficiently enough. In a well-defined static potential, two-body relaxation can lead to the diffusion of stars into the loss cone. This process is further enhanced if the binary is 'wandering' within a vicinity of the galactic centre. In this case the effective solid angle of the loss cone is increased. Depletion of the cone will thus last longer and the rate of stars diffusing into it will also be larger. It was shown recently that the diffusion rate of stars into the loss cone can increase further if stellar orbits are chaotic, which can occur in non-axisymmetric galactic potentials (Merritt & Poon 2003). As a result there is a larger reservoir of stars – potentially by orders of magnitude – that the binary can interact with. In principle this argument should also hold for MBHs in their respective satellites. In fact, due to the higher number of major mergers at early times the potentials of the (mini) galaxies hosting central MBHs are much more likely to be non-axisymmetric.

However, even in the case that binary decay through interactions with stars takes prohibitively long, the presence of gas may be of crucial importance in this context (Milosavljević & Merritt 2001). High densities of gas between the binary MBHs could allow for a much faster evolution and eventual merger of the binary. Several scenarios have been suggested for this, such as a massive gas disk around the binary (Gould & Rix 2000; Armitage & Natarajan 2002) or massive gas inflow (see e.g. Begelman, Blandford & Rees 1980) in the wake of major mergers. Hydrodynamical simulations of galaxy mergers, for instance, find that up to 60 per cent of the total gas mass of two merging Milky Way sized galaxies can end up within a region only a few hundred parsecs across, which is about half the bulge scale radius (Barnes & Hernquist 1996; Naab & Burkert

---

[3] These are stages 4 to 6 of the merger process as described by Begelman, Blandford & Rees (1980), c.f. section 3.2.1.

2001; Barnes 2002).

Here we assume that during major mergers the gas infall will actually lead to all MBHs binaries merging. We also neglect the possibility of triple BH interactions and sling-shot ejections. This issue will be considered in more detail in chapter 6 where we also look at the gravitational wave emission from merging MBHs (section 8.1.3).

5.4 Summary overview of modelling procedure and participating processes

Figure 5.6 summarises the key elements and stages of our computational scheme used to obtain the results discussed in subsequent chapters. On the left we outline the merger tree construction stage. This consists of an Extended Press-Schechter algorithm for the generation of Monte-Carlo merger trees backwards in time departing from a present-day halo of specified mass. The 'real' work only begins once the set of precursor halos is re-assembled, i.e. the tree is sun forward again - with a difference. Whenever the tree indicates a merger has taken place, we now explicitly follow the dynamical evolution of merged sub-structure in the new host halo. By definition this constitutes the main branch of the merger hierarchy; substructure in side branches is assumed to merge on a dynamical time scale.

We stress that the dynamical processes included in the final computational scheme are only those discussed in section 5.1. We have neither included the iterative treatment of density profile erosion and adiabatic back reaction in satellites as outlined in section 5.2, nor the effect of inspiralling MBHs on the density in the galactic centre as discussed in section 5.3.

A satellite is supposed to have been disrupted when, through stripping, its mass has become less than about 0.3 % of its initial mass - the actual criterion refers to a limiting tidal radius of a satellite system, beyond which it is considered disrupted. This criterion roughly corresponds to the 0.3 % mass limit we used. Any MBH contained in such a disrupted halo, is labelled as 'naked'; a small core of material around the MBH may still remain, however, as we discuss in the next chapter.

A satellite and/or MBH is considered to have fallen to the centre of the host halo when its dynamical evolution carries it to radial distances from the host centre less than some infall radius, which is about 1% of the host virial radius at the time. We have already discussed this assumption in section 5.3.2. In the context of gravitational waves from MBH mergers (c.f. section 8.1.3) we improve on this by including an analytical estimate of the dynamical friction time for a satellite/MBH to travel from the infall radius to the host centre.

5.5 Summary

In this chapter we have introduced our numerical scheme that is used to follow the dynamical evolution of satellite + MBH systems in larger host halos throughout a halo merger tree. In this context we have investigated in greater detail by analytical means two aspects of the dynamical evolution that are not included in the numerical scheme. In particular

- We have looked at how the tidal field of a host halo erodes the density profile of a satellite halo when the latter travels on essentially circular orbits. In contrast to the instantaneous tidal truncation used in the numerical scheme, we have also accounted for the adiabatic response of the satellite to initial tidal truncation. Including this effect can lead to significant tidal erosion of the satellite density profile over several orbits.
- We have given a simple scaling approximation for the expected decrease in the central density profile of a halo as a MBH spirals to its centre.

While important for the structure of an individual host or satellite halo, the omission of these effects should not influence the statistical results of our simulations, which are the subject of the next chapter.

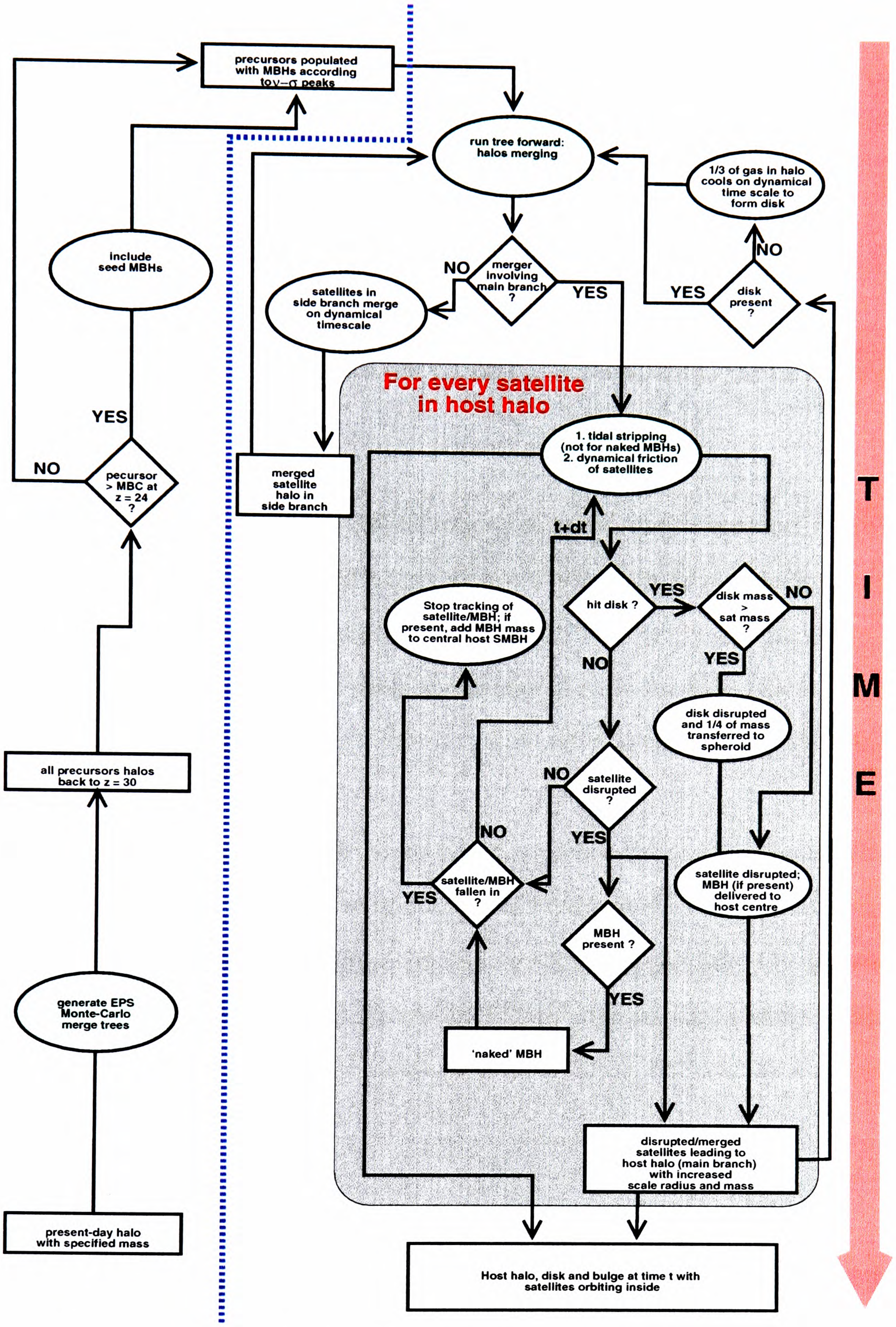


Figure 5.6: Flow diagram of the computational model used. To the left of the dashed blue line lies the tree construction stage. This consists of the generation of Monte-Carlo merger trees backwards in time in the context of the Extended Press-Schechter formalism. The precursor halos obtained such, are then re-assembled taking into account the subsequent dynamical evolution once any two halos have merged. This corresponds to the numerical procedure shown to the right of the dashed blue line. The key stage processes are highlighted in green. The key processes and events during the dynamical evolution are contained in the grey shaded box. Note that the dynamical processes shown here (c.f. green enclosure inside grey box) are *only* those discussed in section 5.1.

Chapter 6

MBHs in present-day Galactic Halos

A fundamental prediction of the hierarchical merging of MBHs is the existence of an abundance of MBHs throughout galaxies and their halos. Other MBHs may have already travelled to the centre and merged there to help build up the central SMBH we see in galactic centres today. In this chapter we determine the number, mass and accretion rates for halo MBHs and the mass of the central (S)MBH. We will use these to estimate observational signatures of MBHs in the next chapter.

In section 6.1 we determine the MBH mass function and radial distribution of MBHs for four final halo masses. We also look at the total number of MBHs in the disk and bulge as well as in Earth-centred volumes. For two different accretion scenarios we compute the accretion rates of these MBHs. This is done in section 6.2. In section 6.3 we estimate the masses of central SMBHs that have build up through merging of smaller MBHs. This is followed by a summary of our results in section 6.4.

6.1 Abundance and Mass of MBHs in galactic halos

6.1.1 Abundance of MBHs in galactic halos

In figure 6.1 we show the abundance of all MBHs for models A, C and D and for all final halo masses. Model B, which only differs from model A by its different MBH seed mass is considered below.

In addition the total number of MBHs in the halo is given in table 6.1. For Milky Way sized

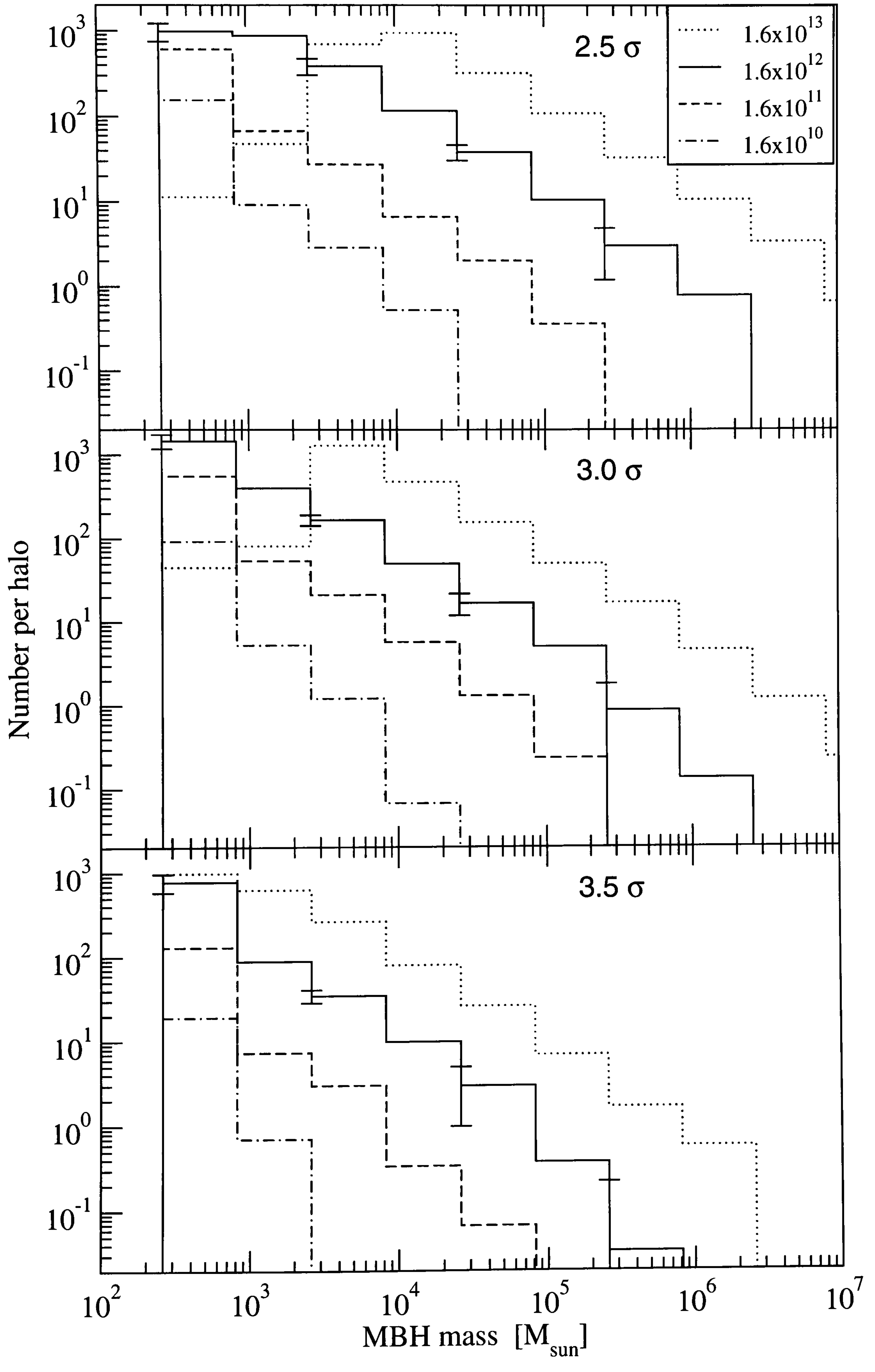


Figure 6.1: Abundance of all MBHs in the halo for models c, a and d (top to bottom panels) averaged over 30 trees with error bars corresponding to the standard deviation.

Table 6.1: Number of MBHs in halo averaged over thirty trees with associated standard deviation. We have shown the total number as well as the number in the bulge within two bulge scale radii and in the disk within two disk scale lengths and scale heights.

Halo mass	model A	model B	model C	model D
total # MBHs in halo				
1.6×10^{10}	100 ± 17	76 ± 13	170 ± 38	20 ± 6
1.6×10^{11}	640 ± 110	570 ± 120	710 ± 76	140 ± 24
1.6×10^{12}	2090 ± 360	1750 ± 170	2430 ± 550	910 ± 220
1.6×10^{13}	2130 ± 230	2250 ± 490	2200 ± 530	1970 ± 210
# MBHs in bulge				
1.6×10^{10}	0	0	0	0
1.6×10^{11}	2.3 ± 1.7	2.3 ± 1.9	4.1 ± 2.6	0.5 ± 0.7
1.6×10^{12}	36 ± 17	35 ± 13	36 ± 11	11 ± 6
1.6×10^{13}	93 ± 30	72 ± 25	77 ± 26	81 ± 29
# MBHs in disk				
1.6×10^{10}	0	0	0	0
1.6×10^{11}	0.9 ± 1.1	1.1 ± 1.2	2.3 ± 1.8	0.3 ± 0.6
1.6×10^{12}	9.3 ± 5.2	9.2 ± 4	9.2 ± 3.5	3.2 ± 2
1.6×10^{13}	14.6 ± 6.6	10.3 ± 4.3	12.6 ± 4.8	11.7 ± 5

halos, for instance, we would expect between 900 to 2100 MBHs to orbit within the galactic halo depending on the model. Here as in all other plots of average abundances and luminosities (c.f. chapter 6) we have stated the standard deviation rather than the error on the 'mean' as a measure of the uncertainty of our results. We say more about this choice in the context of comparison with observations in section 7.7.

We found that the number of MBHs in the galactic disk out to about two disk scale radii is less than 0.2 per cent of the total number of MBHs for all final halo masses. Part of the reason for this low number is that a lot of the MBHs in the disk are orbiting at small distances of less than 1 per cent of the host virial radius and are therefore counted as having fallen to the centre since their dynamics cannot be traced accurately any more as mentioned above. Conversely the high mass end implies that apart from the central SMBH there will be one or two other MBH of about a tenth of its mass orbiting in the halo

While the mass functions display a uniform slope of $N_\bullet$ d $\ln M_\bullet \propto M_\bullet^{-1}$ d $\ln M_\bullet$ it is interesting to note that particularly in model A and C the mass function of MBHs in the most massive halo displays a downward departure from a power law for MBH masses less than about $M_\bullet \sim$

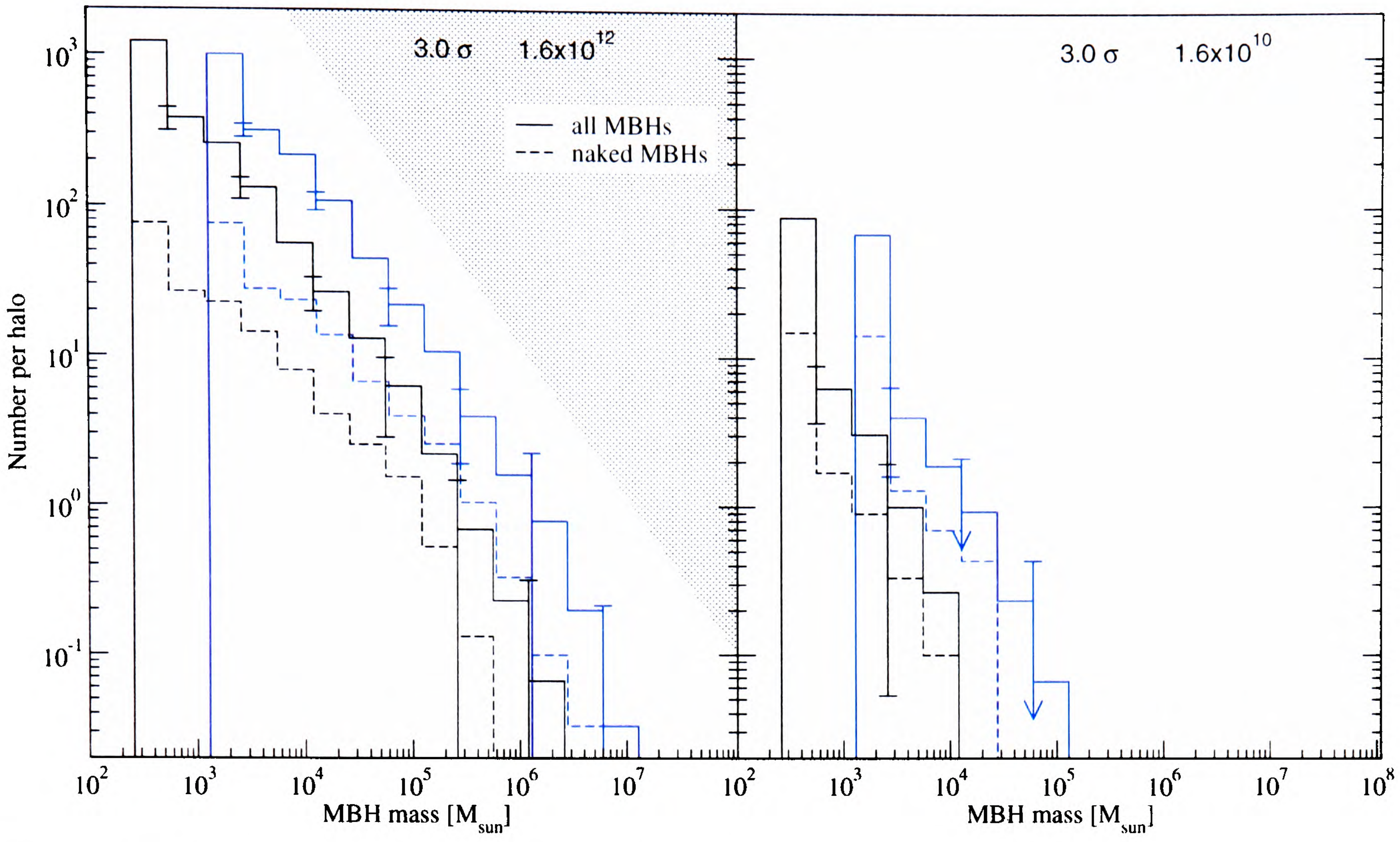


Figure 6.2: Comparison of models A and B: Abundance of all and naked MBHs in galactic halos of mass 1.6×10^{12} $M_{\odot}$(left panel) and 1.6×10^{10} $M_{\odot}$(right). The set of curves to the left and right are for seed MBH masses of 260 and 1300 $M_{\odot}$respectively. Provided the shape of the mass function remains the same for different MBH seed masses, the shaded area in the left panel indicates the prohibited regime where seed MBHs would need to be more massive than the total amount of baryons available in the original MBC halo within which they formed.

$3 \times 10^3 - 10^4$ $M_{\odot}$. As a result the total number of MBHs in the most massive halo is less than expected and is actually quite similar to that in the second most massive halo in models A and C. This is due to the mass resolution limit of the semi-analytical code as explained in section 5.1.1. For the most massive final halos this limit is larger and consequently more halos entering the hierarchy at this limit may already contain more than one seed MBHs, which we then assume have merged. The latter means that the number of MBHs with the original seed mass is reduced.

In figure 6.2 we show the average MBH abundance including that of naked MBHs within the virial radius of the 1.6×10^{12} and 1.6×10^{10} $M_{\odot}$halos respectively for models A and B.

Compared to the mass of the bulge, disk and halo the seed MBH masses are small and so we would not expect them to significantly affect the evolution of substructure within the host. For this reason we find that, except for the high mass end, the MBH mass functions for the two different MBH seed masses are essentially the same but are offset from one another along the

Table 6.2: Total mass in halo MBHs and mas of central SMBH for all final halo masses and all models. The total mass in MBHs is typically 2 – 3 times larger than the mass of the central SMBH.

		Halo mass	
		1.6×10^{10}	1.6×10^{11}
model A	$\sum M_{MBH}[M_\odot]$	$(3.4 \pm 0.59) \times 10^4$	$(3.8 \pm 0.62) \times 10^5$
	$M_{SMBH}[M_\odot]$	$(1.55 \pm 0.67) \times 10^4$	$(1.85 \pm 0.63) \times 10^5$
model B	$\sum M_{MBH}[M_\odot]$	$(1.34 \pm 0.36) \times 10^5$	$(1.68 \pm 0.32) \times 10^6$
	$M_{SMBH}[M_\odot]$	$(7.36 \pm 3.36) \times 10^4$	$(8.83 \pm 3.34) \times 10^5$
model C	$\sum M_{MBH}[M_\odot]$	$(6.48 \pm 0.98) \times 10^4$	$(4.79 \pm 1.49) \times 10^5$
	$M_{SMBH}[M_\odot]$	$(2.84 \pm 0.9) \times 10^4$	$(3.62 \pm 1.5) \times 10^5$
model D	$\sum M_{MBH}[M_\odot]$	$(5.75 \pm 1.84) \times 10^3$	$(5.63 \pm 0.99) \times 10^4$
	$M_{SMBH}[M_\odot]$	$(3.16 \pm 1.48) \times 10^3$	$(3.34 \pm 1.32) \times 10^4$
		1.6×10^{12}	1.6×10^{13}
model A	$\sum M_{MBH}[M_\odot]$	$(3.17 \pm 0.76) \times 10^6$	$(3.42 \pm 0.78) \times 10^7$
	$M_{SMBH}[M_\odot]$	$(1.5 \pm 0.64) \times 10^6$	$(1.14 \pm 0.29) \times 10^7$
model B	$\sum M_{MBH}[M_\odot]$	$(1.27 \pm 0.4) \times 10^7$	$(1.78 \pm 0.31) \times 10^8$
	$M_{SMBH}[M_\odot]$	$(6.93 \pm 2.81) \times 10^6$	$(7.42 \pm 2.93) \times 10^7$
model C	$\sum M_{MBH}[M_\odot]$	$(7.59 \pm 1.04) \times 10^6$	$(7.00 \pm 1.27) \times 10^7$
	$M_{SMBH}[M_\odot]$	$(3.23 \pm 1.21) \times 10^6$	$(2.96 \pm 1.29) \times 10^7$
model D	$\sum M_{MBH}[M_\odot]$	$(6.24 \pm 0.97) \times 10^5$	$(4.9 \pm 1.08) \times 10^6$
	$M_{SMBH}[M_\odot]$	$(2.57 \pm 1.06) \times 10^5$	$(2.33 \pm 1.33) \times 10^6$

ordinate (representing the actual MBH mass) by a constant factor that is more or less equal to the ratio of the initial seed MBH masses. Based on this the line bounding the grey shaded area in figure 6.2 represents the inferred mass function for a seed MBH with a mass of 1.3×10^4 $M_\odot$, that is the case where the entire baryonic mass of an MBC halo collapses into the black hole.

For a final halo mass of 1.6×10^{12} on both model A and B we can deduce from figure 6.2 that the number N of remnant MBHs in the halo follows a power law

$$N_\bullet \ \mathrm{d} \ \ln M_\bullet \propto M_\bullet^{-1.01 \pm 0.04} \ \mathrm{d} \ \ln M_\bullet$$

which is also the basis on which we have determined the line bounding the prohibited region for any MBH mass function. It is difficult to establish a similar uniform power law for halos lighter than this. In this case not as many massive MBHs have formed and the shape of the MBH mass function is thus dominated at the low mass end (near the MBH seed mass) by the discrete nature of the MBH mass increase.

Figure 6.3 shows the number of MBHs as a function of distance from the host centre for model

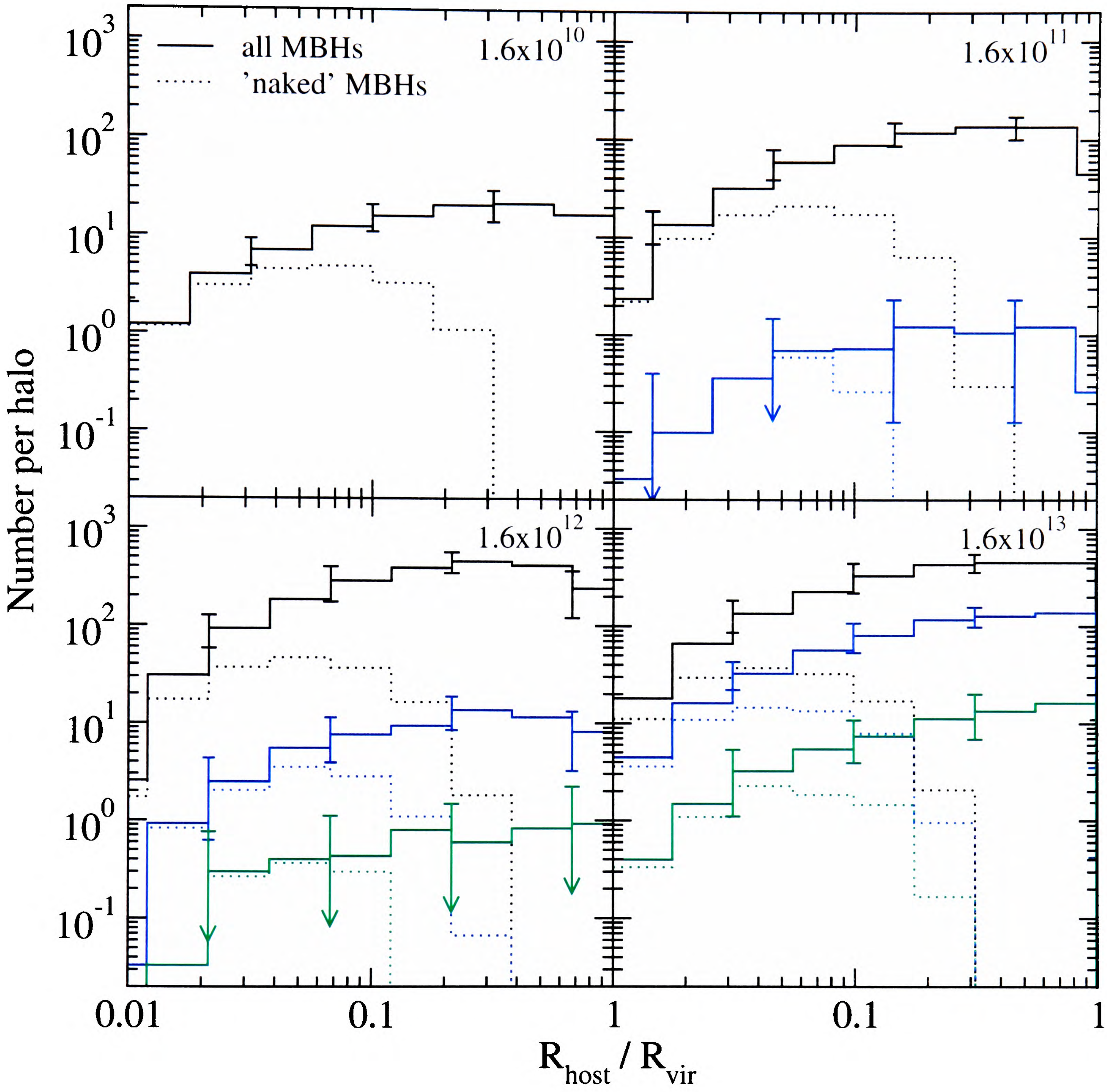


Figure 6.3: Radial distribution of MBHs for model A averaged over thirty trees for all four final halo masses. The contribution to the total of naked MBHs is shown by the dotted lines. Blue and green colours represent the radial distribution for MBHs more massive than 10^4 and 10^5 $M_{\odot}$respectively. The relative distributions of MBHs above the different mass thresholds are similar, implying that there is no obvious mass segregation.

A only. We have also plotted the contribution of MBHs more massive than 10^4 and 10^5 $M_\odot$, shown in blue and green respectively. In addition the plot shows the fraction of naked MBHs for all of the above cases. As expected there are more naked MBHs near the centre where the steeper host gravitational potential results in stronger tidal forces that strip away more matter from any satellite present there. In all cases shown the contribution of naked MBHs typically becomes dominant at distances of less than a few percent of the host virial radius. With increasing halo mass the relative contribution of naked MBHs at small radii decreases. A reason for this is that in small halos the tidal gradient is relatively steeper and MBHs are therefore more likely to be stripped. For the different MBH mass cuts, however, there does not appear to be any significant difference in the relative radial distribution of naked MBHs.

In figure 6.4 we have shown the cumulative abundance of MBHs below a given distance from the halo centre. The distribution is scaled to the virial radii of the respective halos and results are shown for models A,C & D. What this shows us is that the shape of the radial distribution is fundamentally the same in all cases except for the normalisation. The latter corresponds to the difference in the total MBH abundance as shown in figure 6.1. We also note that, while different in all other cases, the radial distribution of MBHs are very similar again for final halo masses of 1.6×10^{12} and 1.6×10^{13} $M_\odot$in models A and C. This is for the same reasons given above in respect of the corresponding mass functions.

An important result that emerges from both figures 6.1 and 6.4 is that both the shape of the mass abundance as well as the radial distribution of MBHs in halos is very similar. The different normalisation of the number of and mass in MBHs for the different halo masses and models is summarised in tables 6.1 and 6.2. For comparison the latter also lists the mass of the respective central SMBHs. In section 6.3 we will see that the mass of SMBHs is tightly correlated with the mass of the galactic bulge component. The table shows that the mass contained in halo MBHs is typically between two to three times larger than the mass of the central SMBH.

Together with the correlations between M_{SMBHS} and M_{bulge} and the more trivial one between M_{bulge} and M_{halo} this implies that a similar correlation exists between the total mass contained in MBHs and their respective halo masses. In fact, for all four halo masses this correlation is

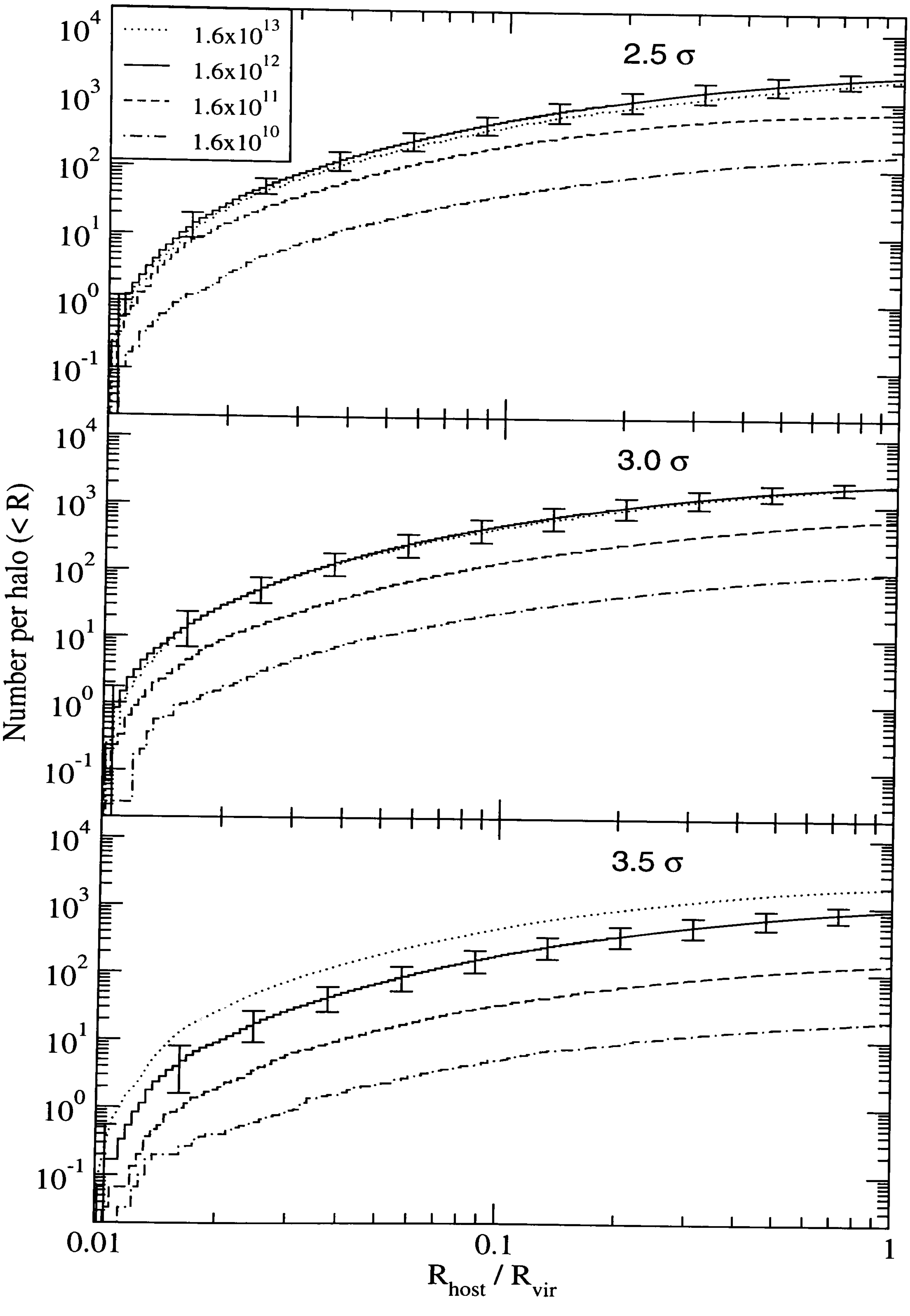


Figure 6.4: Number of MBHs below radius R from host centre, averaged over all trees and scaled to the respective virial radius for the individual halos. Error bars corresponding to the standard deviation are shown for the case of the 1.6×10^{12} $M_\odot$ halo.

Table 6.3: Abundance of MBHs in Earth-centred volumes at 8.5 kpc from the galactic centre in the Milky-Way-sized halo (1.6×10^{12} $M_\odot$). Given are the average over thirty trees with their respective standard deviation.

	Distance from Earth Δr [kpc]		
	2.0	2.5	3.0
Model A	0.87 ± 0.94	1.7 ± 1.29	2.57 ± 1.76
Model B	0.43 ± 0.63	1.03 ± 0.96	2.00 ± 1.56
Model C	0.53 ± 0.73	1.4 ± 1.28	2.47 ± 1.87
Model D	0.23 ± 0.63	0.4 ± 0.72	0.70 ± 0.84

consistent with

$$\sum M_\bullet \propto M_{halo}^{1.0\pm0.03}$$

which is what we would expect, since any halo increasing its mass through the accretion of smaller halos will also inherit all MBHs associated with the latter.

Within the standard deviation quoted we expect the number and mass abundance of MBHs particularly in the 1.6×10^{12} $M_\odot$halo to be representative of Milky-Way-sized galaxies in currently favoured ΛCDM cosmologies.

In table 6.3 we have listed the average abundance of MBHs in a local Earth centred volumes (which we have taken as corresponding to a volume centred at 8.5 kpc from the centre of a 1.6×10^{12} $M_\odot$ halo in our simulations). Virtually all of these will be seed BHs that have not yet merged and in the absence of any growth process other than hierarchical merging their mass will be equal to that of the initial seed BHs. To the extent that the large standard deviations allow for any meaningful comparison between the models, one notable feature is the marked difference between the average number of nearby MBHs in models A and B. We already mentioned that we would not expect any significant difference in the halo (merger) dynamics and thus the final MBH abundance because the models only differ in that they start out with different MBH seed masses. However, the solar neighbourhood at 8.5 kpc from the centre is well within the radial regime where the MBH abundance is becoming increasingly dominated by naked MBHs. At distances smaller (larger) than the sun's orbit there are relatively more (less) naked MBHs and because these are more massive they move towards the centre more quickly. This would explain why the number of nearby MBHs is consistently lower in model B for all local distances considered.

In chapter 6 we will address the question of how these MBHs could possibly be detected.

Our findings for the MBH abundance in halos are in accord with the results of another recent investigation by Volonteri, Haardt & Madau (2002, hereafter VHM03) . We find that the total MBH mass density in a Milky-Way sized galactic in our model C (with a MBC collapse threshold of 3.5 σ) halo agrees to within a factor 2 with their value inferred from the density function of 'wandering' BHs in galactic halos.

6.1.2 Constraints on initial MBH mass function

Figure 6.2 for the two different MBH seed masses gives some indication of the effect of other changes in the masses and numbers of seed MBHs in the primordial halos.

We have seen above that the MBH mass functions are shifted along the M_{MBH} axis in proportion to the mass of the seed MBHs. This mass, however, cannot be higher than the total baryonic mass contained in the original MBC halos. This translates into the grey shaded area shown in figure 6.2 and thus any mass for a single seed MBH between 260 and 1.3×10^4 $M_\odot$will lead to a present-day MBH mass function that lies between the shaded area and the mass function corresponding to model A.

By conservation of mass [1], if the primordial halo contains more than one MBH of different masses in the range 260 $M_\odot < M_\bullet <$ 13000 $M_\odot$then the resulting mass function will again lie between the bottom and the top one shown, but will have a different slope. If initially one or more MBHs were present with masses lower than 260$M_\odot$, the present-day mass function will correspondingly extend to lower masses, but will otherwise still be limited by the shaded area. This means, that even though we had initially made a fairly specific choice for the initial MBH mass function in the primordial halos, any general form for the MBH IMF is expected to lead to results within the limits provided by the MBH mass functions shown, if there is at least one seed MBH of 260 $M_\odot$or larger. This is *provided the seed MBHs form in* 3σ *peaks* as is the case for model A and B. Seed MBH formation in higher or lower peaks changes the overall normalisation and leads to a corresponding scaling of the mass functions. The relative range of possible mass

---

[1] Strictly the masses of two merging BHs are not conserved, but will be lower by a few per cent, since gravitational waves can radiate away some of the BHs' rest mass energy. In the following we assume that this effect only changes our results by a negligible amount, although the mass loss through gravitational radiation accumulated in many mergers for some MBHs may become significant.

functions should nevertheless remain reasonably well defined unless seed MBHs form in MBC halos collapsing from an extended range of peak heights and thus redshifts.

We need to stress that the above depends on the assumption that all MBHs falling to within one per cent of the virial radius merge efficiently in all halos merging along the way to produce the final host halo.

If the only or at least most significant source of seed MBHs is that forming in MBC halos then the total mass contained in halo MBHs can be used to normalise the initial mass function of seed MBHs, to which it is related by the background cosmology. The latter determines the average merger history of halos and thus the average number of MBC halos ending up in more massive halos later on. Note that this is not much affected by the merger efficiency of MBHs since the present-day MBH mass function is dominated by seed MBHs that have not merged, and that contribute a similar amount to the total mass contained in halo MBHs as the few very massive MBHs that have resulted from multiple mergers of seed MBHs. This is just expressing in a different way the $N_\bullet \, \mathrm{d} \ln M_\bullet \propto M_\bullet^{-1} \, \mathrm{d} \ln M_\bullet$ scaling we found earlier, which means that the total mass contributed from successive logarithmic MBH mass intervals is constant.

6.2 MBH mass accretion rates

6.2.1 Bondi-Hoyle accretion

If we consider a MBH travelling within a uniform medium, the steady state accretion rate is given by the Bondi-Hoyle accretion rate (Bondi & Hoyle 1944; Bondi 1952)

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \pi r_{acc}^2 \sqrt{v_\bullet^2 + c_s^2} \rho_g \tag{6.1}$$

Here c_s is the sound speed in the gas and ρ_g its density – both far from the MBH. $v_\bullet$ is the velocity of the MBH and r_{acc} the accretion radius

$$r_{acc} = \frac{2GM_\bullet}{v_\bullet^2 + c_s^2} \tag{6.2}$$

giving an accretion rate

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \frac{4\pi G^2 M_\bullet^2 \rho_g}{c_s^3} (1 + \beta_s^2)^{-3/2} \tag{6.3}$$

where we have used $\beta_s \equiv v_\bullet/c_s$ (Chisholm, Dodelson & Kolb 2002). If the MBH accretes adiabatically from a gas of pure hydrogen this is

$$\begin{aligned} \dot{M} &= 8.77 \times 10^{-12} \left(\frac{M_\bullet}{100 M_\odot}\right)^2 \left(\frac{\rho_g}{10^{-24}\mathrm{g\ cm^{-3}}}\right) \\ &\times \left(\frac{c_s}{10\ \mathrm{km\ s^{-1}}}\right)^{-3} (1+\beta_s^2)^{-3/2}\ M_\odot \mathrm{yr}^{-1} \end{aligned} \tag{6.4}$$

It is implicit that only baryonic matter can get accreted in this way, assuming that a mechanism exists for the dissipation of thermal energy as the material falls towards the MBH (see section 7.3 on accretion processes). Dark matter, on the other hand is by definition not or only weakly interacting. Dissipation therefore cannot occur and dark matter only gets accreted if it approaches the MBH to within a distance that is of the order of the last stable orbit of the MBH. This is much smaller - i.e. it has a much smaller 'cross section' - than the Bondi accretion radius and we will subsequently neglect the possibility of dark matter accretion.

In what follows we assume that while the nature of the accretion process is somewhat uncertain, the overall mass accretion rate is essentially determined by the Bondi-Hoyle formula. This implies, that we neglect the possibility that the mass accretion rate is modified e.g. by a non-negligible mass of an accretion disk that may form around the accreting MBHs.

6.2.2 Accretion environment: ISM vs. baryonic core remnants

As the MBHs orbit through the host they accrete matter from the host ISM. This will only be significant in regions with relatively large amounts of gas, which is the case primarily in the galactic disk and bulge.

Alternatively the MBHs may accrete from a core remnant of the satellite they were originally associated with. In our numerical procedure a satellite is considered tidally disrupted when the tidal radius becomes smaller than the scale radius of the satellite density profile. This condition is sufficient as far as the dynamical importance of the satellite is concerned, since at this stage it will have lost all but at most a few percent of its mass. A core remnant close to the satellite centre may still remain, however. While insignificant for the overall dynamics of the satellite the amount of baryonic matter contained in this core could still contribute significantly to the accretion onto a MBH present at its centre and so boost its accretion luminosity potentially by orders of magnitude. In this case MBH accretion is independent of the conditions and relative MBH velocity in the surrounding ISM of the host halo.

To determine the mass accretion rate we still need an estimate of the gas density in the baryonic core that is essentially acting as fuel supply travelling along with the MBH. Assuming that all baryonic matter has cooled in the satellites before they are subject to tidal stripping and heating in the host the outer radius of the baryonic core assumed spherical is $r_b \sim 0.1 r_{vir}$ where r_{vir} is the virial radius of the unstripped satellite halo of mass M_{vir}. If all baryons are in the form of gas and the baryon fraction in the satellite is cosmological, i.e. $M_b = (\Omega_b/\Omega_m)\ M_{vir}$, then the mean gas density is

$$\rho_g = \frac{3M_b}{4\pi r_b^3} = \frac{1}{0.001}\frac{\Omega_b}{\Omega_m}\frac{3M_{vir}}{4\pi r_{vir}^3} \approx 2.39 \times 10^2 \frac{\Omega_b}{\Omega_m}\frac{M_{vir}}{r_{vir}^3} \tag{6.5}$$

Substituting this into equation (7.1) and using $c_s \approx 10$ km s^{-1} for ISM at 10^4 Kelvin we can determine the mass accretion rate

$$\begin{aligned} \dot{m} &\approx 6.25 \times 10^{-8} \left(\frac{M_{MBH}}{M_\odot}\right) \frac{\Omega_b}{\Omega_m} \left(\frac{M_{vir}}{10^5 M_\odot}\right) \\ &\times \left(\frac{r_{vir}}{\text{kpc}}\right)^{-3} \left(\frac{c_s}{10 \text{ km s}^{-1}}\right)^{-3} \end{aligned} \tag{6.6}$$

and thus the luminosity. Here $\dot{m}$ is the mass accretion rate in units of the Eddington accretion rate

$$\dot{m} \equiv \frac{\dot{M}c^2}{L_E} = 1.53 \left(\frac{\dot{M}}{10^{17}\text{g s}^{-1}}\right) \left(\frac{M_\bullet}{M_\odot}\right)^{-1} \tag{6.7}$$

6.2.3 Distribution of mass accretion rates

In figure 6.5 we have plotted the number of MBHs vs their respective mass accretion rates for ISM as well as baryonic core accretion. The plot serves primarily to highlight the vast difference in accretion rates for the two models. In general the accretion rates for ISM accretion are lower than that for baryonic core accretion by some 7 - 10 orders of magnitude! We can see that for the case of ISM accretion the maximum accretion rate does not exceed $\sim 10^{-7}$ of the Eddington value. It seems then that ISM accretion in the context of our model is completely insignificant.

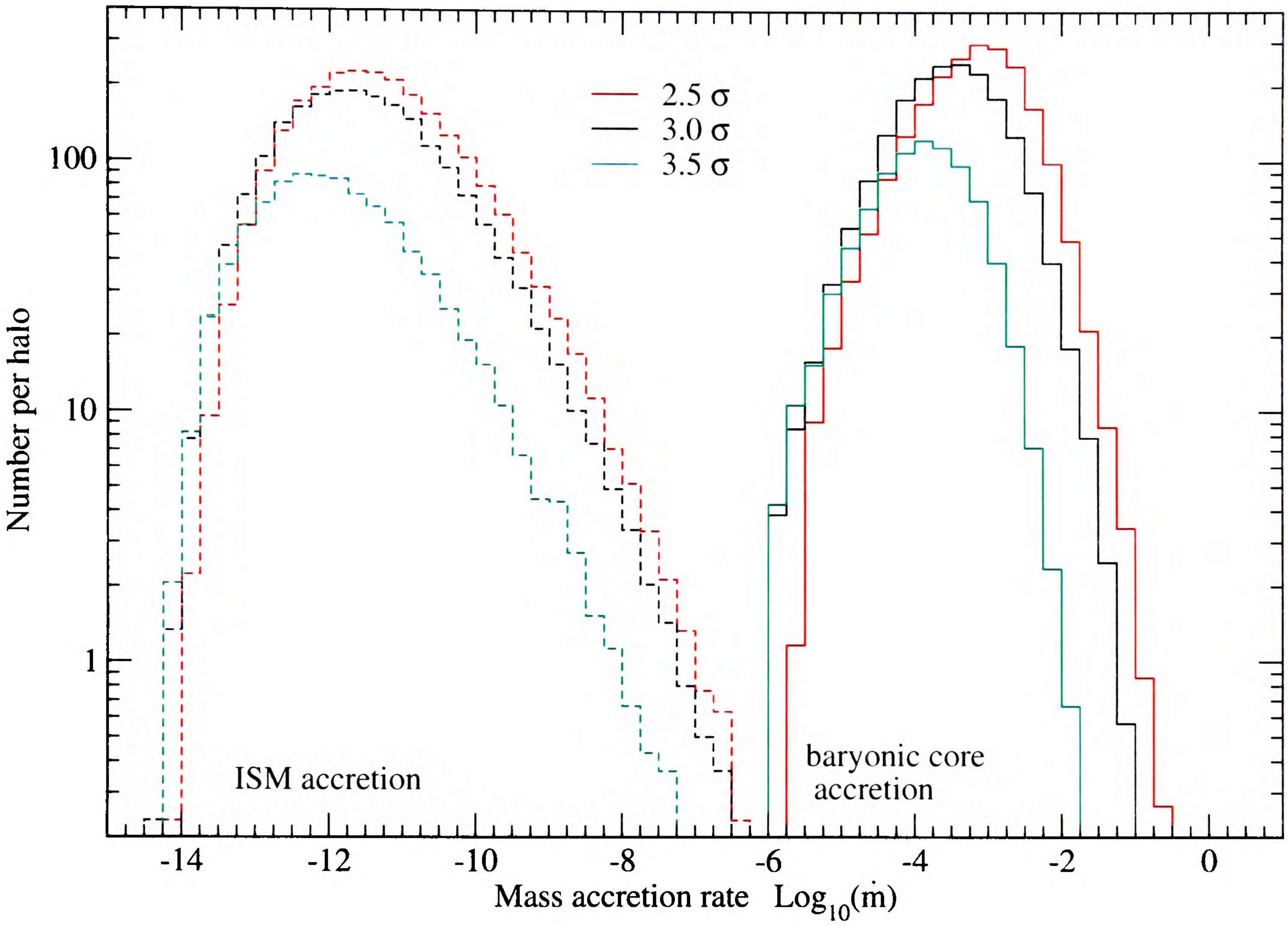


Figure 6.5: Abundance of accretion rates for MBHs in a Milky-Way sized halo for models A, C and D. The curves on the left correspond to ISM accretion, the ones on the right to baryonic core accretion.

There are two reasons for this. Firstly, MBHs are distributed across the halo as we have seen above, with the number of MBHs in the disk and bulge small and certainly very much lower than that of stellar mass BHs. Secondly, even in the disk and bulge the actual density of accretable gas is too low especially at late times. In our analysis of the observable signatures of accreting MBHs we therefore mainly focus on baryonic core accretion for which we get consistently larger accretion rates.

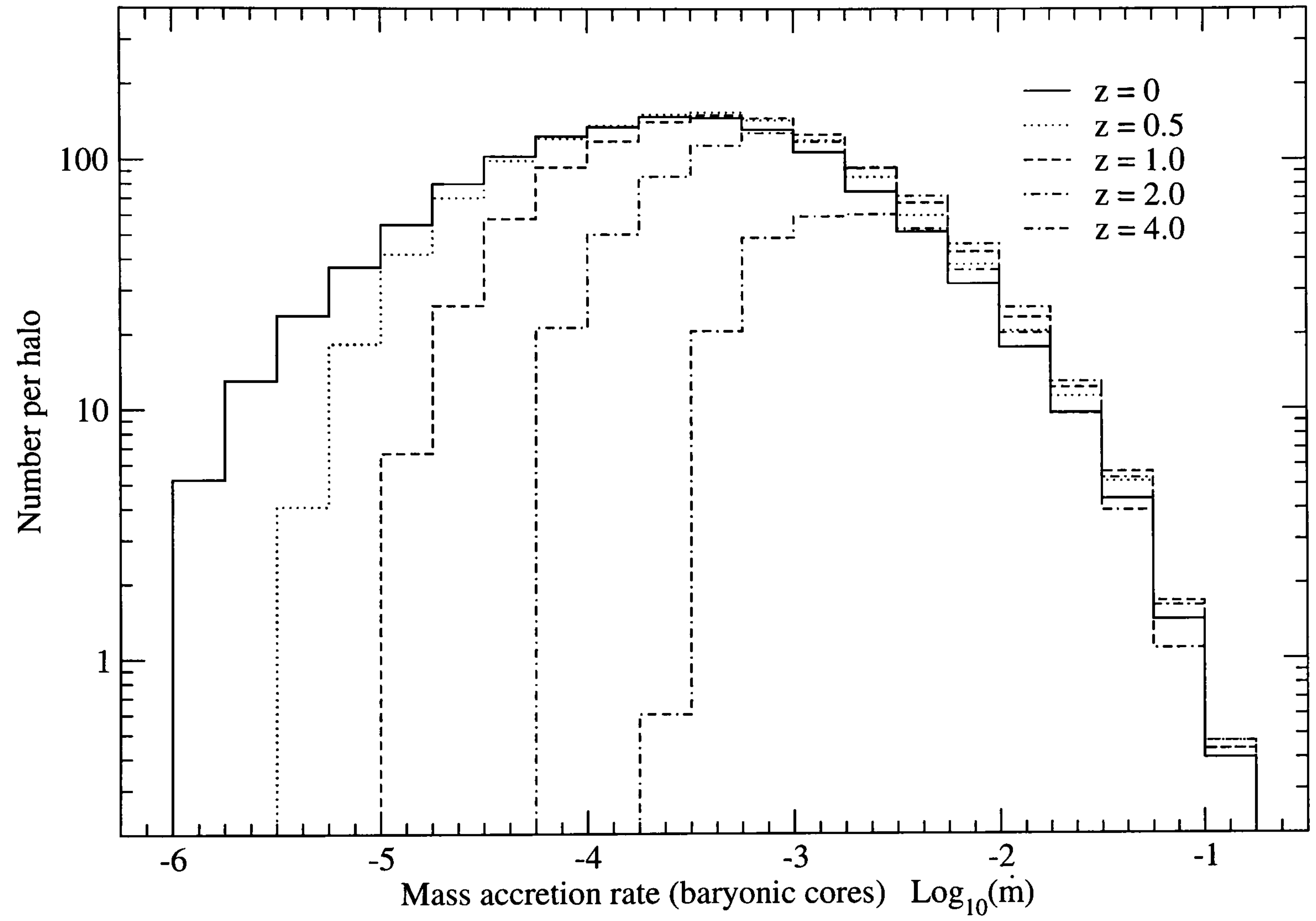


Figure 6.6: Redshift dependence of baryonic core accretion in model A and a final halo mass of $1.6 \times 10^{12} M_{\odot}$. At late times the host virial radius becomes largest. The resulting inclusion of low mass halos with correspondingly small MBHs leads to an increase of objects with low accretion rates. (This particular plot is based on a different set of simulations which does, however, use the same key cosmological parameters.)

For model A we show in figure 6.6 the average distribution of mass accretion rates of MBHs for various redshifts in a halo that grows to mass 1.6×10^{12} $M_{\odot}$at $z = 0$. We see that for increasing redshift the lower end of the range of accretion rates moves towards higher values. The reason for this is that at late times, as the host virial radius becomes larger, more satellites are being incorporated in the outer parts of the host. Most of these will be small satellites with seed mass MBHs

It is also interesting to note that the maximum accretion rate in all cases does not exceed a value of more than about 10 % of the Eddington mass accretion rate. Assuming that the maximum accretion rates have never been larger than this it is obvious that gas accretion from baryonic cores for most MBHs does not lead to significant mass increase over a Hubble time. This is shown in figure 6.7 where we have plotted the accretion rates in Eddington units for all final halo masses in models A, C and D. In particular we have marked various fractions of MBHs that accrete at rates such that their mass would increase by a factor 2, 10 and 100 respectively within a Hubble time provided the rate stays constant. Except for the most massive halos not more than about 10 percent of MBHs accretes even enough to double their mass, while in all cases no more than about 20 percent grow in mass by more than a factor of 100. The factor 2 mass increase is of special importance: In the most conservative baryonic core accretion scenario we would expect an MBH to be embedded in a core that contains an amount of material at least of the order of the initial MBH mass. This is to say the MBH holds on to the material within its initial range of influence. If this applied to all MBHs we would have to discount any MBHs accreting at a rate higher than needed to double their mass within a Hubble time - call this $\dot{m}_{2\times}$, since they would have consumed their core by now. We will not undertake a selection of presently accreting MBHs on the basis of this criterion, because accretion rates for most MBHs are clearly below $\dot{m}_{2\times}$. We also assume that a baryonic core is significantly more massive than the mass of the MBH. This is plausible as the baryonic component has condensed at the centre of halos and thus is much less affected by any tidal stripping of the outer parts of the halo which are dominated by dark matter.

In the context of our baryonic core accretion model realistic accretion rates are likely to be significantly lower, especially at low redshift, when a lot of gas has already been used up in star formation - an effect that we have not accounted for in our simulations.

The distribution of data points for the most massive halos in figure 6.7 can be fit by a power law

$$\dot{m} \propto M_\bullet^{0.68\pm0.02} \Rightarrow \dot{M} \propto M_\bullet^{1.68\pm0.02} \tag{6.8}$$

i.e. the most massive MBHs are also those accreting at the highest rates in general. Due to the

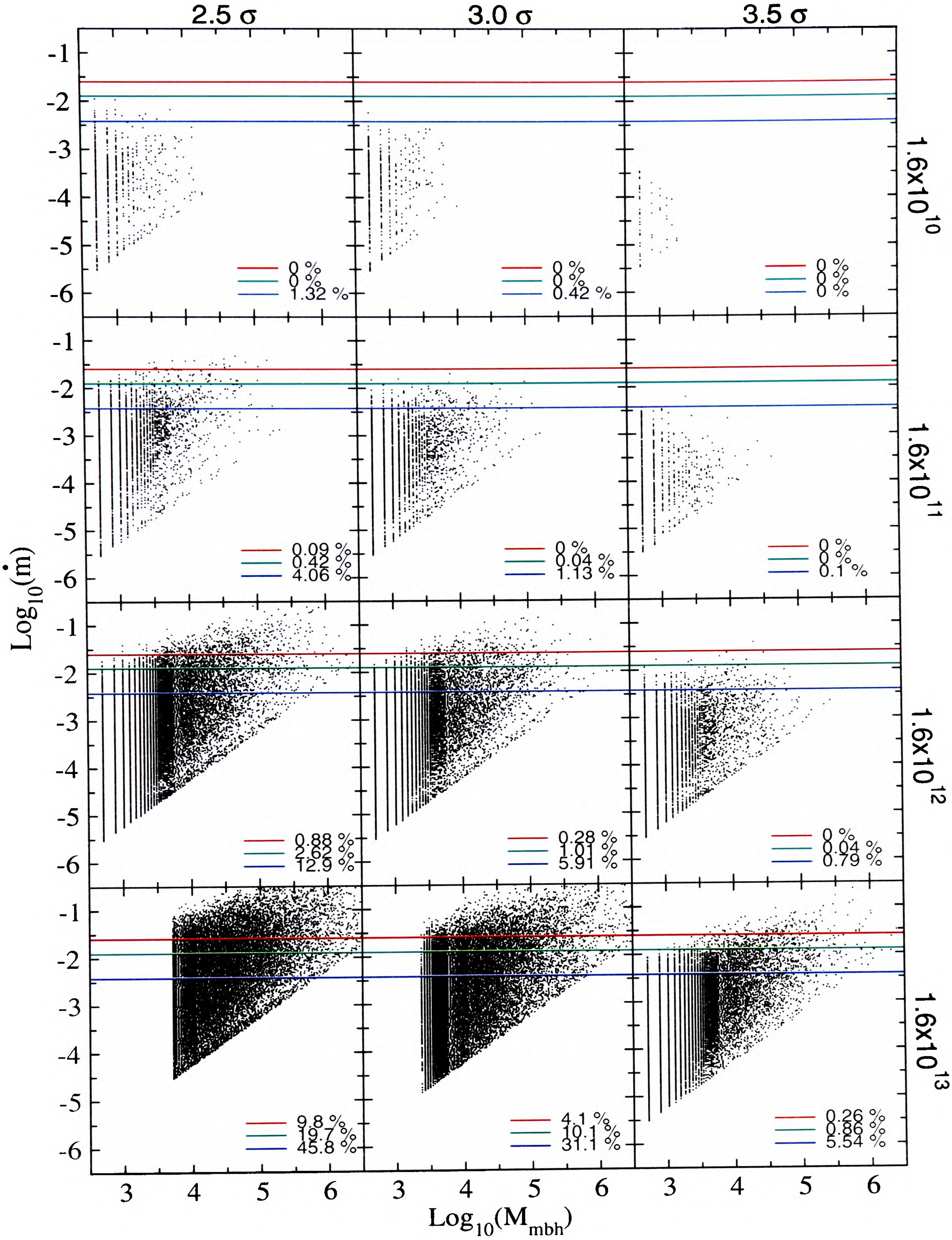


Figure 6.7: Dimensionless baryonic core accretion rates in Eddington units vs. MBH mass for all final halo masses in models C, A and D (from left to right). The results shown are for all 30 trees. The blue, green and red lines denote the accretion rates above which the MBH mass would grow by at least a factor 2, 10 and 100 respectively, within a Hubble time, provided the accretion rate stays constant. The percent figures show the fraction of MBHs with accretion rates higher than the respective cut-offs.

scatter in the plots some of the largest dimensionless accretion rates do actually occur for MBHs that are one or two orders of magnitude lighter than the most massive ones. However, the largest physical accretion rates [2] and thus accretion luminosities are indeed those of the most massive MBHs as we will see in the next chapter.

6.3 Abundance and mass of SMBHs

MBHs that move at distances from the centre less than 1 % of the host virial radius are considered as having fallen to the centre (c.f. section 5.3.2). In line with the assumption of efficient merging the mass of these infalling MBHs is simply added to the mass of a single SMBH growing at the centre.

6.3.1 SMBH from hierarchical merging of remnant MBHs

Figure 6.8 shows the relation between the mass of the galactic bulge and the central SMBH if the latter grows purely through mergers of smaller MBH. The solid line represents the linear relationship between SMBH and bulge mass as determined from observations. To determine this we have used the MBH mass - bulge luminosity relation based on more recent compilation of data (Kormendy & Gebhardt 2001)

$$\left(\frac{M_\bullet}{M_\odot}\right) = 1.24 \times 10^{-3} \left(\frac{L}{L_{B,\odot}}\right)^{1.08} \tag{6.9}$$

where $M_\bullet$ denotes the mass of the SMBH. We have then combined this with the mass to light ratio determined by Magorrian et al. (1998)

$$\left(\frac{M_\bullet}{M_\odot}\right) = 0.33 \pm 0.11 \left(\frac{L}{L_\odot}\right)^{1.18\pm0.03} \tag{6.10}$$

We also assumed that the B band luminosity is approximately the same as the bolometric luminosity $L_{B,\odot} \approx L_\odot$. The observed relation places an upper limit on the allowed masses of seed MBHs and peak heights in the initial density field. The normalisation of the $M_\bullet$ – M_{bulge} mass relation in figure 6.8 is primarily a function of the mass density of MBHs which, in the absence

---

[2] These are not scaled to the MBH mass as is the case for the dimensionless accretion rates.

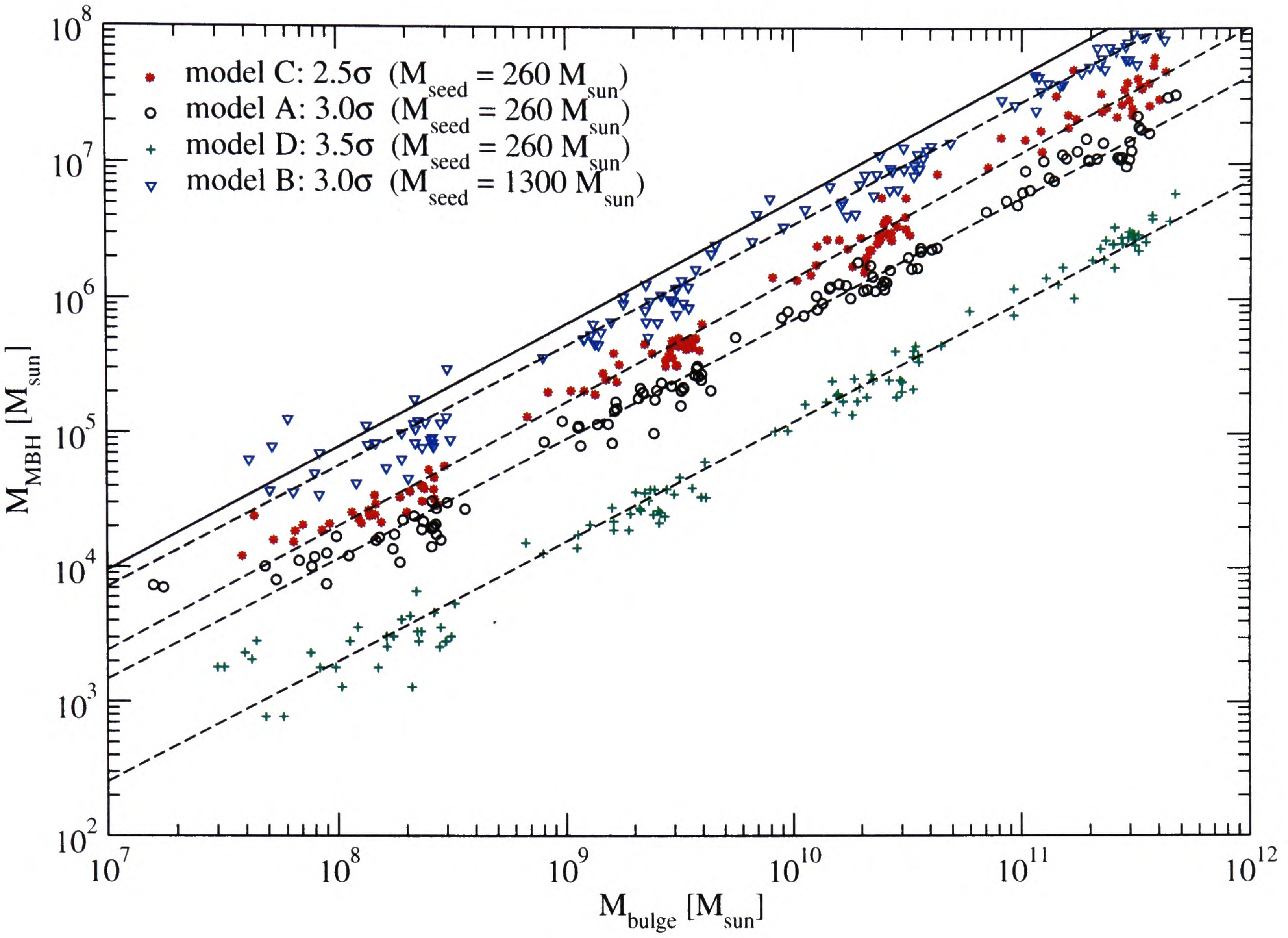


Figure 6.8: Central SMBH masses vs. bulge mass for all seed MBH masses and peak heights considered (models A – D). For each combination the data for all corresponding final halo masses have been grouped together. Dashed lines represent the best fits to the data. The solid line and shaded area are the $M_{SMBH} - M_{bulge}$ relation and associated error from Magorrian et al. 1998.

of accretion, is defined by the number and mass of the seed MBHs in our model. For model B no significant gas accretion is required to match the observed relation and even for models A and C accretion would only be required to increase the SMBH mass by less than a factor of 10. Only in model D (also the fiducial model of VHM03) gas would actually have to raise the SMBH mass by about a factor 100.

The slope of the observed relation, $M_\bullet \propto M_{bulge}^{0.92\pm0.02}$ is very similar to those of the best fits to our data. All the best fits in figure 6.8 have $M_\bullet \propto M_{bulge}^{0.90\pm0.01}$, except for the model with 2.5σ, $M_{\bullet,seed} = 260\ M_\odot$, for which we get $M_\bullet \propto M_{bulge}^{0.93\pm0.01}$. This agreement is surprising, especially since one would intuitively expect a simple linear relation between $M_\bullet$ and M_{bulge} in our merger only scenario. On the other hand, this close agreement raises the question of whether the observed slope is primarily a relic of the hierarchical merger process rather than the result of accretion processes. We speculate that a logarithmic slope less than 1 in the merger induced relation may be due to the evolution of the mass concentration in halos with redshift. At higher redshifts concentrations are typically higher, i.e. the inner scale radii of halos are typically smaller relative to the halo's virial radius. The higher concentration of mass at the centre could thus lead to relatively more efficient merging of MBHs present there and thus to a somewhat larger ratio between MBH mass and that of the bulge.

To the extent that gas accretion occurs it does not alter the slope, i.e. all SMBHs grow through gas accretion by the same factor. At first sight this would seem to be in contradiction to the result for halo MBHs that more massive MBHs systematically accrete at higher rates (c.f. figure 6.7). We can resolve this in the same way as for halo MBHs: At low redshifts a significant if not the largest amount of accretable gas may have already been consumed in star formation.

For the SMBHs our results are also consistent with that of VHM03. For model A, for instance, we found the mass of a central SMBH in a Milky-Way sized halo to be 1.5×10^6 $M_\odot$. Accounting for the difference in seed MBH masses used this agrees with the central SMBH mass of $\sim 1 \times 10^6$ $M_\odot$for a halo of mass $1.6 \times 10^{12} M_\odot$with $\sigma \sim 155$ kms^{-1} as implied by their $M_\bullet - \sigma$ relation (with no gas accretion). This also coincides with the mass determined for the SMBH in the Milky Way, although the Milky Way SMBH is known to lie significantly below the observed $M_\bullet - \sigma$ relation.

However, our slightly non-linear $M_\bullet - M_{bulge}$ correlation corresponds to a $M_\bullet - \sigma$ relation whose logarithmic slope ($\sim$ 4.0) does not match the much flatter one they determined ($\sim$ 2.9) for 3σ collapse and no gas accretion. We believe this to be possibly a result of the different assumptions made about the MBH merger process. In particular the inclusion of triple BH interactions and sling-shot ejections, that they find, would probably lead to even lower central SMBH masses in our analysis.

6.3.2 SMBH growth from gas accretion

A number of studies (Soltan 1982;Yu & Tremaine 2002) suggest that the present day SMBH mass density is consistent with the amount of gas accreted during the optically bright QSO phase. If this is the case then our model B is probably ruled out as it only requires a mass increase of order unity to match the observed abundance of SMBHs today. Models A, C and D would not be affected as gas accretion would still be needed to increase the SMBH masses by at least a factor 3.

On the other hand gas accretion (during the QSO phase) alone cannot explain growth from a stellar mass BHs to the most massive SMBHs ($\geq 10^9$ $M_\odot$). Even if stellar mass BHs are accreting at the Eddington limit, the QSO phase would not last long enough for the BHs to grow sufficiently.

If BHs accrete at the Eddington limit their mass doubling time is given by the Salpeter time scale, t_{salp} (equation 3.6) The quasar epoch lasts about 10^9 years from a redshift of about $z \sim 3.5 - 1.5$ (Richstone et al. 1998). Typical QSO lifetimes are estimated to be $t_Q \sim 10^7 - 10^8$ yrs. Even if we assume that a series of major mergers has triggered individual QSOs repeatedly such that they are active through most of the QSO epoch, the BHs would only grow by a factor of order 10^6 but likely less.

Semi-analytical models of galaxy formation assume that following the QSO epoch the (S)MBHs powering quasars grew further through major mergers (Haehnelt & Kauffmann 2000). However, the latter is unlikely to have raised the (S)MBH mass by more than an order of magnitude. This means that MBHs with masses of at least $10^2 - 10^3$ $M_\odot$must have been present at the beginning of the QSO epoch already. This is confirmed by the presence of luminous quasars at redshifts higher than $z \sim 6$. In fact, the latter means that at least some SMBHs must have been in place by that time as the Universe was hardly old enough to accommodate a long enough Eddington limited gas accretion phase to explain their mass (Haiman & Loeb 2001).

Our model complements this idea. Hierarchical merging involving seed MBHs originating at $z \gtrsim 20$ can produce the number and masses of (S)MBHs needed at the onset of the QSO epoch. To test this idea we estimate the density of SMBHs with a mass of at least 10^8 $M_\odot$- which we take as the mass required to power a QSO - at a redshift $z \sim 6$. Since only some fraction of QSOs are active, the number density of SMBHs has to be at least as large as the density of luminous QSOs at that redshift, which is $\sim 10^{-7}\mathrm{Mpc}^{-3}$ (Richstone et al. 1998). This number density corresponds to the abundance of halos of mass 6×10^{12} $M_\odot$at redshift $z = 6$ in the ΛCDM model we are working with. If we assume that the $M_\bullet - M_{bulge}$ relation in figure 6.8 and the $M_\bullet - M_{host}$ relation in table 6.2 will be the same at $z = 6$ then there is a problem: For a halo mass of 6×10^{12}

$M_\odot$models A and C imply a SMBH mass of 5×10^6 to 1×10^7 $M_\odot$. However, if we allow for gas accretion at only a fifth of the Eddington rate these SMBHs could have certainly grown by an order of magnitude by $z = 6$. The conclusion is then that Eddington limited gas accretion alone onto an initially stellar mass BH or even a moderate mass MBH cannot even remotely produce the number density of SMBHs required to explain QSOs at redshifts of $z \sim 6$.

The need for intermediate mass seed BHs and/or some merging of MBHs/SMBHs is therefore necessary to explain the presence of the most massive SMBHs.

6.4 Summary

Our results from this chapter can be summarised as follows:

- For a range of models that consider different seed MBH masses as well as various collapse thresholds and redshifts for MBC halos we have determined the abundance of remnant MBHs in present day galactic halos.
- We have considered two accretion scenarios for the MBHs. For the case that MBHs accrete from the host ISM we found that the resulting accretion rates are too small for any related observational signatures to be significant. Instead accretion from baryonic core remnants of the satellite halos that MBHs are/were originally associated with yields much larger accretion rates.
- The slope of the $M_{SMBH} - M_{bulge}$ relationship in our model almost exactly matches that of observations. Depending on the model considered, however, various amounts of gas accretion are required to also match the normalisation of the observed relation.
- Our model complements gas accretion based growth models for MBHs. It produces an appropriate number of MBHs and SMBH at high redshifts without which gas accretion alone could not explain the most massive SMBH today as well as the presence of QSOs at redshifts of 6.

In contrast to previous, mostly analytical work, our results are based on more accurate modelling of the dynamics of halo substructure, with the exception of the MBH merger process, that we have assumed is completely efficient due to a number of mechanisms which are, however, as yet not well constrained. The presence of MBHs, their mass density as well as their overall number in individual present-day halos are among the most robust predictions of our model as they are independent of the merger process which only affects the slope of the MBH mass function inside halos. The present-day MBH mass density is entirely determined by the mass density of seed MBHs at high redshifts, which in turn depends on the number density of collapsing MBC halos, their collapse redshift as well as the IMF of pop III stars and their seed MBH remnants.

Predictions of baryonic core accretion rates are not well constrained. We have assumed that MBHs accrete from gas in a baryonic core that is assumed to have a mass of the order of the total baryonic mass in the halo and that the gas is distributed at constant density out to about 10 % of the halo virial radius. An improvement to this calculation would be to assume that the gas distribution has cooled to settle into an isothermal sphere, implying a much higher density at the centre and consequently larger accretion rates near or above the Eddington limit. This could plausibly lead to jets. On the other hand feedback from such a strongly accreting MBH could blow out and heat up the gas in the baryonic core such that an equilibrium accretion rate could be established in much the same way as has been proposed for SMBHs in galaxies (Silk & Rees 1998). If MBH accretion rates were generally much larger in the past, then many or most MBHs would not accrete anymore at present, since their baryonic core would be exhausted. However, as a consequence MBHs would be much more massive, possibly by one or two orders of magnitude - depending on the original size of the core. This vastly increased mass of MBHs would also strengthen the assumption of efficient merging.

The close agreement in slope between the observed $M_{\bullet} - M_{bulge}$ relation and the one based on our merger-only model, raises the question of whether the observed slope is primarily a relic of the hierarchical merger process rather than the result of accretion processes. A slope different from 1, has often been seen as the hallmark of gas accretion induced feedback processes. Our result then shows that merging not only allows for a non-linear relationship to a statistically significant degree, but that it can even match the observed slope.

Chapter 7

Observational Signatures I: Present Day Abundance of MBHs

Previously we have investigated the abundance of MBHs in present day galaxies as a result of this process of hierarchical merging and dynamical evolution (ITS02). We now look at how these MBHs could be observed and thus the model be tested. As the MBHs travel on orbits outside the galactic centre they will emit in the optical, UV and X-ray part of the spectrum due to accretion of matter in their direct vicinity. This matter can be either the inter stellar medium (ISM) of the host or remnant material of the original satellite that remains bound or otherwise associated with any given MBH. While optical and UV emissions could in principle arise from a variety of sources and be subject to a range of absorption and distortion mechanisms, X-ray emissions are more uniquely associated with accreting black holes. In addition hard X-rays are much less subject to absorption. Observations of luminous and ultra luminous X-ray point sources associated with galactic halos in the local Universe could therefore provide constraints on the existence and abundance of a population of MBHs. Optical and UV observations could then be used as a follow up to confirm the case for detection of a MBH.

After a few beginning remarks in section 7.1 we first determine the bolometric luminosity in section 7.2. We go on to discuss two principal types of accretion processes in section 7.3 which we then apply to determine the expected optical and X-ray luminosities in sections 7.4 and 7.5 respectively. We briefly look at the expected optical and X-ray signatures from MBHs in Earth-centred volumes in section 7.6. Finally we compare our results with observations in section 7.7.

We conclude with a summary and a discussion of the wider implications of MBH detections in section 7.8

7.1 Optical and X-ray emissions from accreting MBHs

The steady state mass accretion rate of a black hole given by the Bondi-Hoyle formula above is straightforward because it only depends on gravitational processes specified by a small number of basic parameters that we know in principle how to determine. When it comes to determining how a black hole emits radiation as a result of accreting matter the situation gets significantly more difficult. The efficiency parameter η introduced above only provides an upper limit on how much of the accreted matter is radiated away but essentially encapsulates our ignorance of how the accretion process and subsequent emission work in detail. So far there is no model to comprehensively describe the structure of accretion flows onto black holes as the role of non-gravitational physics in this context is largely unknown. Most models for the accretion flows make assumptions about e.g. the role of magnetic fields, the nature of the viscosity, temperature of electrons, relativistic effects etc., that are difficult to test observationally. In most cases models are primarily testable by comparing their predicted emission spectra to observations.

In choosing particular models to predict the emission spectra in our case we are faced with the additional difficulty that most models were formulated with a view to explaining the accretion processes involving stellar mass BHs, e.g. accretion from a stellar companion, or SMBHs in galactic centres. Our MBHs, however, are both in a different mass range and live in/ accrete from an environment that is different from the ones of stellar mass and super massive BHs respectively.

7.2 Bolometric luminosity

The bolometric luminosity L_{bol} is directly proportional to the physical mass accretion rate times the radiative efficiency parameter η

$$\begin{aligned} L_{bol} &= \eta \frac{dM}{dt} c^2 \\ &= \eta \, 5 \times 10^{35} \left(\frac{M_\bullet}{100 M_\odot} \right)^2 \left(\frac{\rho_g}{10^{-24} \text{ g cm}^{-3}} \right) \\ & \times \left(\frac{c_s}{10 \text{ km s}^{-1}} \right)^{-3} (1 + \beta_s^2)^{-3/2} \text{ erg s}^{-1} \end{aligned} \tag{7.1}$$

and c is the speed of light. η can reach maximum values of $\eta_{max} \sim 0.06$ for non-rotating black holes and up to $\eta_{max} \sim 0.4$ for maximally rotating black holes (Shapiro & Teukolsky 1983), if a mechanism for the effective dissipation of energy exists. This mechanism is provided for by the viscosity of matter in the accretion flow.

7.2.1 Bolometric luminosity for accretion from the host ISM

The effect of ISM accretion and the resulting emission has previously been investigated analytically for stellar mass BHs in our galaxy that are the remnants of ordinary stellar evolution (Fujita et al. 1998). For MBHs larger luminosities are expected, particularly when they travel within or are crossing the galactic disk and bulge regions or in molecular clouds, as more gas is available. But even in less dense regions of the inter stellar medium (ISM) MBHs could generate sizable luminosities if they travel at low velocities and have a correspondingly larger accretion radius.

In particular ISM turbulence can establish a geometrically thin accretion disk around the MBH with an associated radiative efficiency of $\eta_{td} = 0.1$, which is of the order of the maximum radiative efficiency $\eta_{max} \sim 0.06$ expected on theoretical grounds. We will return to this in section 7.3.2. For our computations we have adopted a 'standard' thin disk radiative efficiency of $\eta_{td} = 0.1$, and used the ISM accretion rates determined in the last chapter.

The resulting bolometric luminosity function for the MBHs in individual halos is shown in the top panel of figure 7.1. This also shows the contribution of MBHs above various mass thresholds.

In the last chapter we have seen that ISM accretion occurs at a very much smaller rate than that from baryonic cores and the question is how low the resulting accretion luminosity

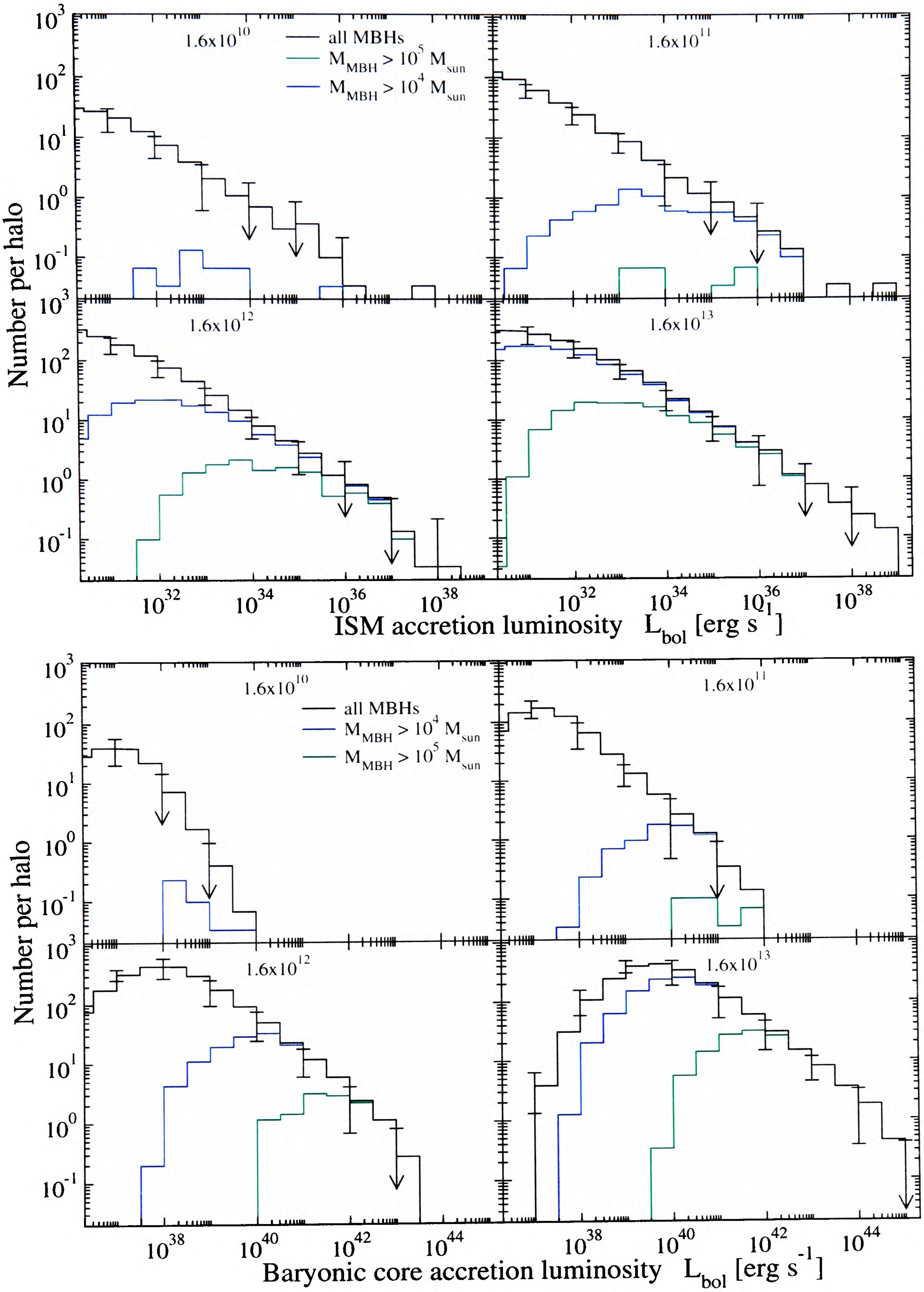


Figure 7.1: Bolometric luminosity functions for model C in the case of accretion from the host ISM (top panel) and from baryonic cores (bottom panel). Results are shown for all final halo masses and MBH mass cuts as shown.

would be. In fact from figure 7.1 it becomes clear that MBHs accreting from the host ISM have luminosities so low that it appears very difficult to detect them let alone identify them as MBHs. We have therefore not attempted to determine any spectral energy distribution or band luminosity functions for this case. It might still be possible to derive statistical constraints were the number of objects large enough [1]. For instance in an inhomogeneous ISM, that we have not accounted for in our model, BHs travelling through dense regions in the galactic disk or bulge would emit at significantly larger luminosities. A MBH accreting within a dense cloud that has a density a factor 10 higher than the average, say, would see its luminosity boosted by about the same factor (c.f. eq. 7.1). By conservation of mass, however, these clouds would only fill $\lesssim 1/10$ of the ISM volume. I.e. if we have 10 MBHs uniformly distributed across the host ISM and accreting at 10^{36}erg s^{-1} only at most one would end up in a cloud with ten times larger density at any one time. However, in our case the numbers of MBHs are just too low for this process to be significant.

Figure 7.1 shows the result for model C, which serves to illustrate that even for the model that produces the highest number of large MBHs in the most massive halos, the number of MBHs accreting at bolometric luminosities larger than $\sim 10^{37}$erg s^{-1} is insignificant. The main cause for these very low ISM accretion rates is that most MBHs orbit at distances larger than the light radius of the galaxy, which means that even if accretion luminosities were very large in the disk and bulge we would still only observe very few sources there.

Regarding this last point baryonic core accretion also offers the advantage that it is essentially independent of the structure and geometry of the host ISM.

To determine actual observational signatures we will from now on focus exclusively on baryonic core accretion.

7.2.2 Bolometric luminosity for accretion from baryonic core remnants

In the previous section we argued that MBHs accrete from a disk assuming that the net specific angular momentum created in the surrounding host ISM through turbulence is large enough.

For accretion from baryonic cores the required net angular momentum comes for the largest

---

[1] See e.g. Fujita et al. (1998) who consider of the order of 10^6 BHs kpc $^{-3}$ (Fujita et al. 1998)

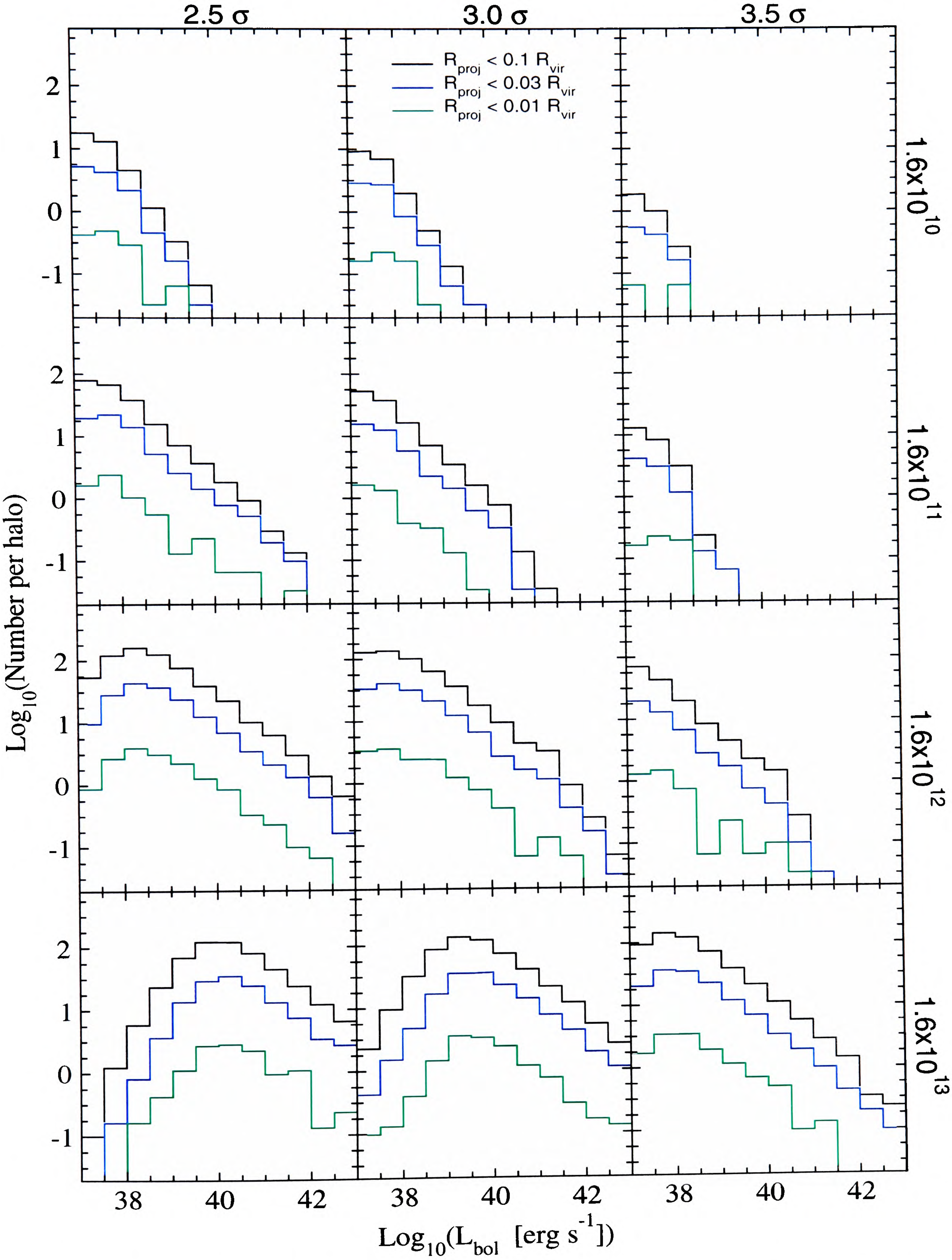


Figure 7.2: Bolometric luminosity of MBHs accreting from baryonic cores for models C, A and D (left to right panels). Shown are the sources whose line of sight falls within some projected distance from the host centre.

part from the angular momentum of the original satellite. Although the outer parts of satellites may have been stripped and their angular momentum deposited in the host halo, the cores will preserve most of their angular momentum. In this sense we can view these MBH - baryonic core systems as the engines of mini-AGN stripped of their halo/galaxy within which they originally resided.

Based on the baryonic mass accretion rates obtained in the previous chapter and using the 10 % standard thin disk efficieny, we obtain a differential MBH bolometric luminosity function which is shown in the bottom panel of figure 7.1 for model C. $\beta_s = 0$ in this case as the MBHs have no relative motion with respect to the baryonic cores from which they accrete. Figure 7.2 shows what we call the projected bolometric luminosity function, that is of all sources within a halo, whose lines-of-sight (LOS) fall within some distance from the centre normal to the LOS. Overall we see that the accretion luminosities are much higher than for the ISM accretion case. They attain values larger than 10^{40}erg s^{-1}, although luminosities will be somewhat lower for models A and D with their lower MBH mass density.

The luminosity functions for the two most massive halos exhibit a logarithmic slope in the declining part of the function.

$$N_\bullet \propto L_{bol}^{-0.6\pm0.02} \tag{7.2}$$

Also from equation 6.8

$$L_{bol} \propto \dot{M} \propto M_\bullet^{1.68\pm0.02} \tag{7.3}$$

The plot also shows that the distribution is very similar for MBHs in halos of mass 1.6×10^{12} and 1.6×10^{13} M$_\odot$, with the slope and the offset of the two identical to within the statistical errors. This is no surprise given that the accretion luminosity is independent of the host characteristics but depends on the properties of the baryonic cores which derive from the same high redshift distribution. A similar slope would therefore also be expected for the two lighter host halo masses, if there were enough larger MBHs to carry out a reasonable best fit. Comparing with equation 7.1 this implies a non-trivial scaling relation between $M_\bullet$ and the satellite gas density

$$M_\bullet \propto \rho_g^{3.13\pm0.2} \tag{7.4}$$

Across all luminosities the largest fraction of the sources is in the host. Many of the brightest ones are in fact at distances larger than 10 percent of the host virial radius R_{vir}. The numbers of sources in the bulge and the disk are similar and relatively low. Inside our galaxy most of the sources should therefore not be affected by strong absorption in the disk and bulge. For other galaxies, however, these problems remain for those MBHs whose line-of-sight passes close to the central region of the galaxy within which they are located.

In the following we take a more detailed look at the accretion geometry and models for the accretion flow with the aim of determining the emission spectrum and band luminosity functions for MBHs accreting from baryonic cores.

7.3 Accretion processes

7.3.1 Accretion geometries: spherical vs. disk accretion

The actual value of the radiative efficiency, η, depends on the mode of accretion. For spherical accretion onto a black hole this is usually substantially lower than for disk accretion. If the gas surrounding the black hole has sufficient net angular momentum about the black hole then an accretion disk may form. An approximate measure for when this occurs is the specific angular momentum,L, of the gas, h around the black hole. If this is equal or larger than the angular momentum of the last stable orbit of the black hole a disk may form

$$L > 2\sqrt{3}\frac{GM_\bullet}{c} \tag{7.5}$$

and the radial extent of the disk, r_{td}, in units of the Schwarzschild radius $r_\bullet$ is (Ipser & Price 1977)

$$r_{td} = \frac{L^2\, c^2}{G^2\, M_\bullet^2} \tag{7.6}$$

In principle the required angular momentum could come from turbulence in the ISM. Larson (1981) finds that the measured turbulence velocity v_{turb} is correlated to the scale r, although with large uncertainties, as

$$v_{turb} \sim 1.1 \left(\frac{r}{1\mathrm{pc}}\right)^{0.38} \ \mathrm{kms}^{-1} \tag{7.7}$$

on scales $0.1 < r < 100$ pc. This implies an angular momentum of the order

$$L \sim v_{turb} r_{acc} = 1.1 r_{acc}^{1.38} \text{km s}^{-1} \text{ pc} \tag{7.8}$$

Using this in equation 7.6 we obtain

$$M_\bullet \gtrsim 240 \left(\frac{r_{td}}{1000}\right)^{1.32} \left(\frac{c_s}{50 \text{ kms}^{-1}} (1+\beta_s^2)\right)^{7.26} \tag{7.9}$$

I.e. for an MBH of mass 260 $M_\odot$travelling at a velocity of 50 km s^{-1} or less, ISM turbulence can establish an accretion disk that reaches out to more than 1000 Schwarzschild radii.

It appears that the condition for the setting up of an accretion disk is satisfied for the range of black hole masses and velocities we are considering. In what follows we will therefore assume that an accretion disk always forms and that purely spherical accretion can be neglected. We note that this only refers to the 'global' accretion geometry on scales of the order of the accretion radius. It does not, however, exclude the possibility of the accretion flow becoming spherical in the vicinity of the MBH's last stable orbit.

7.3.2 Accretion mechanisms I: Thin disk accretion

Shakura & Sunyaev (1973) proposed a model in which BHs accrete from a geometrically thin but optically thick accretion disk, which is considered to achieve the highest radiative efficiencies, typically with $\eta_{td} \sim 0.1$. Because it is optically thick, radiation is emitted locally with a thermal spectrum and black body temperature

$$T(r) = \left(\frac{3GM_\bullet \dot{M}}{8\pi r^3 \sigma}\right)^{1/4} \tag{7.10}$$

where $\dot{M}$ and r are the mass accretion rate and radial distance from the MBH respectively. We can convert these into dimensionless quantities and express the accretion rate in units of the Eddington accretion rate equation (c.f. equation 6.7) and the radial distance from the MBH in units of the MBH Schwarzschild radius, $r_\bullet$

$$\tilde{r} \equiv \frac{r}{r_\bullet} = \frac{r\, c^2}{2\, G\, M_\bullet} = 0.339 \left(\frac{r}{\text{km}}\right) \left(\frac{M_\bullet}{M_\odot}\right)^{-1} \tag{7.11}$$

The corresponding spectrum integrated over the whole disk is given by (Frank, King & Raine 1985)

$$\frac{\mathrm{d}L_\nu}{\mathrm{d}\nu} = A_0 \frac{h\nu^3}{c^2} \int_{r_{lso}}^{r_{td}} \frac{\tilde{r}\mathrm{d}\tilde{r}}{e^{h\nu/kT(\tilde{r})} - 1} \tag{7.12}$$

where $r_{lso} = 3\, r_\bullet$ is the last stable orbit radius of the MBH and $A_0 = 4\,\pi$ is the area of a spherical shell at unit distance from the MBH. However, in further course we treat A_0 more generally as a normalisation constant. All our computations were carried out with $r_{td} = 1000 r_\bullet$ and A_0 was consequently renormalised by setting $\int(\,\mathrm{d}L_\nu/\,\mathrm{d}\nu)\mathrm{d}\nu = L_{bol}$. The resulting spectra are shown in the panels on the right in figure 7.3.

In general the accretion flow near the BH may become optically thin and absorption of photons becomes less important than scattering, which results in an increase of the mean energy of the photons radiated away. The emergent flux is thus non-thermal and has a higher temperature. Consequently in the inner parts of the disk near the MBH, equations 7.10 and 7.12 are not strictly valid anymore. Another prescription to account for the higher frequency photons is needed. This is also desirable from another point of view. MBHs are only clearly identified as BHs via an X-ray signature. However, from equation 7.10 the peak frequency of the photons from a standard thin disk scales as $\nu_{max} \propto T(r_{lso}) \propto M_\bullet^{1/4}$. For the MBHs any resultant flux in the X-ray range is only significant at energies below about 1 keV. A standard thin disk alone would therefore not provide the X-ray flux by which to identify MBHs.

One could tackle this problem phenomenologically by adding a power-law spectrum extending into the X-ray range. A power law X-ray spectrum is well established for AGN that are driven by SMBHs and is also found in recent observations of ultra luminous galaxian off-centre X-ray sources that could well be accreting MBHs (Colbert & Mushotzky 1999; Wang 2002). However, instead we consider a different model that has recently been developed particularly with a view to explaining the emission spectra from accreting SMBH in low-luminosity AGN. This is an advection dominated accretion flow.

7.3.3 Accretion mechanisms II: Advection dominated accretion flows (ADAFs)

In the ADAF model, once the accretion rate drops below some critical rate $\dot{m}_{crit}$, which we define below, accretion switches to an 'advection dominated accretion flow' (ADAF) (Ichimaru

1977; Rees et al. 1982; Narayan & Yi 1994; Manmoto, Mineshige & Kusunose 1997). Because of the low accretion rates the density of the accreted gas is low, too. The gas not being able to cool efficiently stores thermal energy instead of radiating it away and carries it into the BH. On the other hand, because the accretion flow is optically thin the spectrum of the emergent radiation is modified by scattering processes. The main processes are synchrotron emission, bremsstrahlung and comptonisation. For more details the reader is referred to the review by Narayan, Mahadevan & Quataert (1998) and references therein.

The spectrum emitted from a BH accreting in this way depends on a range of parameters such as the mass accretion rate , the viscosity parameter and the relative contribution of magnetic fields. Typically the spectrum has to be determined numerically, but can be approximated by analytic scaling relations (Mahadevan 1997). Using these, the total radiative efficiency for an ADAF, η_{adaf} can be expressed in terms of that of a thin disk, η_{td} as

$$\eta_{adaf} = 0.2\eta_{td}\left(\frac{\dot{m}}{\alpha^2}\right) g(\theta_e) \tag{7.13}$$

for a typical set of parameters as given in Mahadevan (1997). Here $\dot{m}$ is the mass accretion rate in terms of the Eddington accretion rate, α is the viscosity parameter ($\alpha \sim 0.1$ to 0.5), $\theta_e = k_B\, T_e/(m_e\, c^2)$ is the electron kinetic energy in units of their rest mass energy. $g(\theta_e) \approx 11.5(T_e/10^9\mathrm{K})^{-1.3}$ is a measure of whether electrons in the flow move at relativistic speeds or not. Typically $g \sim 1$ to 10 for electron temperatures $2 \times 10^9 < T_e < 10^{10}$ K. We can determine an equilibrium electron temperature which is approximately

$$\left(\frac{T_e}{10^9\mathrm{K}}\right) \approx \left(\frac{\alpha}{0.3}\right)^{-0.06} \left(\frac{r_{min}}{3}\right)^{0.1} M_\bullet^{0.06}\, \dot{m}^{-0.04} \tag{7.14}$$

Generally the effect of ADAFs is to reduce the overall radiative efficiency compared to that from thin disk accretion, although individual regions of the accompanying spectrum, especially the X-ray band, tend to be much stronger than for a thin disk. Although we did not point this out explicitly at the time, the bolometric luminosities from baryonic core accretion in section 7.2.2 extend to very large values which, if emitted in roughly similar proportions across the visible to X-ray bands, are hard to reconcile with observations. The reduced radiative efficiency of ADAFs could therefore also help solve this problem.

As mentioned above the thin disk and ADAF regime can be delineated by the critical mass accretion rate $\dot{m}_{crit}$ which is given by

$$\dot{m}_{crit} \geq 0.28\alpha^2 \tag{7.15}$$

For rates above $\dot{m}_{crit}$ the ADAF mechanism breaks down and in this regime we will work with the thin disk model. For low values of $\alpha \leq 0.1$ accretion may switch to a 'convection dominated accretion flow' (CDAF) (Ball, Narayan & Quataert 2001), which is characterised by much stronger suppression of mass accretion. We will neglect this possibility here as we work with $\alpha = 0.3$ in the computation of spectra and luminosity functions for our model. This is also the fiducial value used in the work of Mahadevan (1997). We will briefly consider changes in α below.

At the end of the previous chapter we have seen that MBHs do not accrete at more than 1/10 of the Eddington rate with most accreting at less than 1/100. I.e. the ADAF model can be applied to most MBHs we are dealing with.

The three dominant contributions to the continuum spectrum of an ADAF are those from synchrotron and bremsstrahlung emission as well as comptonisation, which can all be separately approximated by scaling laws (Mahadevan 1997). In the radio-submillimeter range synchrotron emission dominates and the luminosity (per unit frequency) scales as

$$\frac{\mathrm{d}L_{sync}}{\mathrm{d}\nu} \simeq 3.51 \times 10^{13} \left(\frac{\alpha}{0.3}\right)^{-4/5} \left(\frac{T_e}{10^9\ \mathrm{K}}\right)^{2/5} M_\bullet^{6/5}\ \dot{m}^{6/5}\ \nu^{2/5} \mathrm{erg\ s^{-1}\ Hz^{-1}} \tag{7.16}$$

up to a peak frequency ν_p given by

$$\nu_p = 1.71 \times 10^{17} \left(\frac{\alpha}{0.3}\right)^{-1/2} \left(\frac{T_e}{10^9 \mathrm{K}}\right)^2 M_\bullet^{-1/2} \left(\frac{r_{min}}{3}\right)^{-5/4}\ \dot{m}^{3/4} \mathrm{Hz} \tag{7.17}$$

Photons from synchrotron emission can get comptonised, which raises their energy up to a limit given by the electron temperature $k_B\ T_e$. Since most synchrotron photons are emitted at $\nu = \nu_p$ we can make the simplifying assumption that this is the initial frequency for all photons to be comptonised. The comptonisation spectrum is then simply

$$\frac{\mathrm{d}L_{compt}}{\mathrm{d}\nu} \simeq \frac{\mathrm{d}L_{sync}}{\mathrm{d}\nu}(\nu_p)\left(\frac{\nu}{\nu_p}\right)^{-\alpha_c} \tag{7.18}$$

where the power law index is primarily a function of electron temperature and mass accretion rate

$$\alpha_c \equiv -\frac{\ln\left[23.87\dot{m}\left(\frac{\alpha}{0.3}\right)^{-1}\left(\frac{r_{min}}{3}\right)^{-1/2}\right]}{\ln\left[1+4\ \theta_e+16\theta_e^2\right]} \tag{7.19}$$

There is a constant contribution from bremsstrahlung emission that also tails of exponentially at the maximum frequency implied by the electron temperature. Bremsstrahlung only becomes prominent in the UV/soft Xray range if $\alpha_c > 1$, such that power from Compton emission effectively decreases with increasing frequency

$$\frac{\mathrm{d}L_{brems}}{\mathrm{d}\nu} \simeq 1.02\times10^{17}\left(\frac{\alpha}{0.3}\right)^{-2}\ln\left[\frac{r_{max}}{r_{min}}\right]F(\theta_e)\left(\frac{T_e}{10^9\mathrm{K}}\right)^{-1}e^{-h\nu/k_BT_e}M_{\bullet}\dot{m}^2\mathrm{erg\ s^{-1}\ Hz^{-1}} \tag{7.20}$$

and

$$F(\theta_e) = \begin{cases} 4\left(\frac{2\theta_e}{\pi^3}\right)^{1/2}(1+1.78\theta_e^{1.34})+1.73\theta_e^{3/2}(1+1.1\theta_e+\theta_e^2-1.25\theta_e^{5/2}), & \text{if } \theta_e < 1, \\ \\ \left(\frac{9\theta_e}{2\pi}\right)[\ \ln(1.123\theta_e+0.48)+1.5]+2.3\theta_e(\ \ln1.123\theta_e+1.28), & \text{if } \theta_e > 1 \end{cases} \tag{7.21}$$

The resulting spectrum is shown in the left panels of figure 7.3 for a range of MBH masses and accretion rates ranging from $\dot{m} = \dot{m}_{crit}$ to $\dot{m} = 0.03\ \dot{m}_{crit}$ where $\dot{m}_{crit} = 0.0252$ for $\alpha = 0.3$ which we use for all our computations.

We see that only ADAF accreting MBHs display significant X-ray emission. It is therefore mainly MBHs that could be identified as such by their X-ray signature. Conversely, while optical and UV observations may be used as follow-ups, the sources that are most luminous in this range are accreting from a thin disk and so will not have any associated strong X-ray emission.

X-ray hardness ratios

For our predictions we have computed the X-ray emissions for the energy range between $0.5 < h\ \nu < 2.0$ keV and $2.0 < h\ \nu < 10.0$ keV , which we call the soft and hard X-ray bands respectively.

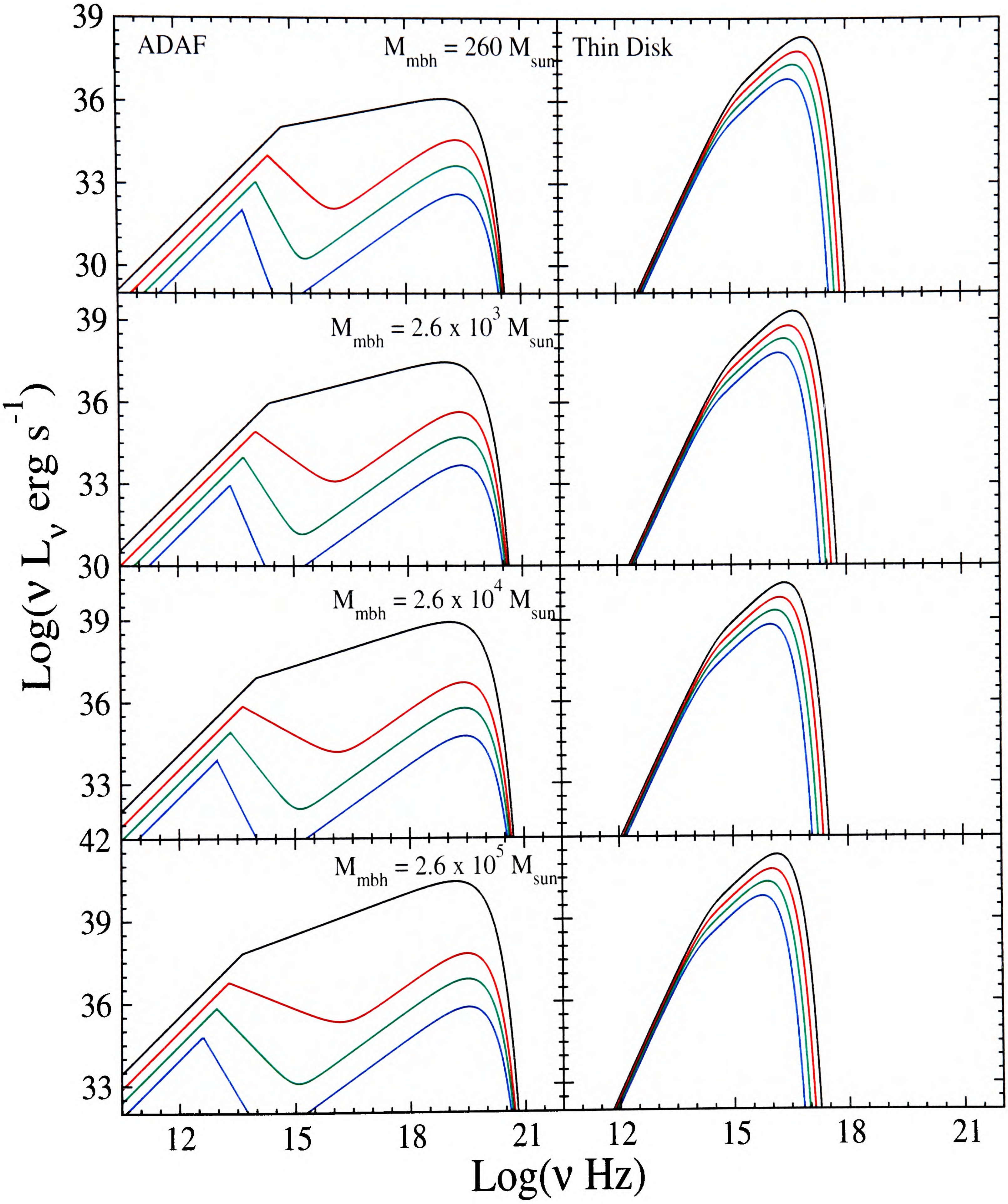


Figure 7.3: Emission spectra from ADAF (left) and thin disk (right) accreting MBHs in the mass range between 2.6×10^2 to 2.6×10^5 $M_\odot$. In each panel the curves from top to bottom correspond to mass accretion rates $\dot{m} = \{1, 0.3, 0.1, 0.03\}\ m_{crit}$ where $m_{crit} = 0.0252$ for $\alpha = 0.3$.

Both bands lie in the Bremsstrahlung part of the ADAF spectrum where $\mathrm{d}L/\ \mathrm{d}\nu = 1$, i.e. the photon counts from both bands are the same; their *hardness ratio* is equal to 1. Unfortunately, observations from XRBs can yield similar values, so it will not be easy to distinguish between MBHs and XRBs on the basis of this alone. The combined thin disk + ADAF model that we used, means that we can also have sources with large soft X-ray luminosities and very small hardness ratios. This is the case particularly for seed MBHs for which thin disk accretion reaches furthest into the soft X-ray regime, and there actually dominates over the respective ADAF contribution (c.f. figure 7.3). In section 7.5 we see that this leads to a distinctive 'bump' in the soft band luminosity function. These sources should be identifiable by their very low hardness ratio.

7.4 Optical emission

We now apply the accretion model outlined above to the present day distribution of halo MBHs as obtained from the semi-analytical code. For the MBH masses we are looking at here the optical region lies within the Rayleigh-Jeans part of the extended black body spectrum that characterises the corresponding thin disk emission. In this part, $k_B\ T(r_{td}) \ll h\ \nu \ll k_B\ T(r_{lso})$ where the spectrum can be approximated by a power law

$$\frac{\mathrm{d}L}{\mathrm{d}\nu} = A_1 \left(\frac{\nu}{\mathrm{Hz}}\right)^{1/3} \tag{7.22}$$

and the constant A_1 depends on the MBH mass and accretion rate as

$$\begin{aligned} A_1 &= 2.57\ A_0\ h^{-5/3}\ k_B^{8/3}\ c^{-2}\ T(r_{lso})^{8/3} \qquad (7.23) \\ &\approx 1.64 \times 10^{11}\ A_0\ \dot{m}^{2/3}\ M_\bullet^{4/3}\ \mathrm{erg\ s^{-1}\ Hz^{-1}} \end{aligned}$$

In the ADAF regime it is primarily synchrotron emission and comptonisation that contribute to the emission in the optical/UV range, with the corresponding scalings given in equations 7.16 and 7.18.

Figure 7.4 shows the B-band luminosity function. The dashed line represents the contribution from ADAF emission with the rest, primarily at large luminosities, coming from thin disk emission. Apart from our fiducial value for the viscosity parameter, $\alpha = 0.3$ we have also given results

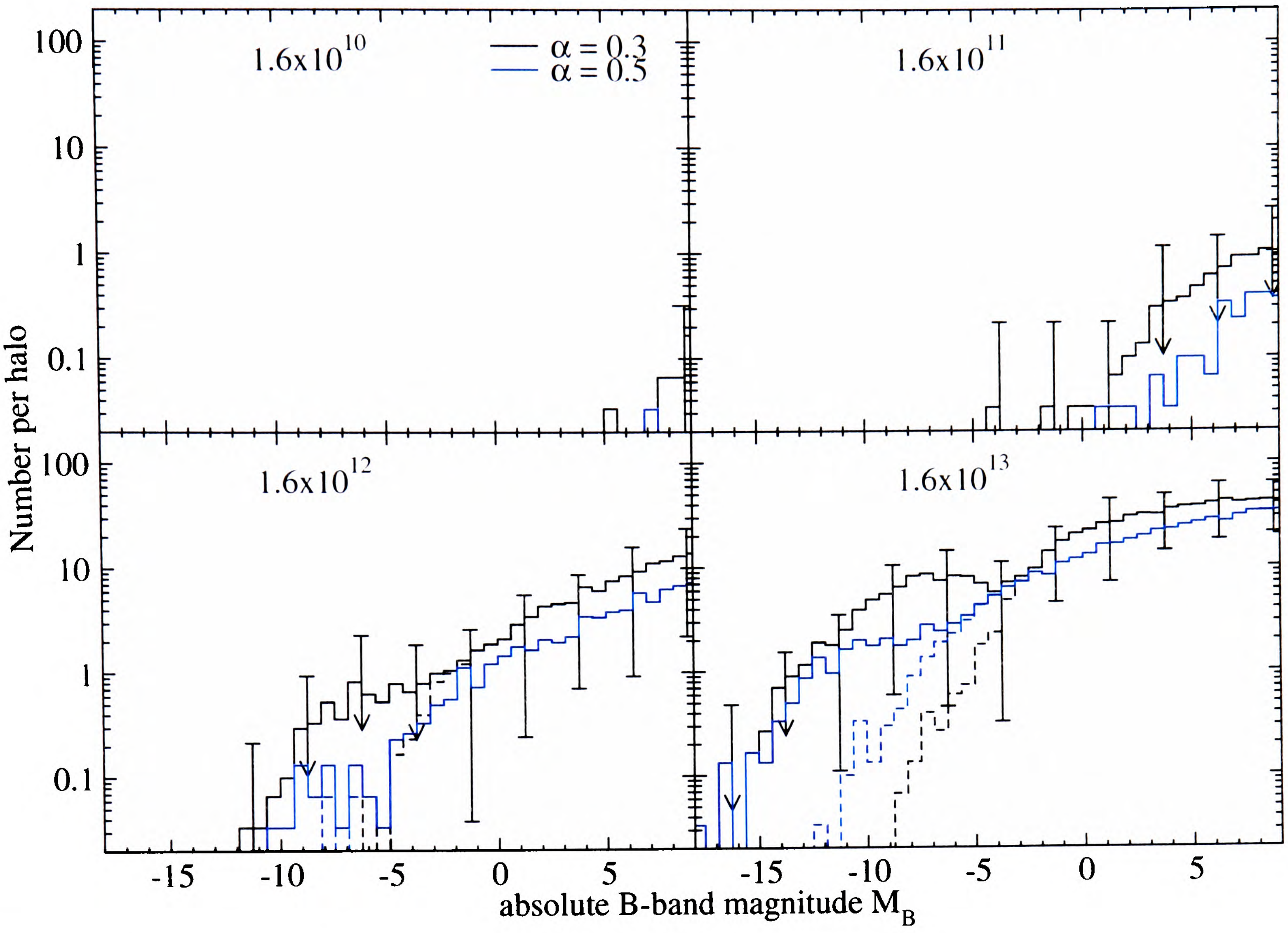


Figure 7.4: B-band luminosity function for MBHs in a $1.6 \times 10^{12} M_\odot$halo. The dashed lines show the contribution from ADAF accretion. Thin disk emission dominates at large luminosities.

for $\alpha = 0.5$ which are shown in blue. While the results for ADAF emission are typically not too sensitive to changes in ADAF parameters, results are affected more strongly by changes in α, but only through its direct impact on the value of $\dot{m}_{crit}$ (c.f. 7.15). As a result the most important effect of increasing α is to extend the number of ADAF emitting sources to larger luminosities. Since a thin disk radiates more efficiently than an ADAF this results in a lower abundance of sources at very large magnitudes in the optical range. For smaller values of α we get a correspondingly larger number of thin disk accreting sources. However, we found that for a value of $\alpha = 0.1$, for instance, the resulting luminosity function in the B-band displays a significant 'dip' which is also somewhat present for $\alpha = 0.3$ in the case of the most massive 1.6×10^{13} $M_\odot$halo in figure 7.4. We will deal with the results for the X-ray range in the next section, but note at this point that a higher (lower) value of α would result in a larger (lower) abundance of sources in the X-ray regime, emission in which is mainly produced in ADAFs.

In figure 7.5 we show the B and V band luminosity function of MBHs within some projected distance from the host centre normal to the line of sight. Here we have included all MBHs within the respective host virial radius along the lines of sight passing within some projected distance from the centre. We will subsequently refer to this as the *projected* luminosity function. The projected luminosity function can more easily be compared with observations, which typically refer to sources within galaxies. In each case the function is plotted for MBHs sources within $R_{proj} = 0.1, 0.03, 0.01\ R_{vir}$ from the host centre, where the latter corresponds roughly to the outer radial extent of the galaxies in the halos. Across all halo masses we find that a fraction of about 30, 9 and 0.9 % respectively of all halo MBHs are located within the projected distances listed above. Since there is no discernible difference in the relative distribution of MBH masses at different radii (c.f. figure 6.3), this means that for any one MBH in a given mass/accretionrate/luminosity range and within $R_{proj} = 0.1 R_{vir}$ there are about two more in the halo.

We have marked with bold black arrows the luminosity range below which ADAF emission dominates.

It is important to note that these luminosity functions only provide an upper limit on what could actually be detected. We have here neglected the effect of absorption which is expected to significantly reduce the flux of optical and UV photons due to dust and neutral hydrogen respectively in the IGM and ISM. The latter affects particularly the light from sources whose line of sight passes closely to the galactic centre, such as those with $R_{proj} < 0.01 R_{vir}$ and located on the far side of the halo. We would expect this effect to be less important in gas-poor ellipticals.

Even if absorption is accounted for, the large luminosity of the accreting MBHs by itself may not be sufficient to identify them as such. As they appear as point sources they could in principle also be distant and/or low-luminosity AGN. For the largest accretion rates, $\dot{m} \sim 0.1$ (c.f. figure 6.7), at which MBHs accrete from a thin disk in our model, it is possible that emission occurs in the form of (mini) jets which may be observable even though the optical and UV output from across the accretion disk may not.

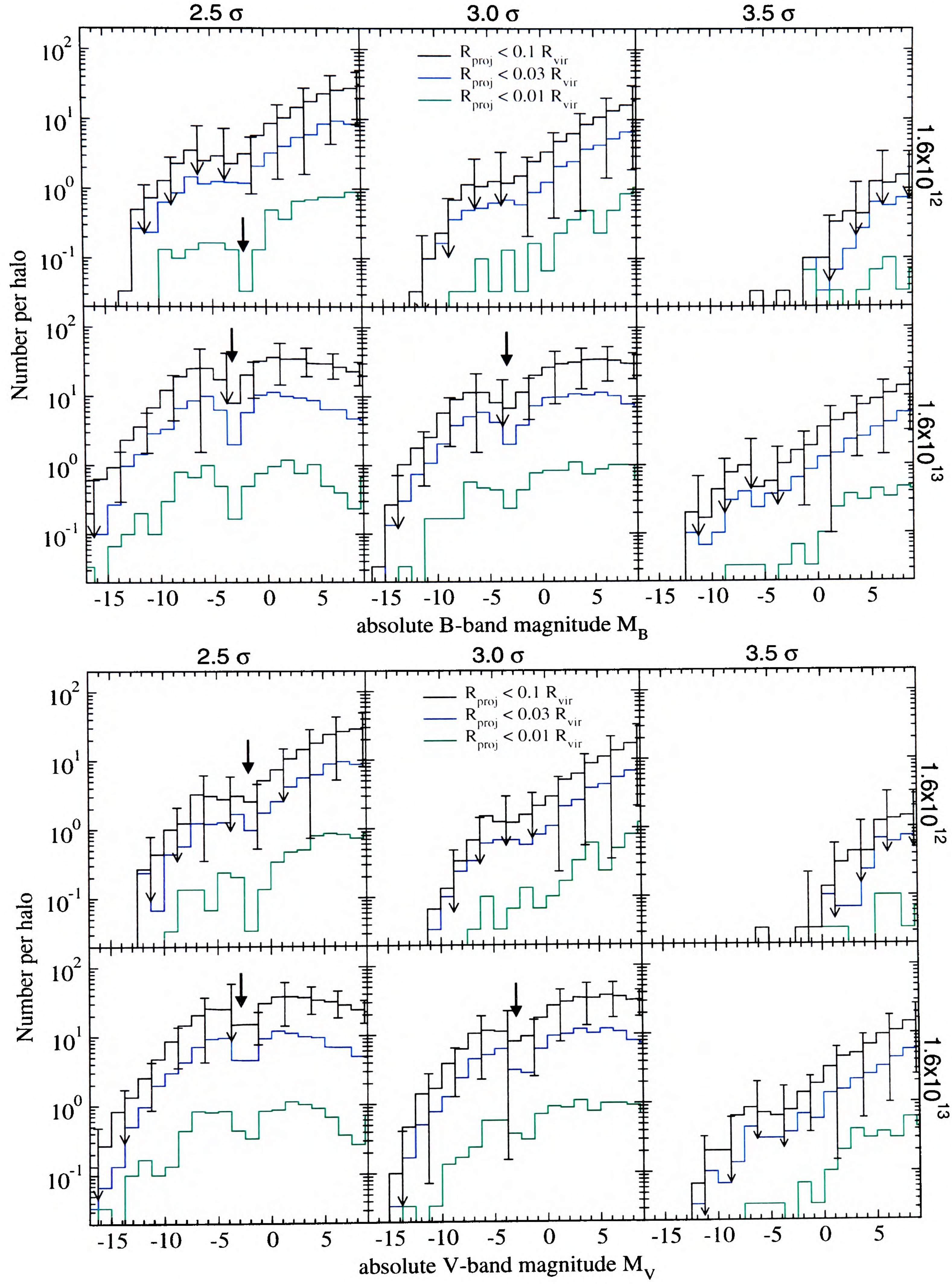


Figure 7.5: Optical luminosity function for MBHs accreting from baryonic cores within various projected distances from the host centre. Results are shown for models C, A and D (left to right panels) in the B and V-band (top and bottom panel respectively). The bold arrows delineate where luminosities are dominated by thin disk emission and ADAF emission respectively.

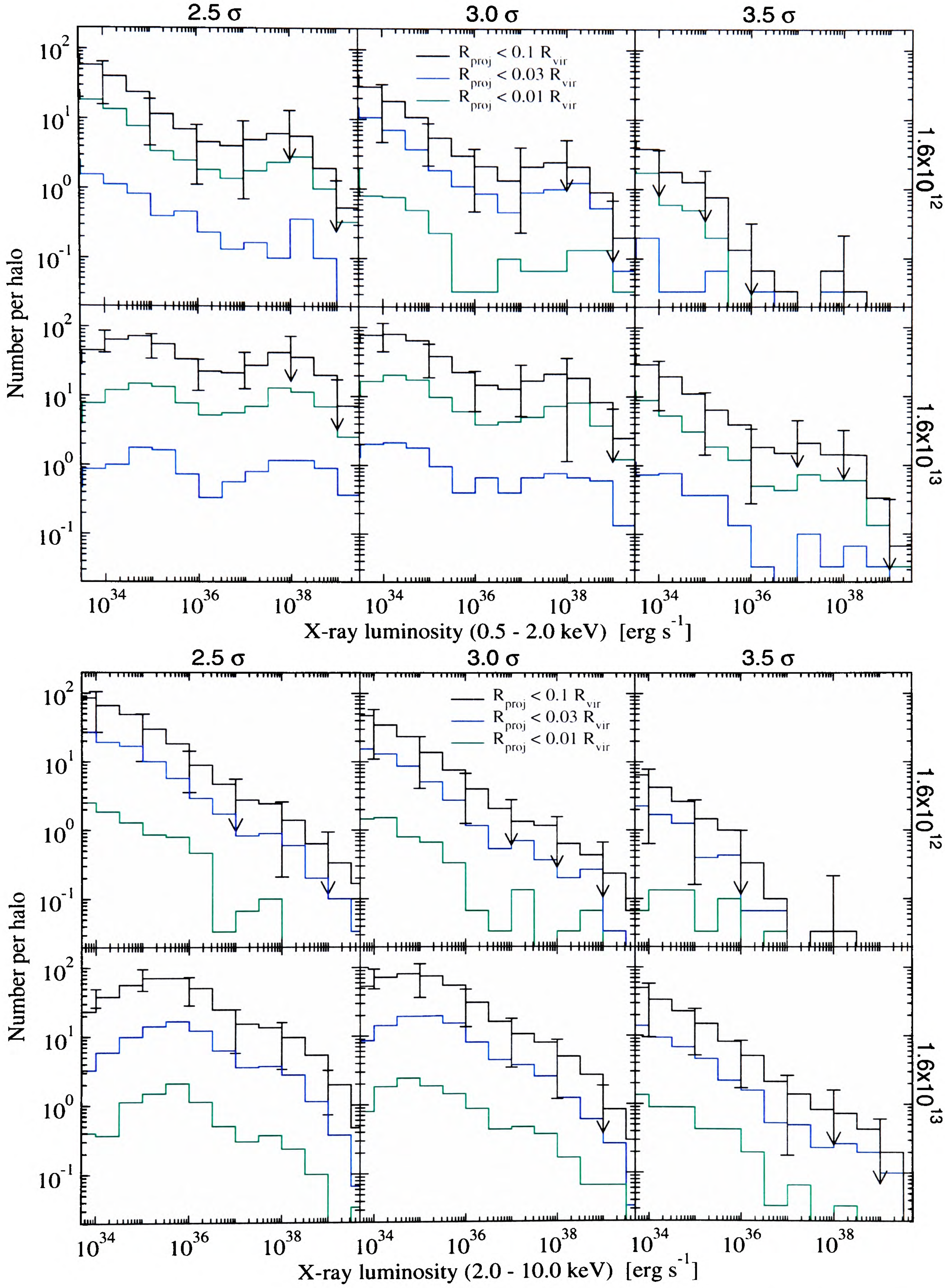


Figure 7.6: X-ray luminosity function of MBHs with (line-of-sight within) various projected distances from the host centre (see also text). Results are shown for models C, A and D (left to right panels) and X-rays in the energy range from 0.5 - 2.0 keV (top panel) and 2.0 - 10.0 keV (bottom panel).

7.5 X-rays

In our model X-rays are primarily emitted by ADAF accreting MBHs. Emission in the X-ray range is dominated by comptonisation or bremsstrahlung depending on the Compton spectral power law index, α_c (eq. 7.19). The two contributions are determined by the scaling relations equations 7.18 and 7.20.

In figure 7.6 we present the projected X-ray luminosity function, both in the 0.5 - 2.0 keV and 2.0 - 10.0 keV energy range. Again we have ignored the issue of absorption here, which can be important primarily for the softer 0.5 - 2.0 keV X-ray band.

X-ray emissions are the primary signature to look for in observations and we already mentioned that optical and UV detections are probably useful only as follow ups due to the range of potential sources and the absorption issue. Another secondary diagnostic would be to see whether any correlations exist between different parts of the emitted spectrum. However, this would be extremely dependent on the accretion model. For instance, in the ADAF model that we used, the shape of the spectrum near the exponential cut-off in the X-ray regime and in the radio/submillimeter range is essentially unaffected by the mass and accretion rate of the MBH (c.f. figure 7.3). If a MBH accretes via an ADAF we would therefore expect a correlation of the radio and X-ray luminosities.

7.6 Optical, UV and X-ray signatures of sources in Earth-centred volumes

Finally in figures 7.7 and 7.8 we have shown the bolometric, optical, UV and X-ray luminosities of sources within Earth-centred spherical volumes of varying radii. As the Earth's position we have taken a point 8.5 kpc from the host centre in the galactic disk. Since we are making predictions only for one particular halo, abundances of less than 1 have only meaning in so far as their associated error bars reach above 1. With error bars corresponding to the $1 - \sigma$ variance, this means that to find one source within the respective distance from Earth is probable to up to ~ 68 %. In either the optical or X-ray bands there is only a statistical chance of finding an accreting MBH in Earth centred volumes.

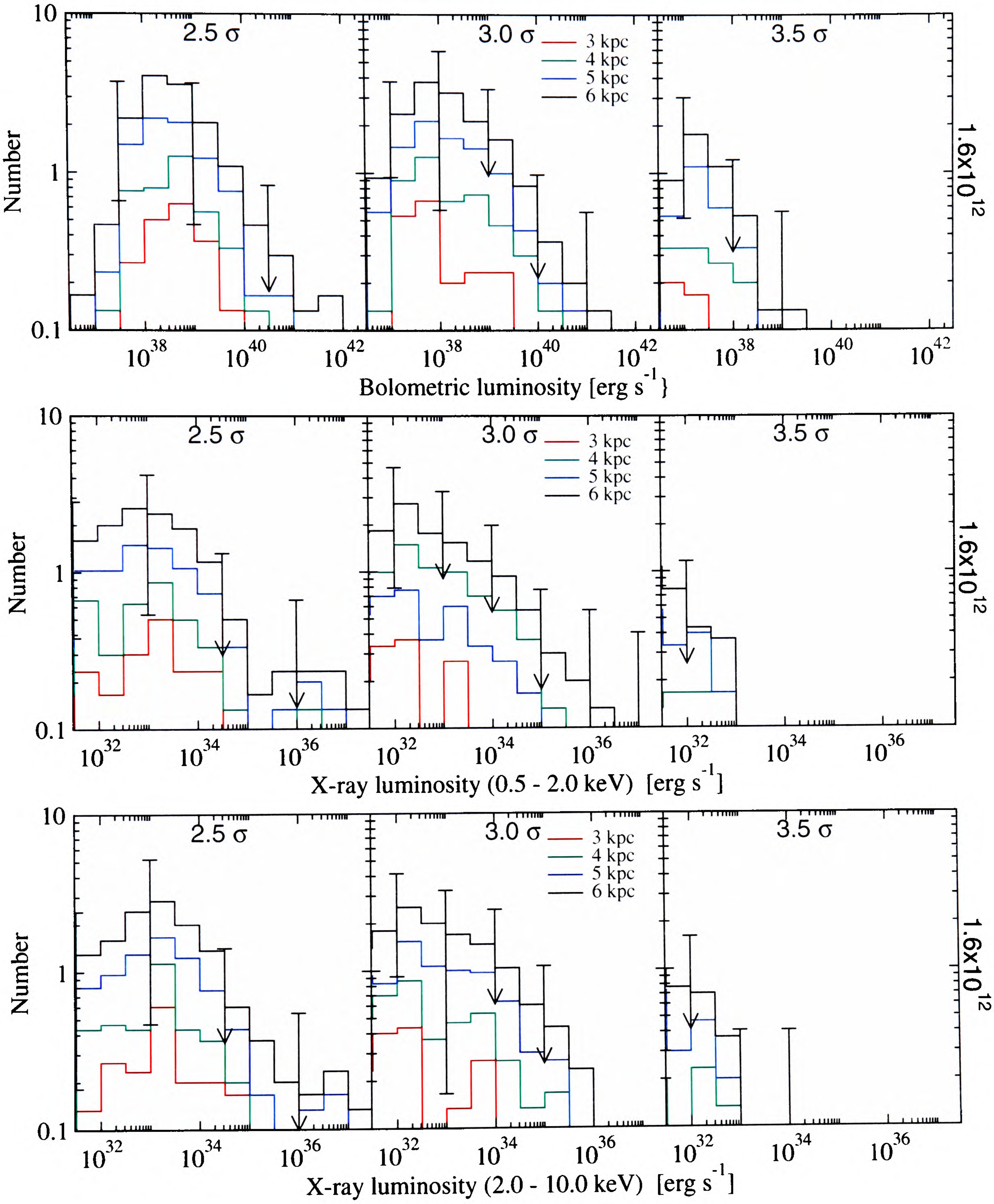


Figure 7.7: Sources in Earth-centred volumes: Bolometric and X-ray luminosity functions for MBHs accreting from baryonic cores. Results are shown for models C, A and D (left to right panels) and X-rays in the energy range from 0.5 - 2.0 keV (top panel) and 2.0 - 10.0 keV (bottom panel).

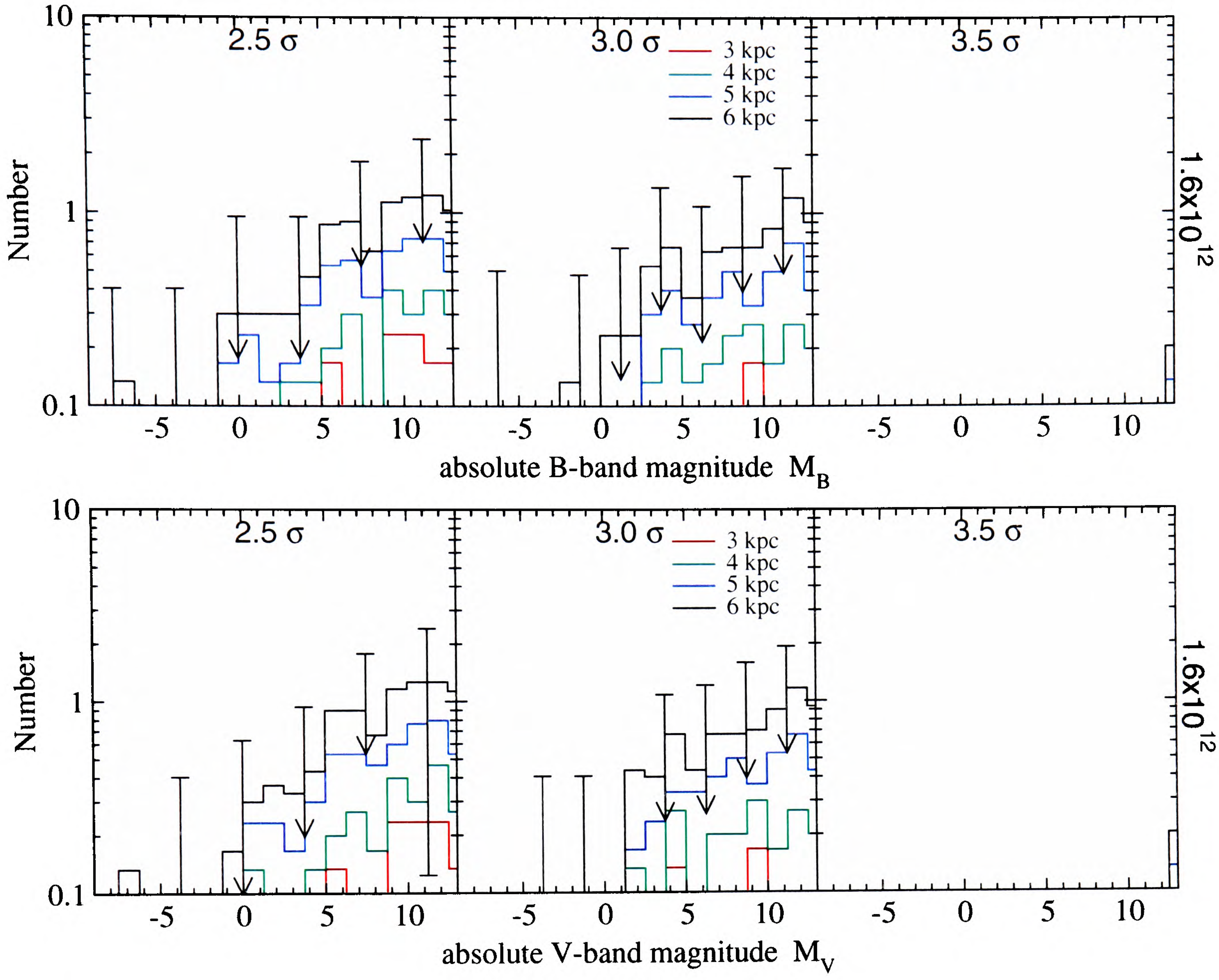


Figure 7.8: Sources in Earth-centred: Optical luminosity functions for MBHs accreting from baryonic cores. Results are shown for models C, A and D (left to right panels) in the B and V-band (top and bottom panel respectively)

7.7 Comparison with observations

We now compare our predictions with observations. As we have already noted, X-ray emissions will be the most important trace to look for. We will therefore confine our comparison to X-ray observational data only and further more to sources which have been identified as belonging to a galaxy.

In our predictions we have generally not taken into account absorption of MBH emission either in the originating galaxy as well as ours. To minimise the former one would want to look for sources in external galaxies that we see face-on. This is especially important for gas-rich spirals. To account for the extinction from the IGM and the Milky-Way's ISM all-sky extinction maps can be used (Schlegel, Finkbeiner & Davis 1998). These allow a determination of the average hydrogen column density between us and the target galaxy.

7.7.1 Observations of ultra-luminous X-ray sources (ULXs)

Our model predicts the presence of ultra-luminous X-ray sources (ULXs) in galaxies and their halos. By ULX we refer to any compact X-ray source with $L_X \gtrsim 10^{38}$erg s^{-1}. Early observations revealed that some galaxies contain ULXs (Fabbiano 1989) outside the nucleus of the galaxy. Subsequently a number of studies discovered more ULXs based on observations from the ROSAT and ASCA satellite missions (Colbert & Mushotzky 1999; Makishima et al. 2000; Roberts & Warwick 2000; Colbert & Ptak 2002). The number of observations has soared with the advent of the *Chandra* and *XMM* satellite missions. These boast a superior resolution and pushed higher the number and distance of galaxies in which ULXs can now be observed. It has initially been argued that the fact that sources emit at larger than 10^{39}erg s^{-1} requires a larger than stellar mass BH, if emission is Eddington limited and occurs isotropically. However, if ULXs do not emit isotropically but rather correspond to beamed emission that happens to be aligned with the line-of-sight, stellar mass BHs can in principle also be the source (King et al. 2001; Zezas & Fabbiano). A departure from isotropy implies that the Eddington luminosity is no longer a limiting factor, too. Radiative emission does not necessarily act along the same axis as the accretion flow anymore. High mass X-ray binaries (HMXRBs) have been suggested as candidates for these sources. Within galaxies ULXs appear to be found predominantly in star-forming regions (Kilgard et al. 2002; Zezas & Fabbiano), i.e. they seem to be associated with star burst galaxies or the spiral arms of disk galaxies. For a number of these galaxies it was found that the ratios of ULX number to massive star formation rates are similar (Smith & Wilson 2003) All this implies that X-ray binary systems could well constitute the ULXs since they are likely to have formed in star-forming regions.

However, multi-colour disk (MCD) spectral fits indicate that the associated maximum colour temperature associated with some sources is significantly lower than would be expected for a stellar mass BH system. If the maximum temperature arises from emission close to the inner radius of the accretion disk and we assume this to be the BH's last stable orbit, a lower temperature implies a larger stable orbit and thus a more massive BH (Wang 2002). Rather large numbers of

ULXs have been detected in some gas-poor ellipticals (Jeltema & Canizares 2003) and Colbert & Ptak (2002) find that the number of ULXs per galaxy is actually higher for ellipticals than for non-ellipticals. The connection between ULXs and star formation established for spiral galaxies therefore does not appear to be universally valid. The alternative of MBHs as ULXs thus appears in a stronger position.

In the baryonic core scenario we have not distinguished whether the accreting MBHs are orbiting in a system with a central spiral/disk or elliptical/lenticular galaxy. We have argued that baryonic cores survive the tidal stripping process as they get incorporated in a larger host halo. Only a passage very close to the host centre induces impulse heating to an extent that may rip the baryonic core away. It is conceivable that a similarly strong shock is imparted to baryonic cores as they penetrate the disk where it exists. This effect has been shown to be important for disk encounters of globular clusters (Binney & Tremaine 1987). In galaxies where the disk has remained relatively undisturbed for a long time, i.e. no recent major mergers have occurred, we would then expect to find a smaller number of baryonic core accreting MBHs. This argument further boosts the case for MBH ULXs in ellipticals.

A number of ULXs seem to be located in globular clusters (Jeltema & Canizares 2003). At first sight this again seems to support binaries as ULX sources. However, in the context of our model, MBHs are consistent with this. Globular clusters could be seen as the visible remnants of the baryonic cores associated with the MBHs.

To distinguish between binaries and MBHs on the basis of fits to ULX spectra alone implies that an assumption is made about the underlying physics of emission that would need to be verified independently. In addition, for some of the more distant sources the detected count rates are in fact so low that spectral fits cannot be carried out. Instead the *hardness* ratio of the number counts in a soft and a neighbouring hard X-ray band are used. The latter is further complicated by the fact that low energy ('soft') X-ray photons are absorbed more strongly than high energy photons, although the observations quoted typically account for this (e.g. on the basis of the extinction maps mentioned above).

Here we will only use spectral information to the extent that they allow to determine X-ray band luminosities from photon counts. However, spectral fits will not be used to make any

Table 7.1: Summary of X-ray observations. All galaxies are spirals except for ellipticals which are denoted '(E)'. All observations were carried out with the ACIS instrument on *Chandra* except for the observation of M31 which was observed with *Chandra*'s HRC instrument.

Galaxy	Energy range	Region surveyed	# ULXs	Reference
M31	0.1 - 10 keV	$R_{proj} < 1.13$kpc	3	(Kaaret 2002)
M81	0.3 - 8 keV	$R_{proj} > D_{25}$	2	(Swartz et al. 2003)
		disk	10	
		bulge	5	
	0.2 - 8 keV	disk ($R_{proj} \lesssim 8.5$kpc	5	(Tennant et al. 2001)
		bulge	3	
NGC 6964	0.5 - 5 keV	$R_{proj} \lesssim 13$kpc	15	(Holt et al. 2003)
NGC 1068	0.4 - 5 keV	$R_{proj} < D_{25}$	40	(Smith & Wilson 2003)
NGC 720 (E)	0.3 - 10 keV	$R_{proj} \lesssim 25$kpc	41	(Jeltema & Canizares 2003)

inference concerning the ULX source object. We simply note that while binaries may explain the majority of ULXs especially in star-forming galaxies MBHs remain a valid candidate, too. In comparing with observations, our aim is thus to show that MBH sources could account for at least some of the ULXs observed.

Table 7.1 gives a summary of the observations of point sources in individual galaxies that we are considering. Their inferred masses roughly place them in a category with the 1.6×10^{12} $M_{\odot}$halo in our model. Observed sources are considered roughly within the light radius of the galaxy. Where observations do cover regions significantly outside this radius it is typically an area to one side of a galaxy that happens to be still covered by the detection chip on the observing equipment. Interestingly a number of ULXs are detected in this area as well. Estimates of the expected number of background objects (AGN, SN etc.) show that they can account for most of the ULXs observed outside the light radius although the error margins are large. A number of sources outside the bulge have projected locations that appear to be in the disk. In this case background sources cannot account for the relatively large number of sources that seem to lie in the disk. On this basis it is maintained that these sources therefore are actually located in the disk, where their number can be explained more naturally. For sources with apparent locations in the disk as well as outside the light radius our model offers the alternative explanation that some of these sources could also be located throughout the halo but inside a column along the line of sight that projects onto the disk.

Before comparing our predictions with observations a note on how to interpret results: A comparison of predictions with observations for any individual halo or galaxy should be viewed with care. Particularly where a number of objects close to one is predicted, only observations from a statistically significant sample of galaxies provides an adequate basis for comparison.

'Consistency' between observations and predictions then means two different things: For large samples of galaxies it means that their averaged properties should compare with predictions to within the error, which is much smaller than the standard deviation we have quoted on all our results so far. On the other hand for any individual galaxy the discrepancy between observations and predictions of a given property should be at most of the order of the standard deviation.

We first look at how our results compare with observations of sources in individual galaxies. Figure 7.9 shows the cumulative X-ray luminosity function from observations and those that we predict for a 1.6×10^{12} $M_{\odot}$halo, which would be expected to host galaxies of the mass we are comparing with. For our predictions we have chosen two different X-ray bands as shown in the figure to match those from the observations more closely. Our predictions are consistent with the data from the individual galaxies if we only consider the ADAF emissions. Inclusion of the thin disk component leads to a much flatter slope and an over prediction of sources at the high luminosity end. The flat slope essentially bridges the gap in the differential luminosity function between ADAF accreting MBHs and a small number of MBHs accreting from a thin disk as shown in figure 7.6. Whether the thin disk is included or not, MBHs cannot account for the large number of sources at low luminosities, which is not a problem as we expect XRBs to certainly dominate this regime.

For luminosities larger than $L_X \sim 10^{38}$erg s^{-1} the predicted average numbers of objects are near or below 1 and we have to compare with observations of a statistical sample of galaxies. The two data sets we are using for this are that of Roberts &Warwick (2000, hereafter RW00) and Colbert & Ptak (2002, hereafter CP02) . Both sets of observations have been obtained using the *ROSAT* High Resolution Imager (HRI) instrument. In comparing with these observations we have now quoted the error on the mean (and not the standard deviation) for our predictions. The two sets of observations vary fundamentally in that CP02 *only* consider galaxies that do contain ULXs. They find a total of 87 sources in 54 galaxies of which 15 are ellipticals. They

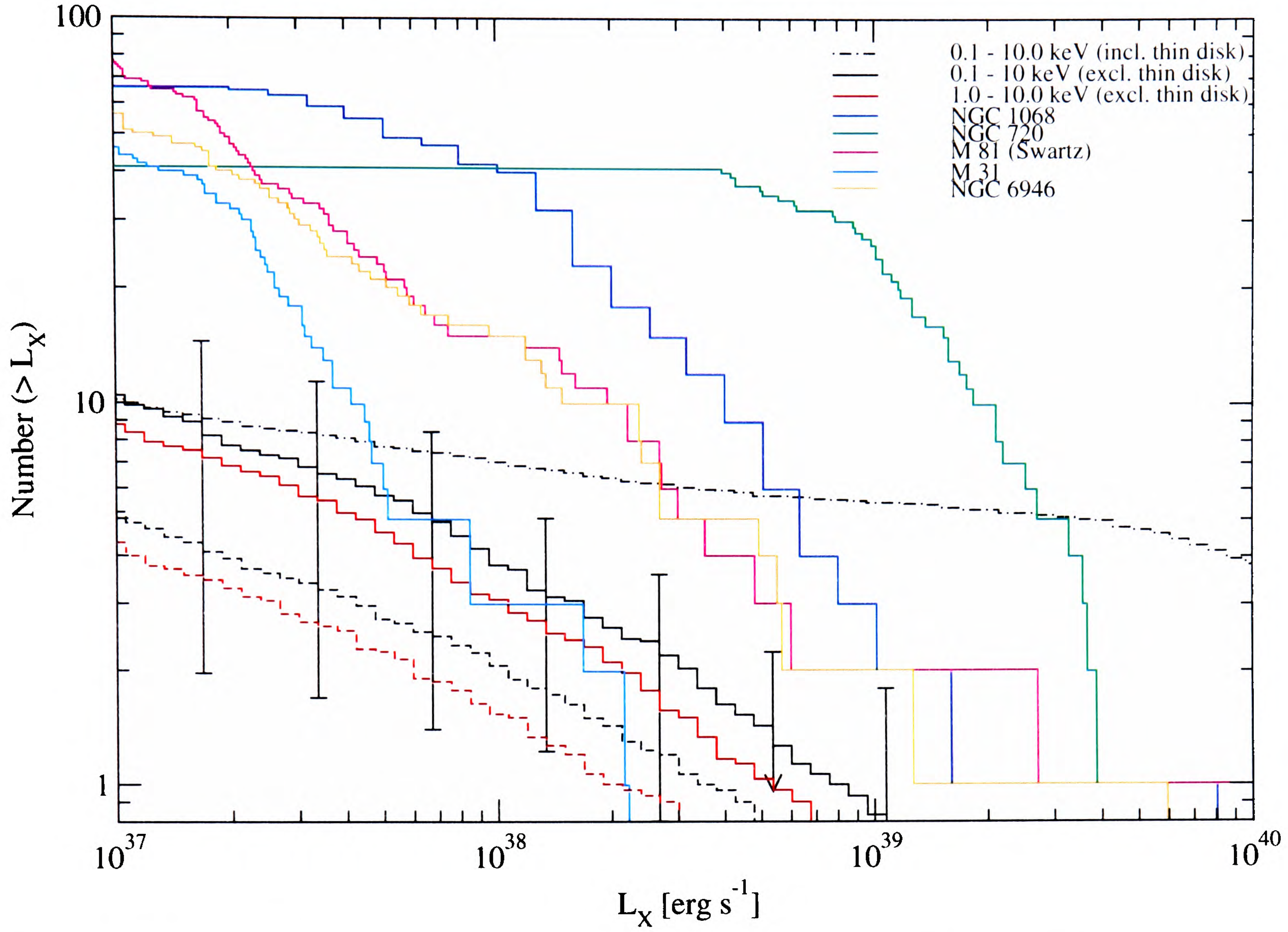


Figure 7.9: Cumulative X-ray luminosity functions for galaxies (see text) observed with *Chandra*. The red and black lines show our predicted cumulative luminosity functions for different X-ray bands. The dashed and solid line correspond to models A and C respectively for sources with projected radii < 30 kpc from the galactic centre. The dashed-dotted line corresponds to the 0.1 - 10.0 keV prediction for model C that includes the contribution from a thin disk.

note that in their sample the number of ULXs per galaxy is larger for ellipticals, which account for a total of 37 ULXs. This corresponds to 2.3 ULXs per galaxy for ellipticals compared to 1.3 per galaxy in the case of non-ellipticals. As the CP02 sample only counts galaxies that contain ULXs the derived luminosity and spatial distributions will have a systematic offset towards higher abundances.

The differential and cumulative luminosity functions are shown in figure 7.10 for the observations and our expected source population for the galaxy sized $1.6 \times 10^{12} M_{\odot}$halo in model A. We note that the luminosity bands used in the two data sets are different; while RW00 quote results for the *ROSAT* 0.1 - 2.4 keV band, CP02 give their data for the 2.0 - 10.0 keV range. To match these we have slightly changed our soft bands from the one used above (0.5 - 2.0 keV). Further we

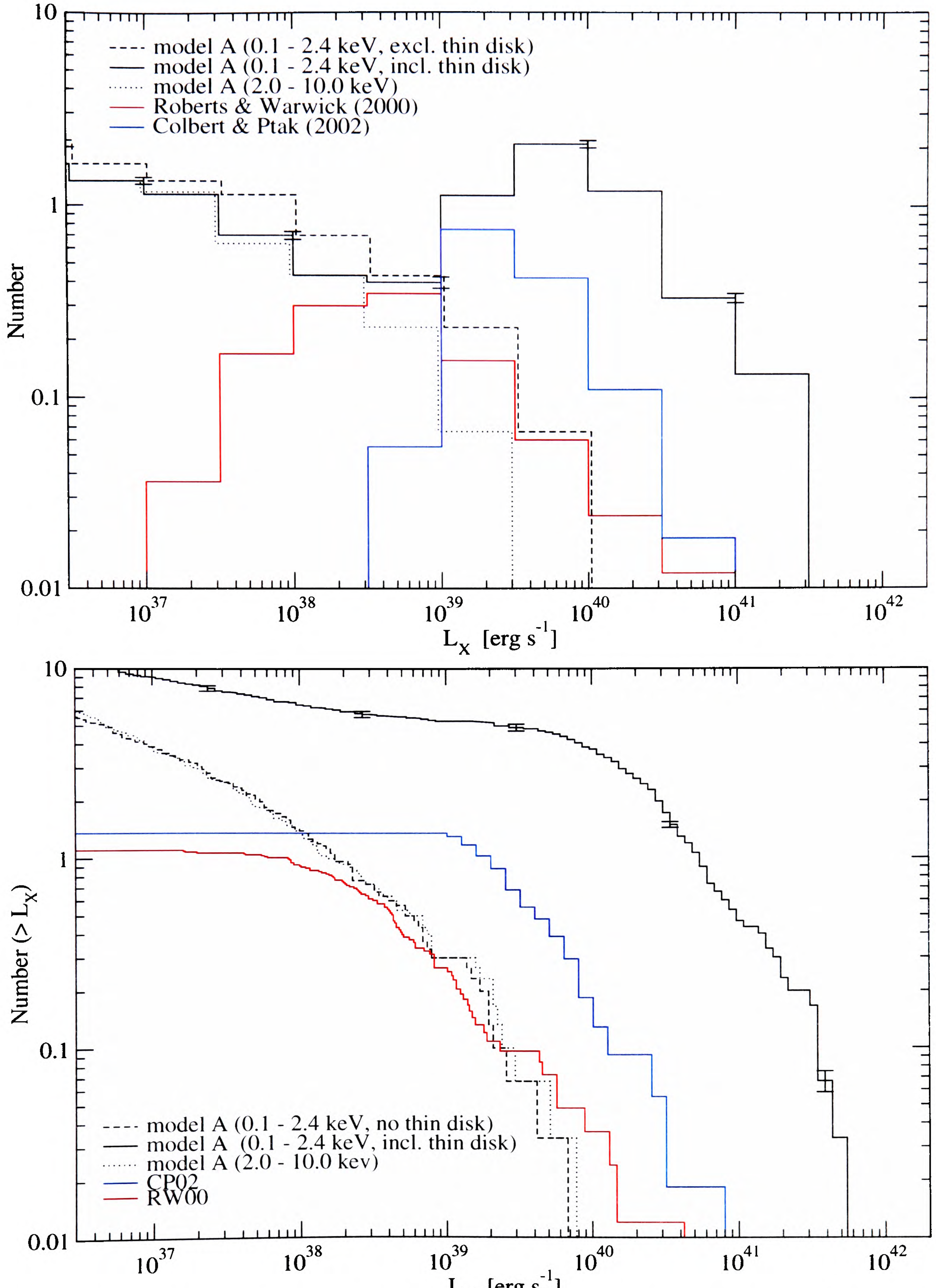


Figure 7.10: Differential (top panel) and cumulative (bottom) luminosity functions for predictions from model A ($1.6 \times 10^{12} M_{\odot}$halo) and observations. Note in particular the similar slope for luminosities larger than 10^{38}erg s^{-1}. Only sources at *radial* distances larger than $R > 0.01\ R_{vir} \approx 3$ kpc were considered. The predicted values only account for sources at projected distances less than $R_{proj} < 0.1\ R_{vir} \approx 30$ kpc.

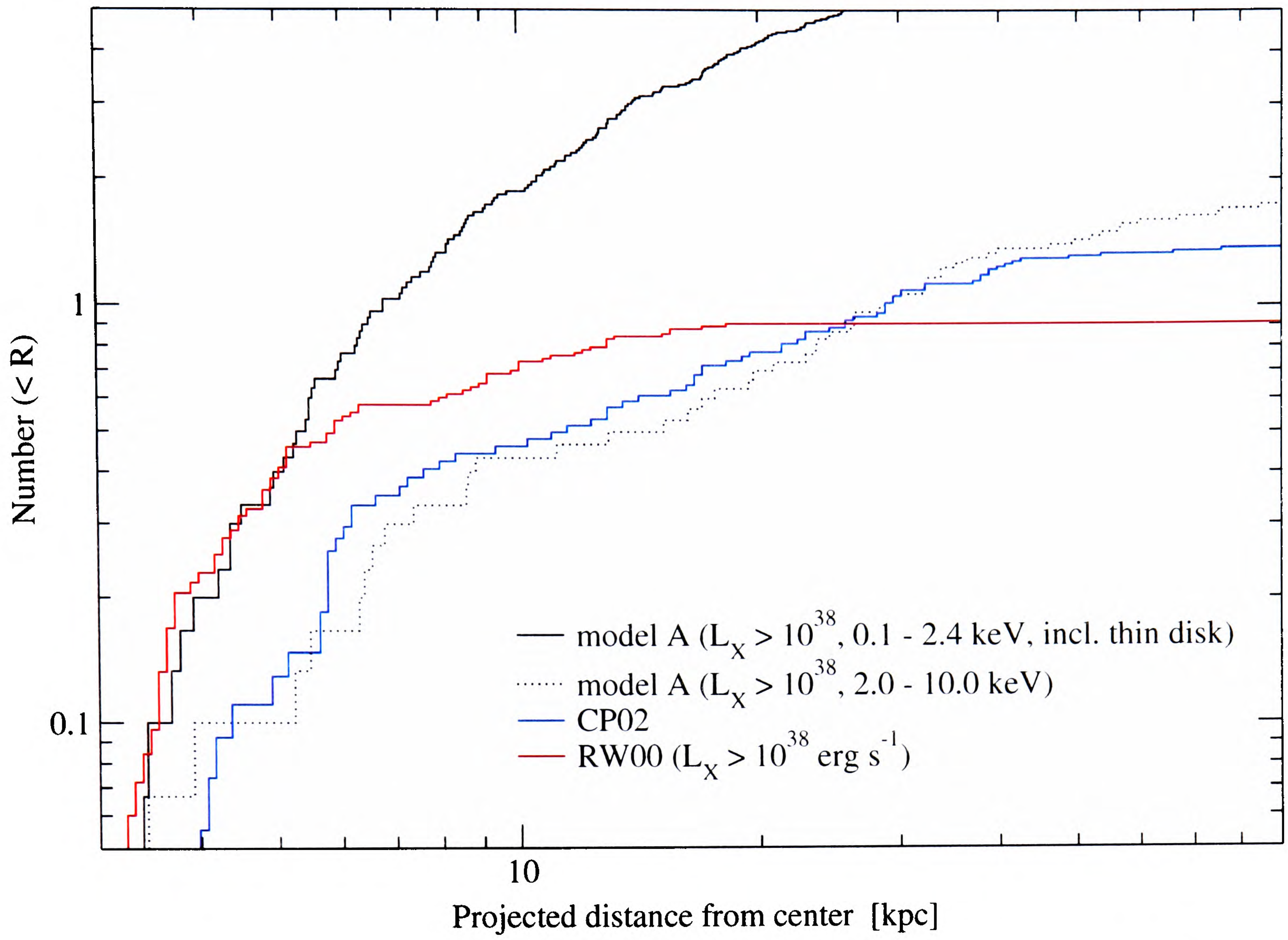


Figure 7.11: Projected radial distribution of off-centre as predicted from model A ($1.6 \times 10^{12} M_\odot$halo)and observed, in both cases for sources with $L_X > 10^{38}$erg s^{-1}. Only sources at *radial* distances larger than $R > 0.01\ R_{vir} \approx 3$ kpc were considered. The predicted values only account for sources at projected distances less than $R_{proj} < 0.1\ R_{vir} \approx 30$ kpc.

have only taken into account sources at radial distances less than 10 % of the halo virial radius, which corresponds to about $R < 30$ kpc for this halo. This is also of the order of the limiting radii below which sources in the CP02 sample are considered as belonging to a galaxy.

Between 3×10^{38} and 10^{40}erg s^{-1} our predictions in the hard band and the soft band without the thin disk contribution match the slope and normalisation of RW00. For lower luminosities the down-turn (tailing off) in the differential and cumulative distribution of the RW00 data is probably due to incompleteness. Even if no MBHs were present there would still be enough XRBs to match or exceed the number of objects at lower energies. Including the thin disk in the soft band leads to the black curve on the right, which is ruled out even by the CP02 data which, as we mentioned above, is biased towards higher abundances.

A power-law best fit yields a logarithmic slope of -0.83 ± 0.03,-0.78 ± 0.01 and -1.08 ± 0.01

for our prediction, RW00 and CP02 respectively.

The problem of too high luminosities when including the thin disk component can be remedied in several ways. One is to simply ignore the thin disk component: The relatively large accretion rates in MBHs accreting from thin disks may have just exhausted the baryonic core if the latter was not large enough to start with. Past merger activity of MBH hosts may also have triggered periods of mini-quasar activity that has lead to a faster depletion of gas in the baryonic core.

A second and more general possibility is that accretion rates were much smaller from the outset. The accretion rates we had determined at the end of chapter 6 for baryonic core accretion were all based on a number of assumptions that yield an estimate accurate to an order of magnitude at most. If the accretion rates were in fact lower by only a factor 5 or so, accretion rates – even for the highest ones with $\dot{m} \approx 0.1$ – would fall below the critical accretion rate (eq 7.15) and all MBHs would accrete in ADAF mode. However, abandoning the thin disk contribution would also remove the corresponding abundance of sources in the high luminosity end of the B and V band luminosity functions.

In figure 7.11 we show the number of sources vs projected distance from the centre. The predictions we show are for model A and the same parameters as in figure 7.10. We did omit the results for the soft band without the thin disk, as this curve is essentially the same as the one for the hard band. The slopes of the spatial distribution of the latter – and by extension also that of the soft band without thin disk – also match that of the observations reasonably well. However, agreement gets increasingly worse for the RW00 sample towards small host distances where its slope gets considerably steeper. This is probably because RW00 have imposed a limiting minimum angular distance below which they did not consider any sources to avoid confusion with a central source. We could not find any indication for MBH 'luminosity segregation' within halos, which is in line with the same negative finding for the radial distribution of MBH masses as noted in figure 6.3.

Apart from the systematic bias in the CP02 data the normalisation is also affected by varying degrees of completeness in the two samples. For instance the CP02 set features no source with $L_X < 10^{39}\text{erg s}^{-1}$, while the RW00 data are restricted to much smaller radii. Both CP02 and RW00 maintain that contamination with background AGN is rather small and accounts for no more than about 11 - 13 % (RW00).

In the near future compilations of better data should be available using the superior resolution of the *XMM* and *Chandra* satellite observatories. One of the first comparable studies is the recent one by Swartz, Ghosh and Tennant (2003) , which is as yet unpublished.

Another way to probe the nature of the objects that constitute ULXs (as well as less luminous MBH powered sources) is to look at variability time scales of X-ray emission. A tentative linear relation between variability time scale and BH mass has been established although with large uncertainties (Markowitz et al. 2003). The relation was found to apply to both the SMBHs powering the AGN of Seyfert 1 galaxies as well as stellar mass galactic X-ray binary systems. Since XRBs display variability across a wide range of time scales, this relation, if confirmed, is probably only usefully applied to large samples of ULXs. If both HMXRBs as well as MBHs power ULXs, variability studies might help establish the existence of two distinct time scales on which ULXs should be found to display variability although with large scatter.

7.7.2 Baryonic core accretion at the centre of dwarf galaxies and globular clusters ?

MBH accretion from baryonic core remnants raises the question of whether dwarf galaxies and star clusters are likely sites for the observation of MBHs and thus luminous X-ray point sources. In the hierarchical structure formation scenario dwarfs orbiting inside the halos of larger galaxies correspond to satellite systems and could in principle contain accreting central MBHs. For a Milky-Way sized halo the semi-analytical scheme we used predicts these dwarfs to orbit at distances $\sim 10 - 100$ kpc, where the latter is of the order of the virial radius of the Milky-Way's dark matter halo. All known satellites of the Milky Way do lie in this range and we look at relevant X-ray observations of these in the next section.

For smaller distances any satellite system is more likely to have been partially or fully stripped. We referred to MBHs in these systems as naked, i.e. MBHs that, although embedded in a baryonic core, are not associated with any extended stellar/gas distribution. Here we would not expect to observe anything other than accreting MBHs unless a significantly large number of stars is contained in a relatively small region around the baryonic core.

This description appears to fit at least some globular clusters (GCs). A lot of the observations we have considered above find that ULXs appear to be associated with GCs. Among the formation scenarios that are considered for GCs are that they represent remnants of high redshift pre-galactic fragments (Beasley et al. 2002) or even that they were formed in MBC halos that gave rise to the first massive stars (Lin & Murray 2002). In the context of our model both could correspond to (remnants of) small sized, high redshift satellite systems. As such they are consistent with the existence of a baryonic core accreting MBH at their centre. The lack of gas in GCs is not really an obstacle, as we already pointed out that a baryonic core of accretable gas can in principle be very small. What is more of a problem is the required mass for an MBH to appear as an ULX inside a GC. The high redshift origin of GCs implies that at their formation MBHs only had a mass equal to the MBH seed mass. On the basis of the accretion rates in figure 6.7, only in few cases this is large enough to power a ULX. However, there is still the possibility that since then the MBHs may have grown significantly through stellar dynamical processes inside the clusters, which we discussed in section 3.3.2.

7.7.3 Sources in and around the Milky-Way

The mass of the Milky Way corresponds to that of a galaxy that would be hosted inside a 1.6×10^{12} $M_{\odot}$halo. For this we predict an average number of up to three (model C) MBH sources with X-ray luminosities exceeding 10^{38}ergs^{-1} within 10 % of the virial radius from the centre. This is the radius which we have taken roughly as the radial extent of the galaxy. A recent survey of X-ray sources in the galaxy was carried out by Grimm, Gilfanov & Sunyaev (2002) . Their work considers sources in the 2.0 - 10 keV band, which gives probably the best representation of the sources; the optical and soft X-ray bands are likely to be severely affected by the large hydrogen column densities in the disk. Crucially the largest part of the sources and all of the most luminous ones above 10^{38}ergs^{-1} have been positively identified as being binary systems. The unidentified sources have luminosities significantly less than 10^{38}ergs^{-1}.

Let us assume for the moment that the X-ray luminosity is close to the maximum in the ADAF spectrum. Even then the absence of any ultra-luminous X-ray source in the Milky-Way that has *not* clearly been identified with a binary system is consistent with our predictions. On average

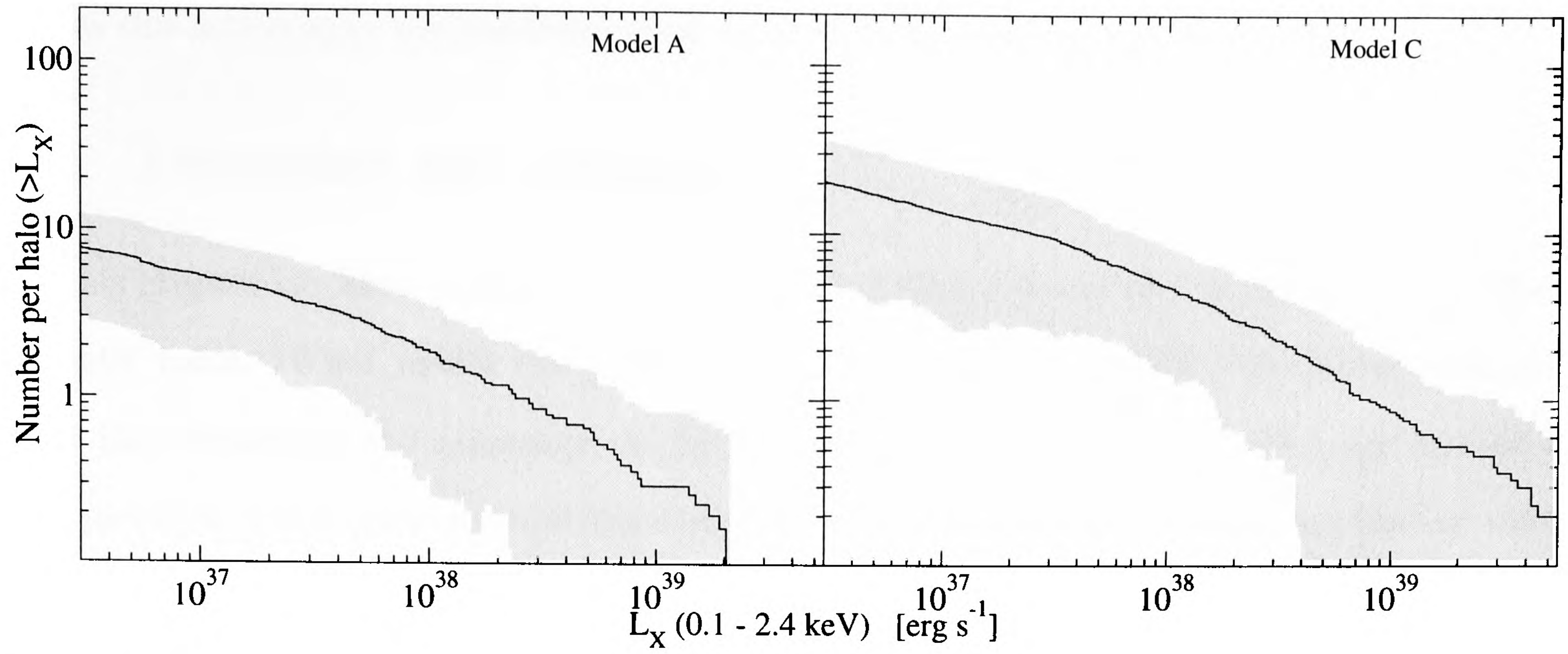


Figure 7.12: Cumulative X-ray luminosity functions for a $1.6 \times 10^{12} M_\odot$halo. The shaded area is the standard deviation around the mean number given by the lines. Only model A is consistent with the absence of sources more luminous than 10^{38}erg s^{-1} in the Milky-Way halo.

we would expect about one to two ULXs as can be seen in figure 7.9, but the standard deviation means that a zero count is possible. That this limit could apply to the Milky-Way is also indicated by the mass of the Milky-Way's central SMBH (2.6×10^6 $M_\odot$), which lies significantly below the $M_{SMBH} - M_{bulge}$ and $M_{SMBH} - \sigma_c$ relations determined from observations (c.f. chapter 3). If we assume that gas accretion has increased the mass of the Milky-Way SMBH by the same factor as in all other SMBH then the lower mass must be due to a lesser number of major mergers. The latter implies that we would then also expect a correspondingly lower number of massive and thus luminous accreting halo MBHs.

Figure 7.12 shows the cumulative X-ray luminosity function for the entire halo of a Milky-Way sized galaxy. We have chosen a soft X-ray band in the energy range 0.1 - 2.4 keV to compare with observations of galaxies in the local group as compiled by Zang & Meurs (2001) . We are particularly interested in those local group galaxies that are satellites of the Milky Way. The observations focus on the core regions of these galaxies, which is where we would expect MBHs to reside, if they are present at all. For all known dwarf satellites of the Milky Way the X-ray luminosity from their cores does not exceed $\sim 10^{37}$erg s^{-1} and typically is much lower. However, for the larger satellites the situation looks different. For the LMC, for example, Kahabka (2002) finds a number of sources at X-ray luminosities larger than 10^{38}erg s^{-1} that have been classified as binary systems. Since the classification of binaries has been done on the basis of X-ray hardness ratios this leaves open the possibility that some of them could be accreting MBHs.

7.8 Discussion and summary

In this chapter we have computed the observational consequences of a population of MBHs in galactic halos. What makes the prediction let alone comparison with observation difficult is that the abundance and masses of the MBHs fall right in the middle of the range marked by the presence of one galactic super-massive BH at one end and that of large numbers of stellar mass binary systems at the other. For MBHs neither their mass nor their abundance and spatial distribution allows for a unique identification by any one type of observation alone. As far as optical and X-ray signatures are concerned MBHs should be just numerous enough to differentiate them from the expected count of background sources. On the other hand the MBH mass and accretion rate is not large enough for their accretion luminosity to be of a significantly different order of magnitude than that of XRBs. In this context the superior data from *Chandra* and *XMM* should allow for better spectral modelling which in turn should enable a clearer identification of the ULX sources.

We have shown that our predictions, particularly from model A, are consistent with X-ray luminosities inferred from observations. The observed X-ray luminosity is determined by fitting a spectral model to the detected count rate. Depending on the model fitted the resulting luminosity can differ by a factor of order unity, which we have deemed sufficiently accurate to compare with predictions. This uncertainty is probably less significant than that arising from the assumption made about the intervening absorbing column densities, which have a strong impact on the counts of soft X-ray photons.

However, while different spectral models may yield similar luminosities they make significantly different assumptions about the underlying emission process and ultimately the nature of the ULX source. Typical fitting models include a multi-colour disk (MCD) and/or a power law. They appeal because of the very few parameters involved. In most cases the colour temperature of the MCD implies a $\sim 1-10$ $\mathrm{M}_{\odot}$BH emitting at or above the Eddington rate. It seems that the ADAF model applied to sub-Eddington accreting MBHs can match not only the X-ray luminosities but also spectral range of detected counts of ULXs. We therefore propose to routinely include ADAF models in spectral fits to X-ray observations.

ADAF models do involve one or two more key parameters, which may introduce degeneracies when used to fit X-ray data alone. The advantage is that for a single (no binary) MBH ADAFs also predict the spectrum in the optical regime and as well as the existence of a correlation of the radio and X-ray fluxes. Consequently observations in these bands could help break the degeneracy arising from the increased number of parameters. In contrast for binaries these bands may carry the spectral contribution of the visible companion, accounting for which adds more parameters, too. To fit the overall spectrum ADAFs and composite disk models for binary systems thus seem equally well placed.

Because of the absorption problem optical follow-ups are likely to yield best results for ULXs that are observed at high galactic latitudes to avoid the effect of the Galactic disk. For such ULXs optical data could help distinguish between MBHs and binary systems for sources placed within galaxies. To test the prediction of MBHs in the dark matter halos outside galaxies and distinguish them from background AGN requires a systematic survey of X-ray point sources and optical follow-ups in fields around galaxies.
Our results in particular are

- We have applied a thin disk and ADAF to model the spectral signature of accreting MBHs. Applying this we predict a few ULXs on average per galaxy sized halo. This is not consistent with observations.

- This problem was found to arise from the thin disk emission which contributes a large part in the soft X-ray band. Removing the thin disk component would remedy this. MBH accretion rates only have to be lower by a factor ~ 5 for all MBHs to switch to ADAF accretion. This is consistent with the much larger uncertainty associated with the baryonic core accretion rate.

- B and V band magnitudes reach -10, or -15 when a thin disk is included. Because of absorption these are probably more useful for follow-ups.

- For ADAF accretion alone we match the slope of the luminosity function for one statistical sample of observations, and both, slope and normalisation for another one. On average we expect to find one ULX in a Milky-Way sized galaxy, although, the actual absence of one in the Milky-Way is consistent with our predictions.

- We did not find any luminosity segregation in the MBH sources. The slope of the cumulative radial distribution of ULXs is also quite similar for observations and our prediction.
- Of the models we considered, model A (first star formation in MBC halos collapsing from 3 σ peaks at $z \approx 25$) appears to be more consistent with observations than the others.

The interpretation of ULXs as HMXRBs relies to a large extent on the lack of alternatives. We believe to have shown that MBHs not only present an alternative but that, in the context of our model, they can even explain (the slope and normalisation of) the high luminosity end of the ULX luminosity function. Furthermore the observed higher prevalence of ULXs in elliptical galaxies compared to spirals is even more difficult to explain by HMXRBs. Their typical prevalence in starforming regions hardly makes them natural candidates for ULXs in ellipticals. In turn the case for MBHs as ULXs in stronger in ellipticals, since small satellites and globular clusters, which would be expected to host accreting MBHs, are less likely to be disrupted due to the absence of a disk component.

The major uncertainties in the MBH luminosities are due to the accretion rate and mass of MBHs which, as we have seen in the last chapter, are based on the assumptions of efficient merging and the presence of a constant density structure of the baryonic core. If merging does not proceed efficiently and we ignore any significant mass increase through gas accretion, no MBHs more massive than the seed MBHs are created and as a result accretion rates and X-ray luminosities would never reach the high values so as to be observable, certainly for the simple structure of the baryonic core we have assumed. However, higher accretion rates , e.g. for baryonic cores with power law density profiles, would change this conclusion.

While the luminosities of accreting MBHs in individual bands depend on the parameters of the particular spectral model used, the predictions of luminosities in the X-ray regime for the ADAF model are generally proportional to the bolometric luminosities. That means we can restate our results more generally, by saying that if MBHs emit in X-rays a some fixed fraction of the bolometric luminosity (from baryonic core accretion) they can account for the observations of ULXs. For the ADAF model that X-ray fraction happened to be in the right region to agree with observations, but more generally we can keep this fraction as a free parameter encapsulating the overall uncertainty arising from the accretion rate and the particular spectral model used.

Chapter 8

Observational Signatures II: Redshift Evolution of MBH Abundance

Complimentary to the observation of optical and X-ray emissions from MBHs in the local Universe, the detection of gravitational waves from coalescing (S)MBH - (S)MBH systems could provide limits on the abundance of these objects at cosmological distances and to very high redshifts.

At the highest redshifts the observation of gamma ray bursts may allow us to directly probe the epoch of first star formation at redshifts larger than 20.

In the context of our model we predict the number of gravitational wave events in section 8.1. Section 8.2 considers the expected number of GRBs from exploding massive population III stars. In section 8.3 we look at any other possible detections. These include the possibility of observational signatures from stars that get tidally disrupted by MBHs and microlensing events from nearby MBHs. We discuss and summarise our findings in section 8.4

8.1 Gravitational waves from SMBH-MBH Mergers

Gravitational wave emission provides another and complementary way of testing our model. In principle it could be used to constrain the merging history of MBHs leading up to the present-day abundance of MBHs. There are essentially two types of gravitational waves that are emitted at different stages in the evolution of the scenario we are considering.

One of these are 'burst' signals associated with the collapse of massive population III stars into

MBHs. In our model these are events occurring at very high redshifts > 20 for a cosmological abundance of massive population III stars and so we expect a stochastic gravitational wave background to be generated (Schneider et al. 2000;de Araujo, Miranda & Aguiar 2002). Although statistical limits can be imposed, the particular number and strength of burst signals created in this way does depend on the initial number and masses of collapsing stars and the collapse mechanism. Particularly the former, however, is not well constrained.

In contrast periodic signals, although with changing frequency, are emitted during the subsequent merger of MBHs to form more massive BHs. These can be further subdivided into signals associated with the inspiral phase, actual coalescence and subsequent ringdown (Flanagan & Hughes 1998). Of these the inspiral phase is the longest and thus most likely to detect. In the following we will be exclusively concerned with the gravitational waves emitted during this phase. These are also referred to as 'chirp' signals which comes from the fact that during inspiral the distance between the BHs in the merging binary continuously decreases, due to the energy loss through gravity waves, resulting in the frequency of emitted waves rapidly sweeping upwards.

8.1.1 Gravitational Wave Amplitude

For any measurement of these the most important pieces of information are the frequency and some measure of their amplitude, both of which will be functions of time. Because of the spin 2 nature of gravity the first non-zero and largest term of the tensor describing the gravity wave induced perturbation of space-time is the quadrupole moment. This can be expressed as an overall dimensionless strain amplitude h_s. For periodic signals such as those emitted by a binary black hole system this is given by (Thorne 1989)

$$h_s = 8\left(\frac{2}{15}\right)^{1/2}\frac{G^{5/3}\mu_{12}}{c^4 D(z)}(\pi M_{12} f_{gw})^{2/3} \tag{8.1}$$

where f_{gw} is the frequency of gravitational waves, which in this case is just twice the orbital frequency of the binary; $M_{12} = M_1 + M_2$ and $\mu_{12} = M_1 M_2 / M$, G is Newton's constant and c the speed of light. $D(z)$ is the distance to the binary and is obtained by integrating the differential distance redshift relation eq 2.25. The effective strain amplitude measured at some characteristic frequency f_c is further enhanced proportionally to the square root of the number of periods, n, that the binary emits at or near that frequency. For an inspiralling binary system, whose frequency sweeps up - or 'chirps' in frequency (c.f. below), $n \propto \mu_{12}^{-1/2} M_{12}^{-1/3} f_c^{-5/6}$ and the characteristic strain amplitude, h_c of this is thus (Thorne 1989; Nakamura et al. 1997)

$$h_c = 2.94 \times 10^{-19} \left(\frac{M_{chirp}}{10^3 \mathrm{M}_\odot} \right)^{5/6} \left(\frac{f_c}{\mathrm{Hz}} \right)^{-1/6} \left(\frac{D(z)}{10 \mathrm{Mpc}} \right)^{-1} \tag{8.2}$$

where f_c is the frequency of the wave signal, r the distance from the merging binary and the 'chirp' mass M_{chirp} accounts for the masses M_1 and M_2 of the binary constituents

$$M_{chirp} = \frac{(M_1 M_2)^{3/5}}{(M_1 + M_2)^{1/5}} = \mu_{12}^{3/5} M_{12}^{2/5} \tag{8.3}$$

Particularly at late times most mergers are between a massive central MBH and much lighter MBHs, which are mostly seed MBHs so that we can consider the problem in the limit that one of the merging binary constituents is much larger than the other. If we assign M_1 and M_2 to the masses of the central and inspiralling MBH respectively, $M_1 \gg M_2$ and

$$M_{chirp} \approx M_1 x^{3/5} \tag{8.4}$$

where $x = M_2/M_1$ denotes the mass ratio between the two masses.

The largest frequency at which a gravitational wave source can emit is determined by its light radius which in turn is dependent on its mass, and so in this case $f_{max} \approx 10\ \mathrm{Hz}(10^3 \mathrm{M}_\odot / M_1)$. A lower limit on the characteristic wave amplitude is then

$$h_c \approx 2 \times 10^{-19} \sqrt{x} \left(\frac{M_1}{10^3 \mathrm{M}_\odot} \right) \left(\frac{D(z)}{10 \mathrm{Mpc}} \right)^{-1} \tag{8.5}$$

8.1.2 Frequency of gravitational Waves

The proposed LISA (Laser Interferometer Space Antenna) gravitational wave observatory will be the only instrument capable of detecting gravity waves in the typical frequency range generated in the inspiral phase of a merging binary of massive black holes. A rough estimate of the maximum frequency of waves emitted by the merging binary is

$$f_{max} \sim 10 \text{ Hz} \frac{10^3 \mathrm{M}_\odot}{M_1} \tag{8.6}$$

where M_1 is the more massive of the binary constituents.

For a binary on circular orbit the frequency of the chirp signal can be described analytically. If the length of the semi-major axis is a_0 at time $t_0 = 0$ then the time for the binary constituents to spiral into each other is (Misner, Thorne & Wheeler 1973)

$$\tau_0 = 0.665 \left(\frac{M_{12}}{10^3 \mathrm{M}_\odot}\right)^{-2} \left(\frac{\mu_{12}}{10^3 \mathrm{M}_\odot}\right)^{-1} \left(\frac{a_0}{10^8 \mathrm{m}}\right)^4 \mathrm{yr} \tag{8.7}$$

The frequency of the waves emitted at a later time t is

$$f_{gw} = 3.11 \times 10^{-3} \left(\frac{M_{chirp}}{10^3 \mathrm{M}_\odot}\right)^{-5/8} \left(\frac{\tau_0 - t}{\mathrm{yr}}\right)^{-3/8} \mathrm{Hz} \tag{8.8}$$

Including the redshift z the waves are then detected by LISA with a frequency

$$f_{gw} = \frac{3.11 \times 10^{-3}}{(1+z)} \left(\frac{M_{chirp}}{10^3 \mathrm{M}_\odot}\right)^{-5/8} \left(\frac{\tau_0 - \frac{t}{1+z}}{\mathrm{yr}}\right)^{-3/8} \mathrm{Hz} \tag{8.9}$$

For the case that $M_1 \gg M_2$, τ_0 and f_{gw} are approximately given by

$$\tau_0 \approx \frac{0.665}{x} \left(\frac{M_1}{10^3 \mathrm{M}_\odot}\right)^{-3} \left(\frac{a_0}{10^8 \mathrm{m}}\right)^4 \mathrm{yr} \tag{8.10}$$

$$f_{gw} \approx \frac{3.11 \times 10^{-3}}{(1+z)} \left(\frac{M_1}{10^3 \mathrm{M}_\odot}\right)^{-5/8} \left(x\frac{\tau_0 - \frac{t}{1+z}}{\mathrm{yr}}\right)^{-3/8} \mathrm{Hz} \tag{8.11}$$

8.1.3 MBH Merger Efficiency

Above we have outlined the main characteristics of the gravitational wave signals received from individual merger events and the formulae given estimate the amplitude and frequency *averaged* over all inclinations of the merger orbital plane.

For a merger event to actually occur the MBHs have to come within a distance where gravitational waves can efficiently reduce the orbital energy and allow the MBHs to spiral together and coalesce. However, this only happens at MBH separations that are very much smaller than the ones we are looking at here.

For a MBH binary on a circular orbit to coalesce through gravitational wave emission within a Hubble time, the MBH separation, a, needs to satisfy (Peters 1964)

$$\begin{aligned} a \leq a_{gw} &= \left(\frac{256\, t_{hubble}\, G^3\, \mu_{12}\, M_1^2}{5\, c^5}\right)^{1/4} \\ &\approx 4.5 \times 10^{-6} \left(\frac{h}{0.7}\right)^{-1/4} \left(\frac{M_1}{260\, M_\odot}\right)^{3/4} \left(\frac{M_2(M_1+M_2)^2}{M_1^2}\right)^{1/4} \text{ pc} \end{aligned} \tag{8.12}$$

At distances larger than this the common environment of the MBHs determines whether and how quickly the MBHs will be driven towards each other. If two MBHs come within a distance

$$a \lesssim a_h = \frac{G\, M_2}{4\, \sigma_c^2} \approx 7.3 \times 10^{-2} \left(\frac{M_2}{260 M_\odot}\right) \left(\frac{\sigma_c}{200 \text{ km s}^{-1}}\right)^{-2} \text{ pc} \tag{8.13}$$

they form a 'hard' binary. Interactions with stars in the binary's vicinity carry away energy from the binary with the result that the binary separation shrinks - the binary is said to 'harden'. The time scale on which this happens is given by

$$t_h = \frac{\sigma_c}{G\, \rho_\star\, a\, H_h} \tag{8.14}$$

where H_h is the dimensionless hardening rate and is of the order $H_h \approx 15$ (Quinlan 1996). In a fixed isothermal stellar background with a number density of stars, $\rho_\star = \sigma_c^2/(2\ \pi G r^2)$ the hardening time scale starting at $a = a_h$ becomes

$$t_h = \frac{\pi\, G\, M_2}{2\, \sigma_c^3\, H_h} \approx 1.44 \times 10^7 \left(\frac{M_2}{260\, M_\odot}\right) \left(\frac{\sigma_c}{200 \text{ m s}^{-1}}\right)^{-3} \left(\frac{H_h}{15}\right)^{-1} \text{ yr} \tag{8.15}$$

As it stands this implies that the hardening time scale is proportional to the lighter MBH in the binary system. This makes sense: as the mass of M_2 increases the fraction of energy lost to any interacting star decreases. A light MBH binary therefore needs less stellar interactions to harden. However, this is only strictly valid if the stellar background is fixed. That is, we have ignored any depletion of the stellar density in the binary environment due to the interactions and resulting ejections of stars from the binary.

We already mentioned in section 5.3.2 that the hard binary stage can represent a 'bottle neck' on the way to MBH mergers - particularly for very massive (S)MBHs, as dynamical friction is no longer significant but gravitational radiation not yet effective enough in further reducing the separation between the MBHs. However, both, the dynamical interaction with stars in the MBH binary's vicinity as well as gas infall are likely to have the potential to quickly reduce the distance between the MBHs. We do not quantify this effect, but note that the role of gas infall in this context is probably more important at early times when MBHs encounter each other more often in the wake of major mergers of two host halos [1].

Dynamical friction at the centre of halos

At still larger distances dynamical friction is the main mechanism by which MBHs are drawn to the centre. This effect has been modelled in the numerical scheme using the dynamical friction time scale eq.(5.5). In the following we assume that this time scale dominates both the time scale for gravitational wave induced coalescence and that for overcoming the hard binary stage, which in principle can be very short due to the aforementioned gas and stellar dynamical processes.

In order to compute the time when a MBH merges with the central (S)MBH in the host we first establish the time when a satellite/MBH system sinks to an orbit with a radius less than the radius of the infall region. This is 1 % of the host virial radius at the time. At this point the satellite/MBH system in question has so far been considered as having 'fallen to the centre'; following their dynamics to smaller radii was too expensive computationally.

However, to this time we now add 'by hand' the dynamical friction time using eq.(5.5) and replacing the total host mass interior to and orbital period at the virial radius with the respective quantities at the outer radius of the infall region. This is also shown in figure 8.2 for a range of satellite to host mass ratios. The circularity parameter ϵ is computed from the last orbit of the satellite system before it crossed into the infall region. The result is a more accurate estimate of the time when the actual merger occurs. As a result of the correction to the infall time we also inevitably change the order in which the MBHs merge with the centre. This in turn leads to a

---

[1] At very high redshifts $z \gtrsim 10$ the mass ratios M_2/M_1 between MBHs are closer to 1 on average as central SMBHs have not yet grown to the massive sizes observed today. The more similar mass of MBHs implies that they encounter each other in the wake of the merger of two halos that are also of similar mass.

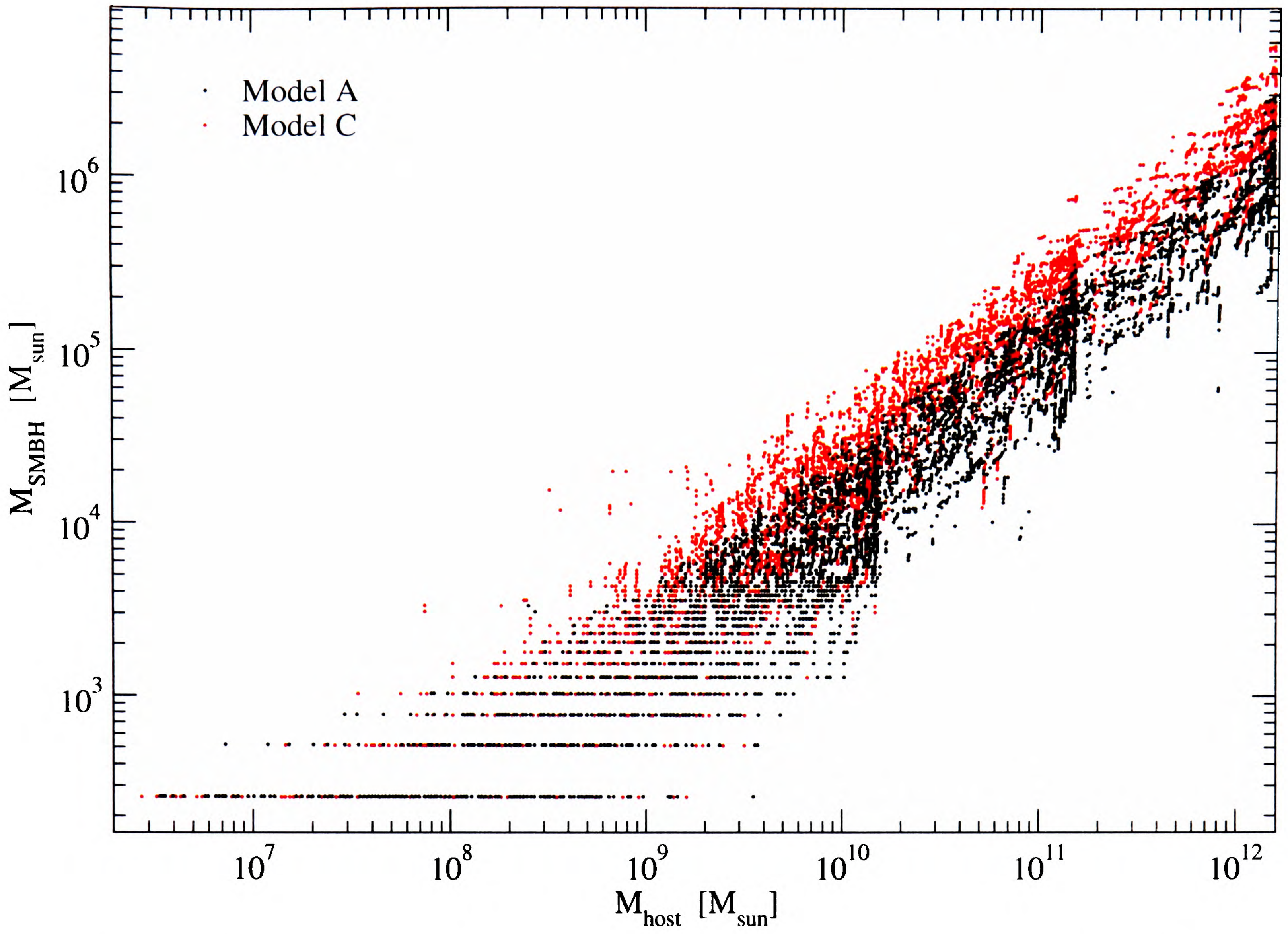


Figure 8.1: The figure shows the mass of the host halo versus the mass of the central (S)MBH. For the (S)MBH mass we have taken account of the slightly altered growth history due to the adding on of the dynamical friction time after satellite MBH systems have crossed the infall radius.

change in the growth history of the central (S)MBH and also to a slightly lower final mass of the SMBH, since there is now a small fraction of satellite MBH systems that do not merge with the centre within a Hubble time. Figure 8.1 shows the host halo mass versus the resulting mass of the central (S)MBH for final halo masses of $1.6 \times 10^{10,11,12}$ $M_{\odot}$and their lighter precursor halos at higher redshifts. The data points from precursors of different final halo masses are recognisable by their accumulation near those final halo masses. Nevertheless, all data points lie on a line with a slope consistent with the one we determined for the $M_{SMBH} - M_{bulge}$ relation in section 6.3. This indicates that the ratio of SMBH to halo mass does not depend significantly on the redshift.

We also made an important assumption regarding the actual mass of the satellite systems that the MBHs are embedded in. The semi-analytical code we used to follow the satellite dynamics considers every satellite naked if its mass has dropped to something less than about 0.3 % of its original mass [2]. For all 'naked' MBH satellite systems arriving at the infall radius we have therefore assumed a mass equal to this limit. We do impose a lower limit of $3 \times 10^3 M_\odot$for the mass of any naked satellite system, roughly corresponding to 1% of the mass of the MBC halos and large enough for the presence of a baryonic core of the order the MBH mass. It is this remnant material of the satellite associated with an MBH that allows MBHs to the delivered to the centre efficiently by dynamical friction. This point has also been mentioned by Yu (2002) .

For the halo of final mass 1.6×10^{12} $M_\odot$we find the following. Of the satellite MBH systems crossing the infall radius about 96 % do arive at the center within a Hubble time. These MBHs contain 99 % of the mass of all MBHs in the infall region. If we reduce the satellite mass to 0.03 % of its original value and the lower limit to $3 \times 10^3 M_\odot$we still find that 88 % of the MBHs in the infall region (accounting for 97 % of the mass) end up at the center within a Hubble time. In halos of lower final mass virtually all MBHs systems crossing the infall radius travel to the centre within a Hubble time.

Dynamical Friction and Gas Infall in Galaxy Mergers

MBHs with masses and orbits that would not allow them to travel to the centre within a Hubble time may still end up in the centre if the environment changes in a way that allows dynamical friction to act more efficiently. Mergers between galaxies of similar mass ('major mergers') through the accompanying violent dynamical processes may induce the dramatic changes to the matter distribution in the central region that we are looking for. Hydrodynamical simulations of mergers of gas rich galaxies (Barnes 2002;Barnes & Hernquist 1996;Naab & Burkert 2001) show that up to 60 percent of the total gas mass of the two galaxies can end up within a region just a few hundred pc across at the centre of the merger remnant, thereby triggering starbursts and initiating the fuelling of quasars[3].

This highly dense gaseous core remains until feedback from star formation, supernovae or quasar activity begins to drive it out again.

We have already identified gas infall as a potential mechanism for accelerating the evolution

---

[2] The actual criterion refers to a limiting tidal radius of a satellite system, beyond which it is considered 'naked'. This criterion roughly corresponds to the 0.3 % mass limit we used.

[3] The galaxy interaction leads to the shock heating of the gas and the formation of a bar in the gas distribution. Radiatively cooled gas then falls towards the centre along the bar.

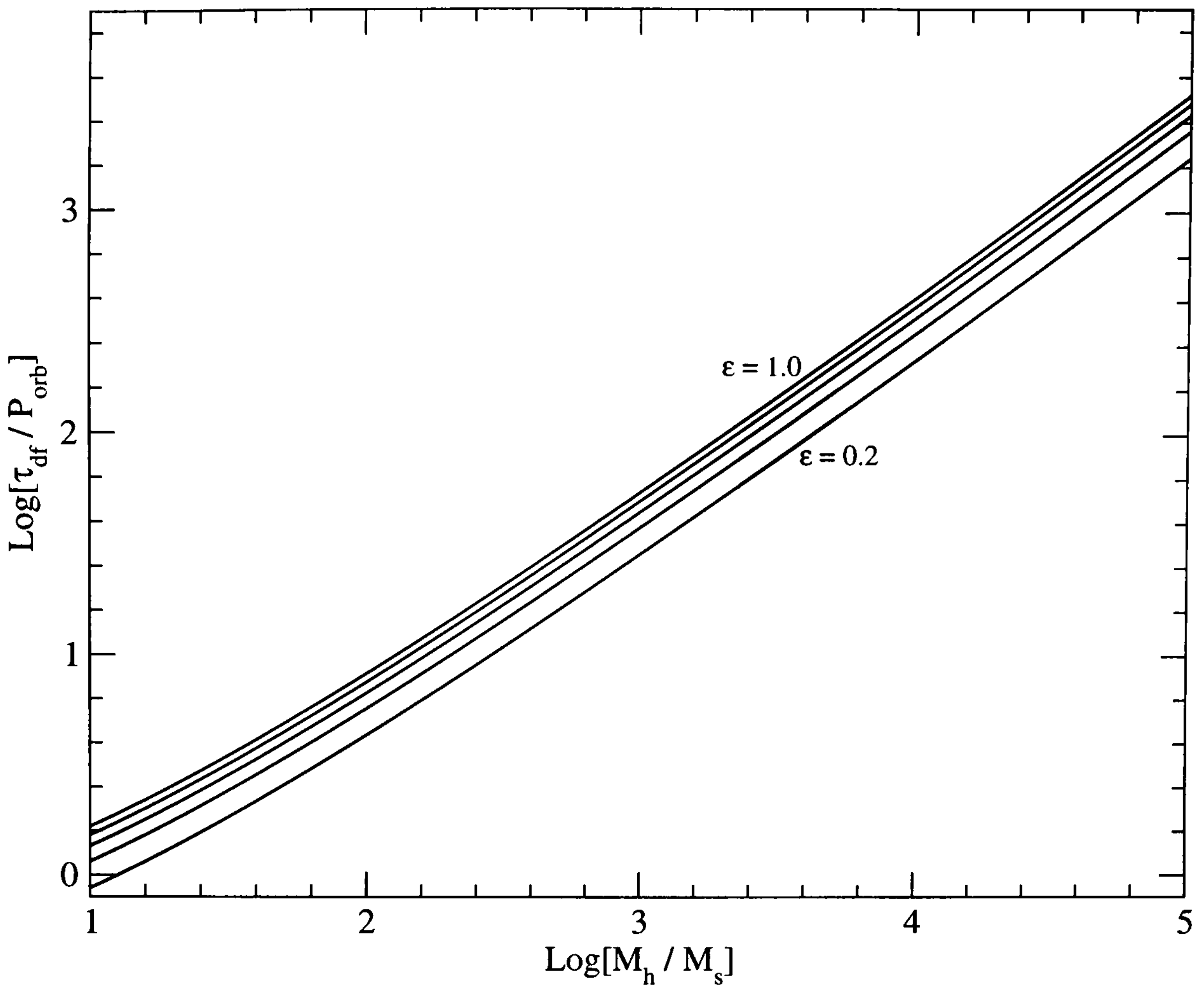


Figure 8.2: Ratio of dynamical friction time to orbital period plotted against ratio of host mass interior to satellite orbit to satellite mass. From top to bottom the curves are for values of the circularity parameter $\epsilon = 1.0, 0.8, 0.6, 0.4, 0.2$.

of a hard MBH binary. We now look at how this effect can also boost the efficiency of dynamical friction as MBHs move in the central region of the merger remnant.

If the respective central-kpc-region of each galaxy ends up within a kpc or so of the new centre of the remnant after merging, then so will any MBHs that were present within a kpc of the original galactic centres. This seems reasonable since the central regions, sitting deepest in the gravitational potential of each galaxy, will be most stable to tidal disruption. It is plausible that MBHs originally present there travel along the line connecting the centres of the merging galaxies, towards the new global potential minimum of the merger remnant.

This migration of MBHs could be accelerated by the rapid infall of large amounts of gas into a small core region. The question then is whether the density and the bulk inflow velocity of the gas is sufficient to accelerate any orbiting MBH towards the centre via dynamical friction. The complex dynamical events and rapidly varying geometry make it difficult to determine accurately any radial bulk inflow velocities and gas densities as they evolve with time and particularly so in the core central region, where things are made more difficult by the resolution limits of the simulations. However, even a rough estimate of these quantities may already yield useful constraints for our problem. In the following we try to estimate the time scale for infalling gas to carry two MBHs towards the centre by dynamical friction before the gas leaves the central region again.

For two disk galaxies, each having a total mass $M_{gal} = 2.75 \times 10^{11}$ M$_\odot$and gas mass $M_{gas} = 5.5 \times 10^9$ M$_\odot$in the disk and initial pericentric separation of $R_p = 8$kpc, BH96 find that roughly 60 percent of the total gas mass end up within a region approximately 100 pc in radius some 1.5 Gyrs later. Assuming the centre of the merger remnant will be half-way between the galaxies at initial separation and the gas flows towards this point along the line connecting the merging galaxy centres, the average inflow velocity of the gas is $v_{flow} \sim 4\text{kpc}/1.5\text{Gyr} \approx 2.7\ \text{kms}^{-1}$. Further we assume that the gas density in the flow is of the same order as the final gas density in the centre, i.e. $\rho_{flow} \sim 7 \times 10^9 M_\odot/(4 \times 100^3\text{pc}^3) \approx 1.6 \times 10^3 \text{Mpc pc}^{-3}$. Putting this into eq. 5.2 we find that the MBHs get accelerated to within 1% of v_{flow} in a time

$$\tau_{df,flow} \sim 2 \times 10^6 \left(\frac{M_\bullet}{260\ M_\odot}\right)^{-1} \text{yr} \tag{8.16}$$

Even for seed mass MBHs this time is negligible compared to the time scale of the inflow of about 1 - 2 Gyrs. Although we do not know what exactly will happen to the MBHs once they have entered the central 100 pc of the remnant, gas inflow seems at least capable of efficiently transporting MBHs from the outer radius of the infall region (~ 1kpc, for galaxies of this size) down to at least 100 pc. This holds in particular for MBHs that previously would not have been able to spiral anywhere near the centre.

Multiple MBH interactions

Up to now we have implicitly assumed that a MBH binary merges before it forms a new binary with another incoming MBH. However, given the number of inspiralling MBHs involved it is possible that there are cases where incoming MBHs will find another MBH binary that has not yet merged. If the binary system and the incoming third MBH come close enough then there is a possibility that one of the MBHs will be ejected. The required proximity of incoming MBH and MBH binary is facilitated primarily by major mergers. This will be the case particularly at high

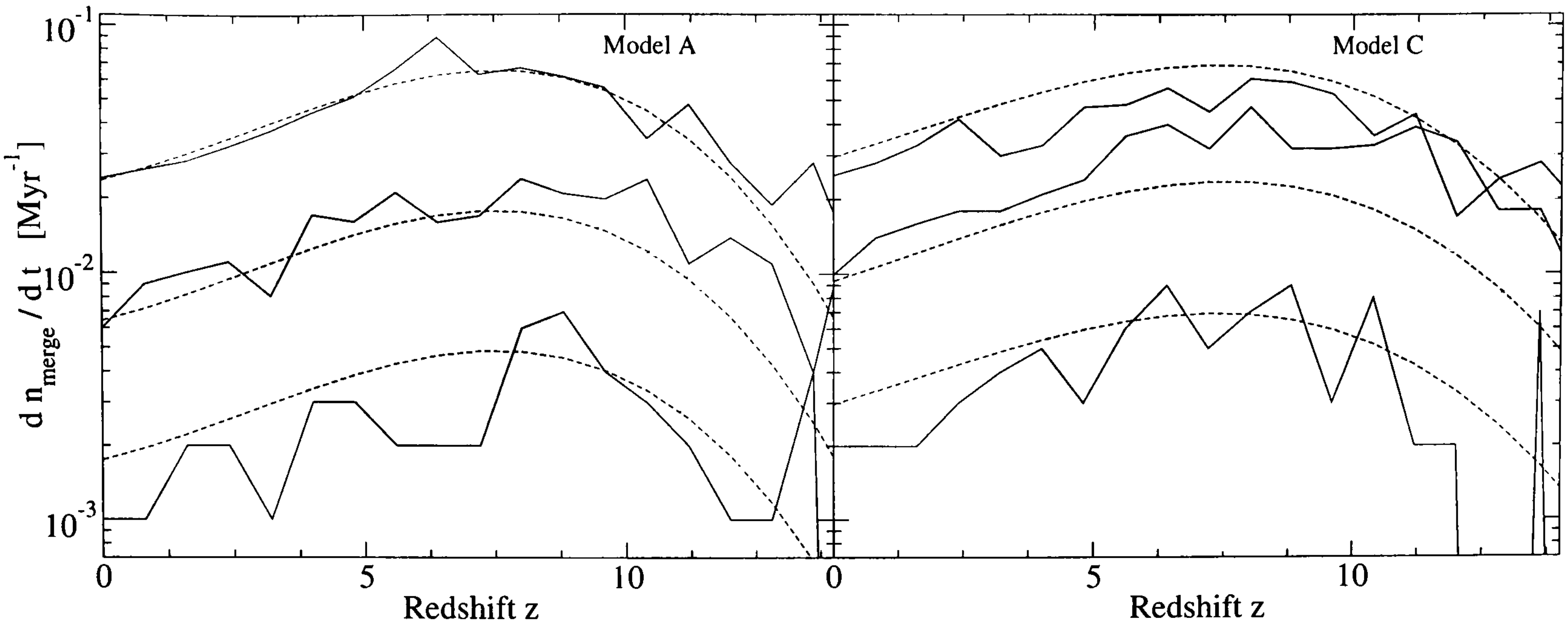


Figure 8.3: Rates of halo MBHs merging with central (S)MBH as a function of redshift. Results are shown for models A and C and final halo masses of $1.6 \times 10^{12,11,10} M_\odot$(top to bottom curves). The dashed curves represent the same best fits to the data scaled by a factor corresponding to the different final halo mass (see text).

redshifts, as most MBH interactions will be between MBHs with similar masses (of the order of the MBH seed mass), which also implies similar masses of their associated halos.

As far as our model is concerned we have not incorporated the possibility of triple interactions in our results. We assume that stellar and gas dynamical effects in the wake of major mergers always lead to swift merging of MBH binaries, primarily through accelerating evolution of the hard binary stage. Here we consider what happens if we drop this assumption. For the case that stellar and gas dynamics have no mitigating effect, we can determine a rough limit for the number of triple interactions and resulting sling-shot ejections of MBHs.

In figure 8.3 we show the number of mergers per unit time (10^6 yr) for model A and C. From top to bottom the curves are for $1.6 \times 10^{12,11,10} M_\odot$halos. The dashed lines are the best fits. Let us consider the largest merger rate, which we find at $z \sim 5 - 10$ in the precursors of halos with final mass $1.6 \times 10^{12} M_\odot$. The rate is $dN_{merge}/\ dt \sim 0.07\ \mathrm{Myr}^{-1}$ implying that two MBHs arrive at the centre within $\sim$ 15 Myrs. For the lowest rates $dN_{merge}/\ dt \sim 0.001\ \mathrm{Myr}^{-1}$ two MBHs would reach the centre within about a Gyr. These time scales would have to be compared with the hardening time of a MBH binary at the centre to determine the likelihood of a triple interaction. Yu (2002) finds that the hardening time scale can be significantly larger for equal mass binary systems, which again is more likely to be the case at very high redshifts.

Although the central halo MBH merger rate as shown in figure 8.3 is highly uncertain at high redshifts, there is a clear declining trend towards high redshifts. Consequently we would expect multiple MBH interactions to be less important. This is different from Volonteri et al. (2003) , who find that there is a very significant probability for triple interactions. One important reason for this discrepancy is that they explicitly did not take into account tidal stripping of satellite systems and the resulting increase in the dynamical friction infall time scale. For those MBHs that are involved in triple interactions they find that those MBHs that get ejected MBH will actually leave the galaxy (remnant) altogether (Volonteri, Haardt & Madau 2003). These ejected MBHs could thus constitute a population of inter-galactic MBHs. Due to the nature of the process these MBHs will be mostly seed mass MBHs and will have been stripped naked, so that they will be virtually impossible to detect in the IGM.

The overall result is then, that while the formation of multiple MBH systems is possible, the probability of such events is likely less important than analytical arguments suggest. Previous estimates typically did not account for the tidal stripping of infalling satellites let alone their detailed dynamical evolution, but also the role of gas dynamics at the centre. The first effect implies that the rate at which MBHs arrive from the outer parts of the halo at the halo centre is lower , whereas the latter means that once MBHs have arrived at the centre, they form a binary that coalesces faster.

In the light of this we believe our assumption of efficient merging to be justified, particularly at redshifts $\lesssim$ 15 where MBHs mergers actually produce a strong enough gravitational wave signature that could be detected as we will see in the next section.

8.1.4 Rate of SMBH-MBH mergers

We now attempt to obtain the cumulative signal of all mergers. To do this we essentially need to integrate over the mass function of halos times the number of mergers per halo and finally integrate again over the relevant redshift ranges.

In what follows we will only consider models A and C. In model D the rate of mergers is too low to derive a meaningful estimate of the average merger rate per halo, and consequently the total rate of events received from across the sky. Model B only differs from A in that it has a

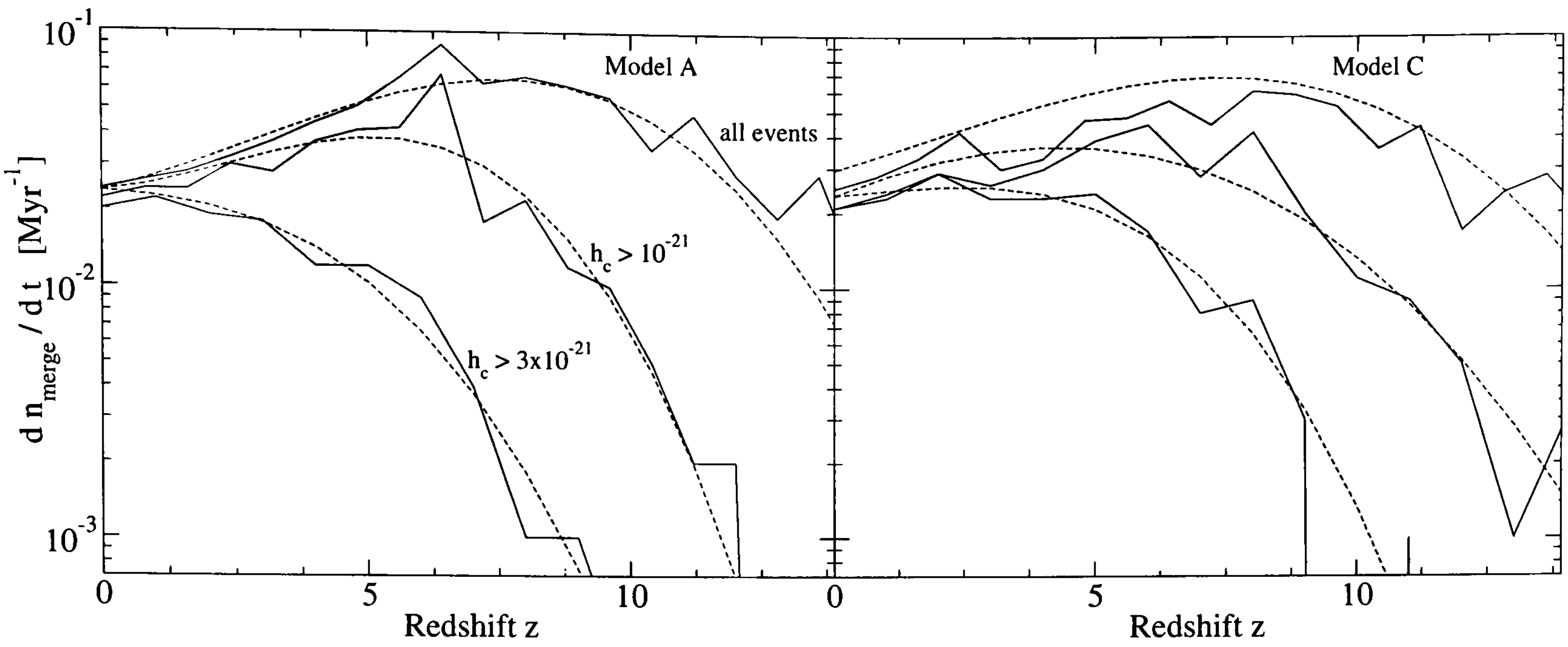


Figure 8.4: Rates of halo MBHs merging with central (S)MBH as a function of redshift. Results are shown for models A and C and final halo mass of $1.6 \times 10^{12} M_\odot$with the dashed line representing a best fit. From top to bottom the sets show the contribution of events with dimensionless strain amplitudes above a given value as shown.

larger seed MBH mass. We do not expect this to make a significant difference when it comes to mergers with the centre. When calculating the dynamical friction time from the point the MBH systems cross the infall radius (c.f. section 8.1.3), we assumed the MBH satellite system's mass to be ~ 0.3 % of the original satellite mass. The difference in the satellite's MBH mass in model A and C is thus insignificant. The larger MBH mass in model B will also only lead to very slightly increased strain amplitudes when MBHs eventually merge.

The number of MBH mergers per halo is shown in figure 8.3 for models A and C. The dashed lines represent a single cubic-polynomial fit in log space that is multiplied with a simple scaling factor for different halo masses.

$$\frac{\mathrm{d}n_{merge}}{\mathrm{d}t}(M_{h,0}, z) = \left(\frac{M_{h,0}}{1.6\times10^{12}M_\odot}\right)^{\beta} \exp[\mathrm{b}_0 + \mathrm{b}_1\, z + \mathrm{b}_2\, z^2 + \mathrm{b}_3\, z^3]\ \mathrm{Myr}^{-1} \qquad (8.17)$$

where $M_{h,0}$ denotes final halo mass. For model A we find $(\mathrm{b}_0, \mathrm{b}_1, \mathrm{b}_2, \mathrm{b}_3, \beta) = (-3.75, 0.12, 0.022, -2.63\times 10^{-3}, \frac{5}{9})$ and $(-3.52, 0.145, 7.5\times10^{-3}, -1.57\times10^{-3}, \frac{1}{2})$ for model C. In this form the merger rate can now simply be divided into a product of a redshift dependent part, $u(z)$, and another one scaling with final halo mass

$$\frac{\mathrm{d}n_{merge}}{\mathrm{d}t}(M_{h,0}, z) = \left(\frac{M_{h,0}}{1.6\times10^{12}M_\odot}\right)^{\beta} u(z) \qquad (8.18)$$

Figure 8.4 shows the merger rate per halo with strain amplitudes $h_c > 10^{-21}$ and 3×10^{-21} for a final halo mass of 1.6×10^{12} $M_{\odot}$only. While the best fit parameters change accordingly, we have assumed that the scaling with halo mass as parametrised by β remains the same. Although we have not explicitly stated any error bars, uncertainties will be quite large, since the merger rate per halo is determined by averaging over individual merger events occurring over large time scales. The only way of reducing this uncertainty is to run a significantly larger number of merger trees.

When determining h_c we assume that the merging MBHs emit waves at the maximum frequency which is only expected to occur very briefly towards the end of the inspiral phase due to the rapid sweeping up in frequency. This, however, should not affect our result significantly, since h_c depends only weekly on the frequency (c.f. equation (8.2)).

Using the merger rate per halo we can already determine as a function of epoch a lower limit to the rate, $\mathrm{d}\nu_{rec}$ of received SMBH - MBH merger events.

$$\begin{aligned} \frac{\mathrm{d}\nu_{rec}(z)}{\mathrm{d}V} &> \frac{1}{1+z}\int_{M_{h,min}}^{\infty} \frac{\mathrm{d}n_h}{\mathrm{d}M_{h,0}} \frac{\mathrm{d}n_{merge}}{\mathrm{d}t}(M_{h,0}, z)\ \mathrm{d}M_{h,0} \\ &= \frac{u(z)}{1+z}\int_{M_{h,min}}^{\infty} \frac{\mathrm{d}n_{h,0}}{\mathrm{d}M_{h,0}} \left(\frac{M_{h,0}}{1.6 \times 10^{12} M_{\odot}}\right)^{\beta} \mathrm{d}M_{h,0} \end{aligned} \tag{8.19}$$

where V denotes volume and (1 + z) accounts for the redshift induced reduction in the rate of received events. $\mathrm{d}n_h/\mathrm{d}M_{h,0}$ is the differential mass function of halos of mass $M_{h,0}$ (c.f. eqn. 2.52 from chapter 2)

$$\frac{\mathrm{d}n_h}{\mathrm{d}M_{h,0}} = \sqrt{\frac{2}{\pi}}\frac{\rho_0}{M_{h,0}}\left|\frac{\mathrm{dln}\sigma}{\mathrm{d}\ \mathrm{ln}M_{h,0}}\right| \frac{\delta_c}{\sigma(M_{h,0})}\exp\left[\frac{-\delta_c^2}{2\sigma(M_{h,0})^2}\right]$$

where ρ_0 is the present day cosmic matter density, $\sigma(M_{h,0})$ is the variance in the linear matter density field on scales corresponding to mass $M_{h,0}$ and δ_c is the critical overdensity for collapse.

Equation 8.19 is a lower limit as it only considers mergers in the main trunk of the tree that grows to become a single halo of mass M_h at $z = 0$. To account for the mergers in side branches before they become incorporated in the main trunk of the tree the PS mass function, which we had only computed for final halo masses above, needs to be replaced by the redshift dependent mass function. In addition we need to re-express the redshift dependent MBH merger rate per halo in terms of the halo mass at the redshift concerned. To this end we use the relation

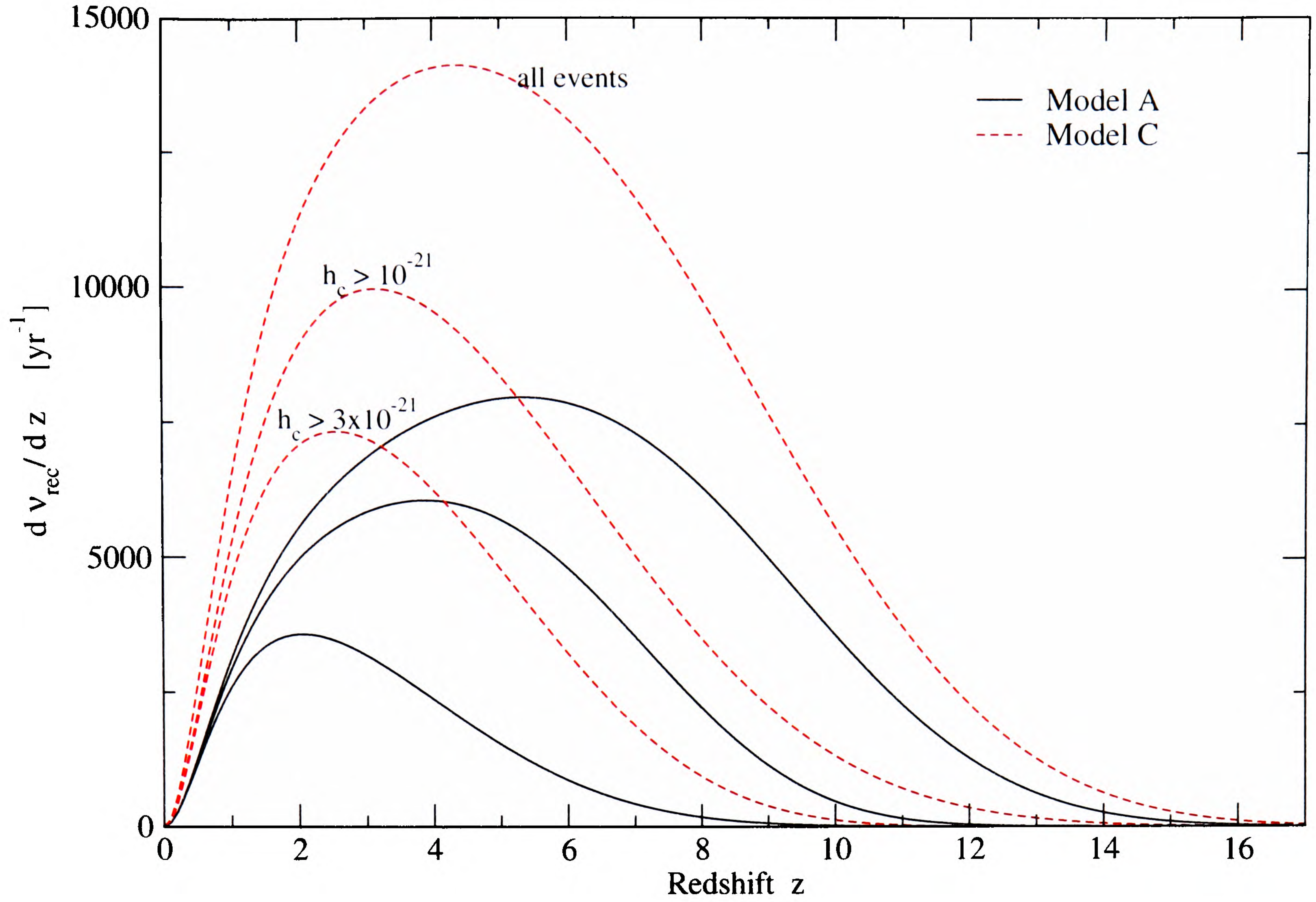


Figure 8.5: Redshift dependent rate of received merger events for models A and C. From top to bottom the sets show the contribution from all events and those with dimensionless strain amplitudes above 10^{-21} and 3×10^{-21}.

$M_h = M_h(M_{h,0}, z)$ (eq 5.1) in eq 8.18 to eliminate the dependence on $M_{h,0}$ and obtain

$$\frac{\mathrm{d}n_{merge}}{\mathrm{d}t}(M_h, z) = \left[10^{-7} \times \left(\frac{M_h}{1.6 \times 10^5 M_\odot}\right)^{\frac{1}{1-0.04(z-0.01\ z^2)}}\right]^{\beta} u(z) \qquad (8.20)$$

The lower integration limit in eq 8.19 is determined by the minimum mass of halos that harbour central MBHs. This mass has to be at least of order 10^7 $\mathrm{M}_\odot$,since we argued above that only at this mass has every halo acquired one seed MBH. This is important since we have implicitly identified the number density of halos with that of central MBHs.

If we now multiply $\frac{\mathrm{d}\nu_{rec}(z)}{\mathrm{d}V}$ with the volume of a spherical shell at radius $D(z)$ we obtain the total rate of events, d $\nu_{rec}(z)$, received from sources at distance $D(z)$ to $D(z) + \mathrm{d}D$

$$\mathrm{d}\nu_{rec}(z) = \frac{\mathrm{d}\nu_{rec}(z)}{\mathrm{d}V} 4\pi D(z)^2 \mathrm{d}D \qquad (8.21)$$

The resulting differential redshift distribution of the received rate of events is shown in figure

8.5, where we have also shown the contribution from events with associated dimensionless strain amplitudes h_c above 10^{-21} and 3×10^{-21}. The latter is based on the fits to the merger rate per halo shown in figure 8.4. Again, the errors are quite large and are difficult to quantify precisely, primarily because of the uncertainties in the merger rate per halo to which we now also have to add the uncertainty in the halo mass scaling, which we described by a power law with index β.

This differential distribution can then be integrated over all redshifts to obtain the total number of merger events received.

8.1.5 Detections and the distribution of strain amplitudes

For actual detections with LISA we need to take into account its restrictions on the minimum detectable strain amplitudes. For the MBH masses under consideration here, and the resulting wave frequencies, LISA should be able to see events certainly with $h_c \sim 10^{-21}$ and as weak as $h_c \sim 10^{-23}$ near $f_{gw} \sim 0.01$ Hz.

For the combination of MBH masses and the redshift range considered, all events yield strain amplitudes larger than 10^{-23}, and LISA may therefore be able to detect most if not all of these events. Adopting the more conservative LISA limit of $h_c \sim 10^{-21}$ the number of detections is somewhat smaller as shown by the corresponding curve in figure 8.5.

We are interested in detections of events at cosmological distances and to do so the conservative LISA detection limit requires a minimum chirp mass of about 10^4 $M_\odot$ (Haehnelt 1994). MBHs of this size, however, will typically be hosted inside halos of at least 10^9 $M_\odot$, which could therefore be taken as the effective lower limit of integration in equation 8.19. In fact $d\nu_{rec}$ is not very sensitive to the particular choice of integration limits. The event rate of mergers in halos as determined by our simulations is very small at the low mass end which therefore does not contribute significantly to the overall rate. Similarly the steep decline in the mass function at high masses imposes an effective upper limit.

The redshift distribution of merger events can now be integrated across the redshift range to obtain the total number of events received. The result is shown in figure 8.6, where we have shown the total number of events above a given strain amplitude. As in the previous section, we have used a fit to the merger rate per halo above a given strain amplitude to determine the

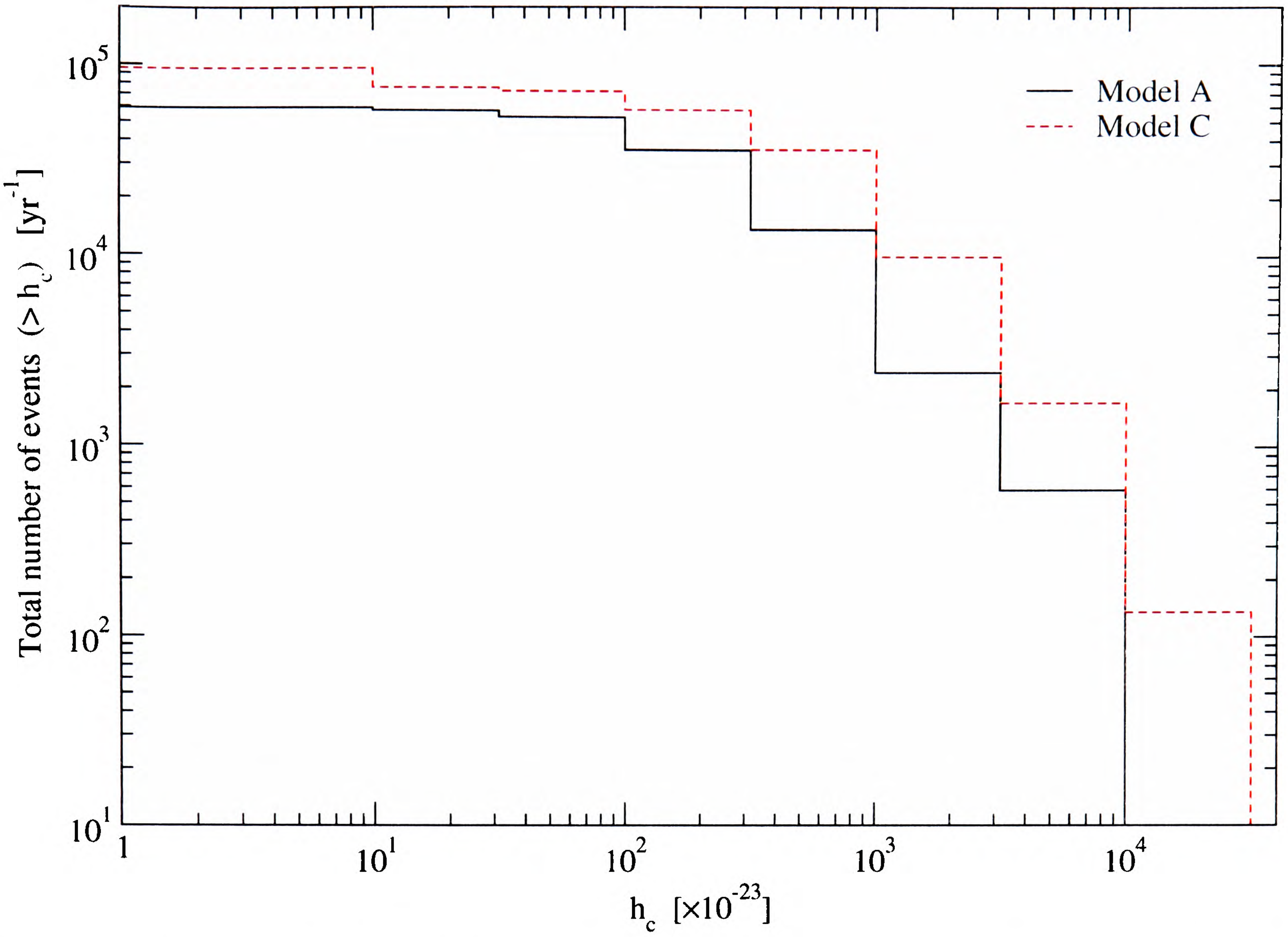


Figure 8.6: Distribution of dimensionless strain amplitudes for events received from across all redshifts.

merger rate per unit redshift. From this we see that LISA would be expected to detect some $10^4 - 10^5$ events per year. The difference between model A and C in this respect is too small to be detectable in practise as the overall uncertainties in our results are at least of the same order.

We have seen above that the strain amplitude is rather insensitive to the frequency of the waves emitted. The frequency does matter for LISA detections as its peak sensitivity lies in a rather narrow band around 0.01 Hz. Merging seed MBH binary systems have very low strain amplitudes, especially at high redshifts, but peak frequencies well above 1 Hz. The only way of detecting these is during a significantly earlier phase of their coalescence when their frequency is lower.

8.2 Observations of Gamma Ray Bursts from exploding population III stars at high z

An exciting prospect is the possibility that hypernovae concluding the lives of the very first and massive stars at $z \gtrsim 20$, may be associated with gamma-ray bursts (GRBs). These could potentially be observed with upcoming satellite missions and would so allow to observe the epoch of first star formation itself. In contrast to the other observations mentioned above, this cannot be used to verify the hierarchical MBH merging scenario. Instead GRB observations could possibly provide constraints on the abundance of the first massive stars and in this way allow to directly test this part of our model.

GRBs are the brightest explosive events in the Universe. Typically they involve the emission of large amounts of radiative energy of the order $\sim 10^{52-54}$ erg s in γ rays, over a period of only a few seconds. A GRB is followed by an afterglow in the X-ray, optical or radio parts of the spectrum and lasts from days to weeks. The total energy emitted in an afterglow is typically two or three orders of magnitude below that of the actual burst. Spectroscopy has been carried out on the afterglow emissions as well as the emission of what appear to be host galaxies of the GRBs once the afterglow itself has faded away sufficiently. In this way it has been possible to constrain the redshift of a number of GRBs below $z < 3$. If GRBs could be used as standard candles in a similar way as type Ia supernovae this would facilitate their use to calibrate the GRB distance - luminosity relation. On this basis one could then determine the maximum redshifts at which GRBs could still be detected by present and future observing missions, such as with the BATSE detector on the *Compton Gamma-Ray Observatory* or the *SWIFT* satellite due to be launched late in 2003 (Lamb & Reichart 2000). For a number of GRBs for which redshifts could be established Lamb & Reichart (2000) found that *SWIFT* could have in principle detected these out to redshifts of $z \sim 20$ and some to $z \sim 70$. For further details of GRBs their afterglows the reader is referred to the review by Piran (2000) and the extensive work by Bloom (2002) and references therein. For the following we assume that GRBs are in principle detectable out to redshifts beyond $z > 20$.

A range of processes have been discussed that could be the cause for GRBs. As mentioned, one of these is the possibility that GRBs are produced by the collapse of massive stars. In the following we exclusively focus on this possibility and attempt to estimate the number of GRBs from massive population III stars in our model.

8.2.1 Total number of GRBs

The procedure to compute the total number of GRBs across all relevant redshifts is very similar to the one we used to determine the gravitational wave events. It is different in that we only deal with the very first stars and are not concerned with the subsequent merging of the MBHs these stars leave behind. This is therefore a purely analytical argument with no recourse to our simulation data concerning the merging and dynamical evolution of MBHs.

Again we estimate the abundance of MBC halos at the redshifts in question ($z \sim 25$). Depending on how many massive stars form per halo we can then determine the expected number of GRBs. A similar route has been followed in previous work for the determination of GRBs at $z \lesssim 5$. In this case the abundance of massive stars per halo/galaxy was assumed to trace the star formation rate at the respective redshift (Bromm & Loeb 2002;Roy Choudhury & Srianand 2002). We use the mass function of massive stars discussed in chapter 4. To start with we assume that every MBC halo between a maximum redshift z_{max}, corresponding to a $2.5 - 3.5\sigma$ peak, and a minimum redshift determined by the end of first star formation, $z_{min} \sim 20$ (c.f. section 4.3.5), produces one star that explodes as a hypernova with an associated GRB.

We have seen that a hypernova leads to the expulsion of all gas from MBC halos. Each MBC halo can thus only ever produce one GRB. This is a problem as we now require each halo to retain a memory of whether and when it produced a GRB. For this reason it is not strictly correct to use the Press-Schechter mass function to compute the number density of halos at a given redshift and equate the result with the number density of GRBs. A halo more massive than M_{mbc} may have already lost all its gas through a hypernova at some earlier stage when it was significantly less massive. In practice, however, the Press-Schechter approach is more than adequate for two reasons. First, the PS mass function is steepest for large masses corresponding to halos collapsing from high peaks as is the case for all MBC halos in our redshift range. The number density of halos with mass $M_h \geq M_{MBC}$ is therefore vastly dominated by halos with masses very close to

M_{MBC}. Secondly, the survival time [4] for halos collapsing from high peaks is very short (Lacey & Cole 1993). That means that by far most halos with $M_h \geq M_{MBC}$ will have formed at or very shortly before the redshift in question and the probability of previous star formation and a GRB is very small.

If we identify a GRB event with the collapse of a MBC halo, the rate of GRBs at any given redshift is then approximately

$$\frac{\mathrm{d}\nu_{GRB}}{\mathrm{d}z}\,\mathrm{d}z \approx 4\pi D(z)^2\,\mathrm{d}D(z)\int_{M_{MBC}}^{\infty}\frac{\mathrm{d}^2 n_h}{\mathrm{d}t\;\mathrm{d}M}\,\mathrm{d}M \tag{8.22}$$

where we have used the rate of change of the MBC halo mass function

$$\begin{aligned}\frac{\mathrm{d}^2 n_h}{\mathrm{d}t\;\mathrm{d}M_h} &= \left(\frac{2}{\pi}\right)^{1/2}\frac{\rho_0}{M_h^2}\left|\frac{\mathrm{d}\ \ln\sigma}{\mathrm{d}\ \ln M_h}\right| u\ \mathrm{Exp}\left[-\frac{1}{2}u^2\right]\left[\frac{1}{u}\frac{\mathrm{d}u}{\mathrm{d}t}-u\frac{\mathrm{d}u}{\mathrm{d}t}\right] \\ &= \frac{\mathrm{d}n_h}{\mathrm{d}M_h}\frac{\mathrm{d}u}{\mathrm{d}t}\left[\frac{1}{u}-u\right]\end{aligned} \tag{8.23}$$

and $u = \delta_c/\sigma(M_h) * D_{grow}(t)$. We have shown the mass function and its rate of change in figure 8.7.

8.2.2 Rate of GRBs received

For GRBs at the very high redshifts we are considering, no afterglow emission can be detected anymore. An accurate redshift determination is thus not possible unless a unique distance - luminosity relation for GRBs could really be established. If all GRBs do have the same total luminosity and peak photon flux we have a unique distance - luminosity and thus redshift luminosity relation. Eqn (8.23) above can then be converted into a flux distribution $F_{GRB}(z)$ using $\mathrm{d}z \propto -2\pi D(z)^3/(10^{52-54}\ \mathrm{erg})(\ \mathrm{d}D/\ \mathrm{d}z)^{-1}\ \mathrm{d}F_{GRB}$.

However, the case for a unique GRB distance-luminosity relation has yet to be made. To test our model we therefore focus on a prediction of the rate of GRBs to be compared with those detected by e.g. the *SWIFT* mission. We integrate eq. (8.23) over the redshift range of population III star formation to arrive at the total rate of GRBs received. This is shown in figure 8.8, together with rate of GRB events originating from any given redshift interval for

---

[4] The survival time is the time by which the mass of a halo has more than doubled.

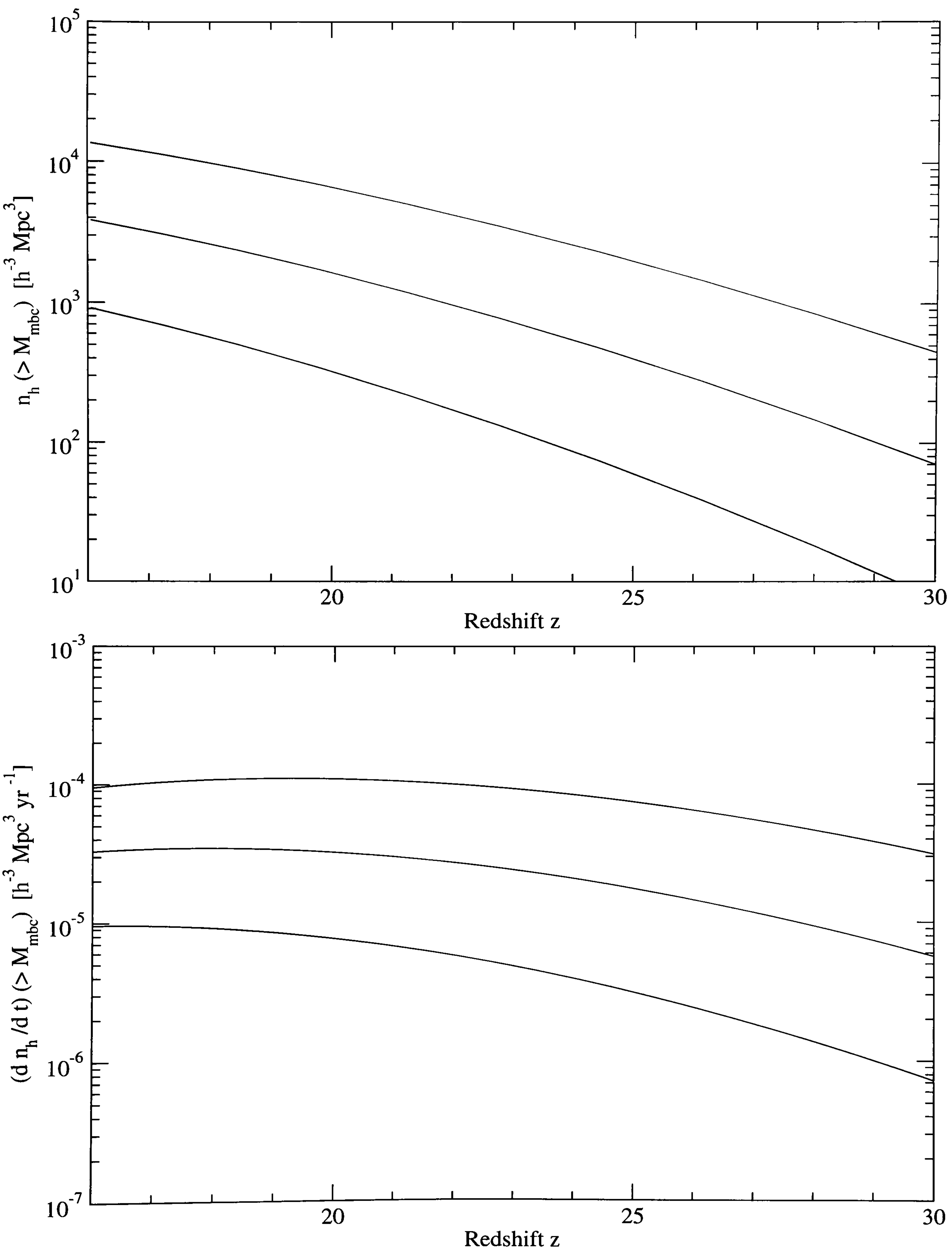


Figure 8.7: *Top*: Redshift dependent comoving number density of halos with mass $M_h > M_{MBC}$. From top to bottom the curves correspond to a $M_{MBC} = \{1, 3, 10\} \times 10^5 M_\odot h^{-1}$. *Bottom*: Rate of change of halo MBC halo number density. The curves are in the same order.

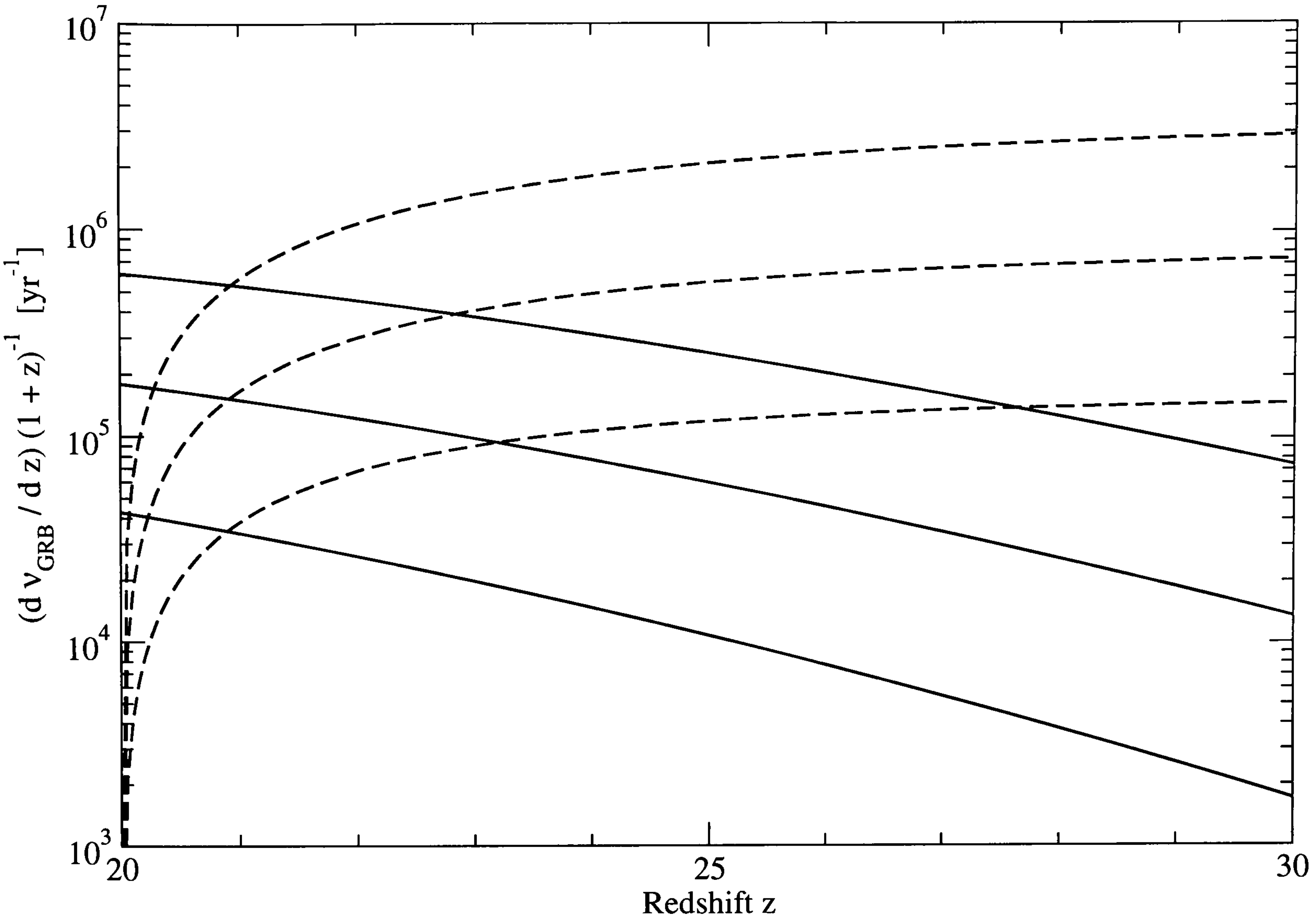


Figure 8.8: Number of GRB events received per year originating from redshift z to $z +$ dz. The dashed curve shows the total number of GRB events from $z \geq 20$. From top to bottom the curves correspond to a $M_{MBC} = \{1, 3, 10\} \times 10^5 M_{\odot} h^{-1}$.

three different values of M_{MBC}, assumed constant over the redshift range $20 < z < 30$. We find that the total number of GRB events received per year from this redshift range varies between $\nu_{GRB} \approx 3 \times 10^6 - 2 \times 10^5 \mathrm{yr}^{-1}$ for $M_{MBC} = 10^5$ to 10^6 M$_{\odot}$respectively. Even for $M_{MBC} = 3 \times 10^5$ M$_{\odot}$at $z = 25$, it may still make more sense to consider a larger halo mass, since by definition baryons - and therefore stars - in MBC halos only cool on a Hubble time scale, whereas we have assumed that GRB events coincide with the collapse of the halos within which they occur. At $z = 25$, for instance, a halo mass three times larger than the MBC halo mass we used would result in a cooling time that is about a tenth of the Hubble time at that redshift and so more in line with this assumption.

$$\nu_{GRB,rec} = \int_{z_{max}}^{z_{min}} (1+z)^{-1} \frac{\mathrm{d}\nu_{GRB}}{\mathrm{d}z}\, \mathrm{d}z \tag{8.24}$$

A conservative estimate for the total number of GRB events received is $\nu_{GRB} \approx 2 \times 10^5 \mathrm{yr}^{-1}$, assuming that halo collapse, star formation and GRB are coinciding and that first star formation ends at $z = 20$. How many of these could actually be observed depends on whether or not GRBs are luminous enough to still be detected at these distances.

8.3 Other detections

8.3.1 Gamma rays from MBH mergers

If a small fraction of the gravitational energy from MBH mergers is released in the form of gamma rays - maybe due to the fast accretion of large amounts of material in the very close vicinity of the MBHs - we may be able to see these. For an inspiralling MBH binary the energy radiated away in gravitational waves, L_{gw} is (Shapiro & Teukolsky 1983)

$$\begin{aligned} L_{gw} &= \frac{32}{5}\frac{G^4}{c^4}\frac{M_{12}^3\,\mu_{12}^2}{a_0^5} \\ &= 1.63 \times 10^{49} \left(\frac{a_0}{10^8\ \mathrm{m}}\right)^{-5} \left(\frac{M_{12}}{10^3\ \mathrm{M}_\odot}\right)^3 \left(\frac{\mu_{12}}{10^3\ \mathrm{M}_\odot}\right)^2 \ \mathrm{erg\ s^{-1}} \end{aligned} \tag{8.25}$$

and combining this with the characteristic merger time scale eq 8.7 this becomes

$$L_{gw} = 6.4 \times 10^{58} \left(\frac{\tau_0}{\mathrm{s}}\right)^{-5/4} \left(\frac{M_{12}}{10^3\ \mathrm{M}_\odot}\right)^{1/2} \left(\frac{\mu_{12}}{10^3\ \mathrm{M}_\odot}\right)^{3/4} \ \mathrm{erg\ s^{-1}} \tag{8.26}$$

If we now multiply this with $1/f_{max}$ (eq 8.6) we can estimate the energy radiated away on the last orbit before final coalescence.

$$L_{gw} \approx \times 10^{59} \left(\frac{M_{12}}{10^3\ \mathrm{M}_\odot}\right)^{1/4} \left(\frac{\mu_{12}}{10^3\ \mathrm{M}_\odot}\right)^{3/4} \ \mathrm{erg} \tag{8.27}$$

It is interesting to compare this with GRBs. A typical GRB lasts between a fraction of a second to minutes and emits total energies between $10^{51} - 10^{54}$ erg (Piran 2000) and about a factor 100 less if we assume beamed emission. As far as the duration of GRBs is concerned this is matched by the period of the last orbits of MBH binaries with mass $M_{12} \sim 1000\ \mathrm{M}_\odot (\tau_0 = 1/f_{max} \sim 0.1$ seconds) to $M_{12} \sim 10^6\ \mathrm{M}_\odot ((\tau_0 \sim 100$ seconds). This means that if some fraction $\epsilon_\gamma \lesssim 10^{-10}$ of gravitational wave energy is emitted in gamma rays, merging MBHs could account for some (beamed) GRBs. At the lower limit the detection limit of the *BATSE* instrument on the *Compton Gamma Ray Observatory*[5] implies $\epsilon_\gamma \gtrsim 10^{-14}$ for MBH mergers to be detected out to cosmological distances. For isotropically emitting GRBs ϵ_γ would be a factor 100 or so higher.

8.3.2 Flares from the tidal disruption of stars

In addition to accretion from the ISM, MBHs could in principle also emit radiation as a result of an encounter with stars in the bulge or disk. The case of MBHs feeding on a captured binary through an accretion disk is probably ruled out. For the MBH to binary mass ratios we are considering (e.g. $M_{mbh}/M_\star \approx 10^{4-6}$ for a MBH with a one solar mass star companion) the tidal forces due to the MBH are so strong at the Roche radius that the stellar mass companion would be tidally disrupted and a Roche lobe would not actually be established [6]. However, this tidal disruption would also create a very strong but short term 'X-ray flare' with a largely thermal spectrum peaking in the UV/soft X-ray (Komossa 2001).

We have investigated the possibility of X-ray flares generated in the tidal disruption of stars in the bulge through MBHs orbiting there. These flares are extremely luminous (up to $\approx 10^{44}\mathrm{ergs}^{-1}$) but last only a few weeks or so and are followed by a longer term lower level emission from the debris of the disrupted star that gets accreted from close orbits around the MBH. In this way some 10 to 50 percent of the gaseous debris of the star accretes onto the MBH with the rest getting unbound. This process of tidal disruption has already been computed for stars in the vicinity of the central SMBH (Rees 1988; Syer & Ulmer 1999; Magorrian & Tremaine 1999) and potentially been observed (Komossa 2001). In principle the extension to off-centre MBHs is straightforward. The event rate is directly proportional to the density of stars at some tidal radius r_t of the order of 10^{4-7} kilometres from the black hole depending on its mass. For a central SMBH r_t also coincides with the distance from the bulge density distribution. The latter has a power law profile and so for MBHs orbiting at distances of 1000 pc, say, from the centre the number density of stars is lower by some 9 - 12 orders of magnitude, assuming a $1/r$ profile. The number of X-ray flare events is lower by the same factor and thus negligible if for a central SMBH a few to a hundred events are observed per year. Even if we assume there are hundreds of MBHs in the central region

---

[5] *BATSE*'s burst sensitivity is quoted as $3 \times 10^{-8}\mathrm{erg\ cm}^{-2}$.

[6] For MBHs more massive than this stars would be swallowed whole.

that have not merged to form a single SMBH the odds are improved by the same factor, but the resulting event probability is till far too low..

8.3.3 Microlensing

It is not clear if microlensing could yield a significant detection of MBHs. A fundamental quantity in the detection of microlensing events is the time it takes the lensing object to cross its own lens radius ('Einstein ring radius'). This timescale governs the variability of the flux from a background light source lensed by an MBH. The crossing time depends on a number of parameters such as the mass of the lens, the distance between us, the lens, the lensed object and the relative velocities of all three. For MBHs this crossing time is typically too long to be observed, certainly for MBHs in our own galactic halo. It may be possible, though, to pick up flux variability from MBHs crossing our line of sight to an AGN in the MBH's own galaxy. In this case both, lens and lensed object are at similar but large distances from us and the crossing time is proportional to the square root of the distance between the MBH and the AGN. To maximise the probability of observing such an event, we are therefore looking for MBHs crossing very closely in front of their galaxy's AGN in a large number of galaxies. However, any such observation is further complicated if the AGN displays intrinsic variability on a timescale similar to that of the crossing time.

8.4 Summary

In this chapter we were mainly concerned with probing our model at high redshifts. The merging of MBHs in the context of the hierarchical merging of galaxies and halos would lead to the emission of gravitational waves that could be detected with the LISA mission due to be launched in the next decade. We can draw the following conclusions:

- The merger rate per unit redshift peaks at $z \sim 4-6$. The maximum shifts to lower redshifts if events with higher dimensionless strain amplitudes are considered.
- All merger events have strain amplitudes above $h_c > 10^{-23}$ and most above 10^{-22}, so that LISA should be able to detect most of these.
- The total number of MBH merger events received are about 10^4 – 10^5 per year. This is at

least an order of magnitude larger than typical estimates for ('major') mergers of SMBHs in hierarchical structure formation scenarios. Our larger number arises from the inclusion of a lot of mergers with fairly large binary mass ratios, i.e. what we may qualify as 'minor' mergers.

- In contrast to previous analytical work our predictions are based on a more accurate and explicit modelling of the dynamical evolution of MBHs in galactic halos. This has allowed us to follow MBH merging in halos at higher redshift and therefore also down to lower precursor masses.

Since we have assumed efficient MBH merging, the calculated merger and detection rates only provide upper limits. In addition, since we consider MBH merging, the resultant wave amplitudes are expected to be smaller by one or two orders of magnitude than those (albeit much rarer) events involving SMBHs.

If the collapse of the first massive population III stars into MBHs is accompanied by gamma ray bursts these might allow us to probe the abundance of the first stars

- We predict $10^5 - 10^6$ GRBs per year originating at redshifts higher than 20, if there is one massive star per star-forming (MBC) halo associated with a GRB.

We have also briefly looked at other possible observational signatures of both the present abundance and mergers of MBHs.

- A fraction of the gravitational wave energy in MBH mergers may be released in gamma rays, e.g. through interaction with remnant gas in the close vicinity of the coalescing binary. If GRB emissions are beamed this fraction must be in the range $10^{-14} < \epsilon_\gamma < 10^{-10}$ for the resulting bursts of gamma rays to be detectable at cosmological distances but still to be smaller than those of proper GRBs. For isotropic GRB emission ϵ_γ is about a factor 100 larger. However, the short event duration makes it difficult to identify them unless a GRB style after-glow is associated with them.

- The likelihood of tidal disruptions of host stars in the vicinity of MBHs orbiting in the central few kpc or so of a host halo is negligible. If MBHs are associated with a dense star cluster, which might be a remnant of a previously more extensive satellite system, there is a greater probability of tidal disruptions occurring.

- Prospects of MBH detection through microlensing are very uncertain. While the lensing timescale is typically too long for MBHs, it might just become short enough in the case where MBHs in other galaxies pass closely in front of their galactic AGN as well as right across our line of sight to that AGN.

While the prediction of gravitational wave events and GRBs both depend sensitively on the initial assumption of massive star formation, the additional assumptions of efficient merging for gravitational waves and that GRBs are created in the process of massive star collapse, further increase uncertainty. All these assumptions are in principle verifiable primarily with improved numerical simulations [7]. For the time being, the prediction of event rates of 10 to 100 a day should provide motivation to look out for them in present and upcoming observations. Qualitatively the most important thing is to note, that we could potentially observe the highly energetic events accompanying both the formation (via GRBs) and merger (by gravity waves) of MBHs, and therefore probe the high redshift regime.

---

[7] The role of gas in BH binary mergers, improved resolution in star formation and collapse simulations in a cosmological context

Chapter 9

Summary and Outlook

We conclude here with a summary of our findings and briefly discuss them in the light of the recent results from the *Wilkinson Microwave Anisotropy Probe* (WMAP) and QSO observations. We close with a few remarks on immediate extensions of the work presented here and briefly highlight future prospects in the area of our investigation.

In short, this work can be characterised as an attempt to link the apparent present day abundance of super massive black holes found at the centres of nearby galaxies to much lighter precursor BHs that are the collapsed remnants of the very first stars in the Universe. One of the key predictions of our model is that in addition to central SMBHs a number of massive black holes should be present in the halos of present day galaxies. A detection of these MBHs would thus provide a key test for our model.

9.1 First star formation and seed MBHs

A pre-condition for stars to form inside dark matter halos is that the baryonic component can cool and settle to the halo centre fast enough for a star-forming cloud to form. Using analytical arguments we looked at the minimum halo mass required for baryonic cooling to occur within a Hubble time. We subsequently referred to halos with this mass as MBC halos. In hierarchical structure formation MBC halos, in principle, form at all redshifts although with different probabilities that depend on the height of the peak in the initial matter density field from which these halos are supposed to have collapsed. For instance, at a redshift of $z \approx 30$ MBC halos collapse from peaks about 3.5 standard deviations ('σ') denser than the mean, while at $z \approx 20$

they collapse from the more frequent 2.5 σ peaks.

We used a simple argument from the energy of supernova outflows to constrain the redshift range for star forming halos to $16 < z < 27$ corresponding to collapse from $2.6 - 3.3\sigma$ peaks. This is the range where halos are still light enough for supernova ejecta to leave them, yet frequent enough for the ejecta to metal enrich the IGM to a minimum level required for a switch to the present-day mode of star formation. This redshift range is also in line with more detailed analytical predictions by other workers.

High resolution simulations of collapsing baryonic clouds with primordial chemical abundances and associated long cooling times show that cloud fragmentation scales are very much larger than today. This indicates that the first stars may have been few and rather massive. Extrapolating from the most recent simulations it appears, that possibly not more than one of these *population III* stars forms initially with a mass upwards of $\sim 10^2$ $M_\odot$.

The first stars may suppress star formation in their vicinity both through radiative and supernova feedback. We considered the impact of radiative feedback on the stars host cloud. Looking at the clustering properties of MBC halos we determined the fraction of halos in which star formation could be affected by feedback from stars that are already in place in other halos. We conclude that star formation may be suppressed in as much as a third to a half of all MBC halos.

At the end of their life time the massive stars may either collapse directly to form a BH or explode as a super- or hyper nova leaving behind a black hole. In either case the black hole remnant is expected to be massive, too, and likely of the same order as the mass of the original star.

9.2 Hierarchical merging of MBHs

To follow the MBHs from their initial formation sites into present-day galactic halos, we used a semi-analytical approach that combines a hierarchical merger tree algorithm with explicit prescriptions for the dynamical evolution of merged substructure in halos.

We investigated in more detail two dynamical processes that were not included in the semi-analytical approach. One is the tidal erosion of satellite density profiles through repeated tidal stripping and adiabatic expansion of the satellite even if it moves on fixed circular orbits within its host. We have shown that this process can lead to significant mass loss. Secondly we derived a simple scaling relation for the reduction of the central density as a MBH spirals to the centre of a halo. The omission of these processes from our semi-analytical approach, should not affect the statistical results for the evolution and abundance of MBHs as outlined in the following.

We considered a number of different sets of initial conditions for the collapse redshift and mass of the seed MBHs and followed their evolution in a fiducial ΛCDM cosmology. Although MBC halos may form more than one massive star, we have always assumed that only one massive BH remnant is produced to act as a seed in the subsequent merging process. The main effect of considering different MBC collapse redshifts and seed MBH masses is to change the overall mass density contained in MBHs. In addition different seed MBH masses affect the low mass end of the present-day halo MBH mass function.

We find that present day galactic halos should contain of the order 10^3 MBHs of which less than 5 % are located in the disk and bulge. The number of halo MBHs is almost exactly inverse proportional to their mass. MBH masses range from the initial seed mass 260 $M_\odot$to about $10^5 - 10^6$ of which we would typically expect one or a few to be present in the halo. Those MBHs that reach the centre of a halo merge to form a SMBH with a mass of $3 \times 10^5 - 3 \times 10^6$ $M_\odot$in Milky-Way sized halos. This is about 1/2 to 1/3 of the total mass contained in halo MBHs. The relation between the mass of the galactic bulge and that of the SMBH has exactly the same slope as that found in observations, although the normalisation is typically too low. This underscores the need for gas accretion to increase the SMBH mass to that observed in the nuclei of nearby galaxies. However, gas accretion alone does not seem to be able to build up stellar mass BHs to become SMBHs as massive as $\sim 10^9$ $M_\odot$today. An early BH merger component therefore seems plausible, and is further encouraged by the fact that some of the most massive SMBHs were already in place at redshifts larger than 6 as deduced from QSO observations.

Two accretion scenarios were considered for the MBHs orbiting in halos. First, MBHs will accrete from the host's ISM, however this will occur at very low rates and is essentially insignificant. This is primarily due to most MBHs orbiting outside the bulge and disk, where not sufficient gas is available. The hierarchical merging process raises the possibility that halo MBHs remain associated with a baryonic core remnant of the satellite they were previously associated with.

This core, which can be as little as 10^{-5} of the initial satellite mass would survive tidal stripping in the host halo, and could thus act as a 'portable' fuel supply for the MBH embedded at its centre. Making a few simple assumptions about the structure of this baryonic core, the result is that MBHs accrete from this core at very much higher rates. Accretion rates are typically 0.1 % percent of the Eddington rate but can be as high as 10 %. At these rates MBH mass increase through gas accretion remains significant. At most 5 - 20 % and typically a much lesser number of MBHs increase their mass by more than a factor 10 through gas accretion.

The uncertainty in this result, as well as the observational signatures we derive from it, depend primarily on the assumptions made about the structure of the baryonic core. A more realistic description of the baryonic core has to be the starting point for a more accurate determination of accretion rates and derived quantities.

9.3 Observational signatures

Observations to test our model fall into two main categories. These are observations of the present-day abundance of MBHs in halos, and those that probe the redshift evolution of MBHs. In the context of the former we have predicted the bolometric luminosities of halo MBHs accreting from baryonic cores. The luminosity function exhibits a -0.6 logarithmic slope in its declining part, while the luminosity and MBH mass are on average related as $L \propto M_{\bullet}^{1.7}$

We estimated the optical and X-ray signatures of MBHs in halos using thin disk and advection dominated accretion flow models. Particularly the latter is suitable as it applies to sub-Eddington accreting systems as is the case here. In the X-ray regime we predict luminosities exceeding 10^{39}erg s^{-1}. Comparing with observations of ultra-luminous galaxian X-ray point sources (ULXs) above 10^{38}erg s^{-1}, our X-ray luminosity function reproduces the slope and in one case also the normalisation of the observed one. It implies that accreting MBHs could at least account for a significant fraction of luminous X-ray sources. This is particularly the case in early type galaxies, where there is a smaller number of X-ray binaries that could produce similar X-ray luminosities. Observations also indicate that many ULXs are associated with globular clusters. In the context of our model this can naturally be explained, as globular clusters could may constitute the remnants of a tidally disrupted satellite + MBH system. The non-detection of ULXs in our galaxy as well as any of its known satellites is consistent with our predictions.

We have also computed the U and V band luminosity functions However, optical observations are probably only useful as follow-ups to X-ray detections, primarily because they are much more affected by extinction processes, but also because there are more competing types of sources. In contrast high X-ray luminosities are a hall-mark of accreting BH systems. Optical and X-ray luminosities have also been determined for MBH sources in Earth centred volumes of 3 - 6 kpc in size. It appears, though, that the number of sources is too low to yield any distinctive luminosity in either band.

To probe the MBH redshift evolution the primary observational signature are gravitational waves generated by mergers of MBHs. The LISA gravitational wave observatory appears to be the only instrument capable of detecting mergers between MBHs in the mass range we are considering. Our model predicts that LISA should detect some $10^4 - 10^5$ merger events per year, which is at least an order of magnitude larger than estimates of merger rates for SMBHs in other semi-analytical models of SMBH + galaxy formation. While mergers occur in principle at all redshifts up to where the seed MBHs initially formed, the bulk of events received at Earth originates at redshifts between 3 to 6. In estimating the merger rate we have assumed efficient merging. Dynamical friction can deliver MBHs wrapped in remnants of their satellite systems to the centre of the host halo within a Hubble time. In addition the massive gas infall in major mergers may drag MBHs in the central regions still closer to the centre as well as accelerate the time for hard MBH binaries to decay. As a result multiple MBH interactions appear less significant.

If the collapse/ explosion of the first stars is associated with gamma ray bursts and GRBs are in principle observable out to these redshifts the epoch of first star formation can be probed. Of the order $10^5 - 10^6$ GRBs should then be received per year.

Among other possible observations we have looked at gamma rays from MBH merger events. If only a tiny fraction $\sim 10^{-14}$ of the binding energy of merging black holes is released in gamma rays we may be able to detect them out to cosmological distances, however, the short event duration makes it difficult to identify them.

Flares from tidal disruption of stars by MBHs are unlikely to be observed for stars in the host galaxy. If MBHs are associated with dense star clusters tidal disruptions are more likely. The prospect for detection of MBHs by microlensing are also very uncertain. For the masses of MBHs the lensing time scale is typically too long but may be short enough in special circumstances.

9.4 A note on recent data from WMAP and QSO observations

The beginning of 2003 saw the first data release from the *Wilkinson Microwave Anisotropy probe*. In most respects the present understanding of the Universe is confirmed by the data and the constraints on many of the main cosmological parameters have been tightened. However, some surprising results emerge. In our context the most interesting one of these is the fact that the Universe appears to have been reionised much earlier than previously thought (Bennett et al. 2003; Kogut et al. 2003). Depending on the model assumed for the reionisation history the WMAP group estimate a reionisation redshift of $z_{reion} = 17 \pm 5$ for the simplest models to $z_{reion} = 20^{+10}_{-9}$ for the general case, where the latter quotes 95% confidence limits. If population III star formation had been a somewhat remote concept before this finding has now moved it to centre-stage. QSOs cannot account for reionisation at this early epoch. Some other source of reionisation is needed and attention has turned to population III stars as a possibility.

Additional support for this comes from the observation of iron emission in QSOs at $z \approx 6$ (Freudling, Corbin & Korista 2003). If type Ia supernovae were responsible for this metal enrichment stars would have had to have formed at $z \approx 20 \pm 10$. Massive population III stars may be able to account for reionisation, but would have to form with a top-heavy IMF (Wyithe & Loeb 2003). However, a single or at most a few massive stars per MBC halo would not be sufficient. Yet, this is what first results from the most recent simulations indicate (private communication G.Bryan). Our model provides another possibility: If seed mass MBHs already accrete from baryonic cores at redshifts $z \gtrsim 15$ their emission in the UV and X-ray may constitute a much more efficient source for reionisation.

On the other hand, the reionisation redshift from the WMAP data implies that massive star formation does not occur at redshifts lower than this. Our model C, where the first stars form in MBC halos collapsing at $z \approx 20$ from 2.5σ peaks is thus just about consistent with this; collapse from 2σ peaks (at $z \approx 15$), however, is probably ruled out.

9.5 Outlook and future work

Interest in star formation at $z \gtrsim 20$ is set to increase further in the light of recent observations and their implications for the high redshift Universe. In this context early MBH formation and subsequent merging may also become an increasingly important area of investigation. This is not least because future observations will open new possibilities to probe MBH abundances and their previous merging in complementary ways.

Dedicated campaigns to survey luminous X-ray sources in a large number of galaxies with *Chandra* and *XMM* are underway. Their superior resolution should allow to constrain much better the nature of these sources and thus the abundance of MBHs. Observations in other bands could help establish correlations as would be expected for spectral emission models for MBH accretion such as the ADAF model we used. Our model predicts the presence of MBHs not only within the visible extent of galaxies but also in their halos. A relatively simple way of testing this would be to observe larger fields around galaxies. However, at increasingly larger distances from the galaxies the distinction of MBH from background sources will become more difficult.

Abundances of luminous X-ray sources appear to vary depending on galaxy morphology, in particular implying that more sources in early type galaxies are possibly MBHs. An extension of our work would then be to see if this can be reproduced, e.g. by including prescriptions for the disruption of baryonic cores by encounters with galactic disks, which would result in a lower number of luminous MBH sources. Extinction has a large impact and, although included in many observations, our predictions may as well be modified to account for this, too. Given the variability in the merger trees we analysed, statistical errors in our predictions could be reduced by considering a larger number of trees.

Tests of the formation and redshift evolution of MBHs still appear to be some way away. The more speculative idea that the first stars explode and produce GRBs could possibly be tested fairly soon with upcoming gamma ray observations such as from the *SWIFT* mission. Detections of gravitational waves from the merging of MBHs will begin in earnest probably not before LISA which is planned to be launched in 2011.

Two of the largest uncertainties in our model are the efficiency for the formation of seed MBHs and of their subsequent merging once any two form a hard binary. Regarding the former it seems only hydrodynamical simulations with much higher resolution and including relativistic fluid dynamics will be able to yield the required constraints on the initial mass function of massive population III stars and resulting MBHs. To determine the evolution of MBH binaries more accurately we need to better understand the role of gas especially during and after mergers of the MBH's hosts. This will also require considerably better resolved hydrodynamical simulations.

Index of symbols used

The following is a list of the most important or commonly used quantities and variables in this thesis. Throughout the text a subscript '0' denotes the present day value of the respective quantity unless stated otherwise.

c	speed of light
c_s	speed of sound
$D(z)$	comoving distance to redshift z
D_{grow}	linear growth factor
E_{bind}	halo binding energy
f_{gw}	gravitational wave frequency
$f_{pk}(> \nu_{pk}$	fraction of mass in Universe contained in peaks above height ν_{pk}
f_z	$= M_{sn}/M_b$
G	Gravitational constant
H	Hubble constant
h_s	gravitational strain amplitude
$\dot{m}$	dimensionless accretion rate (in units of Eddingtonaccretion rate)
k_B	Boltzmann constant
L_{bol}	bolometric luminosity
m_p	Proton mass
M_b	baryonic mass inside halo
M_{MBC}	minimum halo mass allowing efficient baryonic cooling
M_h	halo mass
M_{host}	mass of host halo
M_J	Jeans mass
M_{12}	total mass of binary constituents
M_{chirp}	BH binary 'chirp' mass
M_{sat}	mass of satellite halo
M_{sn}	mass of super nova ejecta ('metals')

$M_\star$	stellar mass
$M_\bullet$	MBH mass
n_b	baryon number density
n_h	halo number density
n_{pk}	number density of peaks
N_γ	number of ionising photons per star
$P(k)$	cosmological matter power spectrum
R_{vir}	halo virial radius
r_{acc}	accretion radius
r_b	radius of baryonic component (in halo)
r_{lso}	radius of BH last stable orbit
r_t	satellite tidal radius
r_{td}	thin disk radius
t_{cool}	baryonic cooling time scale
t_{hubble}	Hubble time
t_{rlx}	two body relaxation time scale
t_{salp}	Salpeter time scale
$t_\star$	stellar life time
T_{tf}	matter transfer function
T_{vir}	halo virial temperature
T_{td}	thin disk temperature
V_{vir}	circular velocity at halo virial radius
v_c	circular velocity
$v_\bullet$	BH velocity

α	ADAF viscosity parameter
α_c	Compton spectrum power-law index
β_s	$= v_\bullet / c_s$
δ_c	fractional overdensity
$\Delta_c(z)$	mean overdensity in virialised halo
η	radiative efficiency of accretion
γ	power-law index for halo density profiles
Λ	cosmological constant (or dark energy)
Λ_c	Coulomb logarithm
Λ_{cool}	cooling rate
Λ_J	Jeans length
μ_{mol}	mean molecular weight (of primordial baryons)
μ_{12}	reduced mass of binary constituents
ν_p	synchrotron peak frequency
ν_{pk}	height of peak in matter density field in units of field variance σ
ν_{rec}	rate of received gravitional wave merger events

Ω_m	cosmic matter density parameter
Ω_Λ	density parameter corresponding to dark energy component (cosmological constant)
Ω_r	density parameter corresponding to radiation component
Φ	host gravitational potential
ϕ_{sat}	satellite gravitational potential
$\rho_\bullet$	(S)MBH mass density
ρ_{crit}	critical matter density in Universe
ρ_g	gas density
$\sigma^2(r)$	variance of matter field on scales r
σ_c	galaxy bulge velocity dispersion
σ_h	halo velocity dispersion
τ_{df}	dynamical friction time scale
τ_{gw}	gravitational wave induced inspiral time scale
$\xi(r)$	two-point correlation function of linear matter density field
$\xi_{pk}(r)$	two-point correlation function of peaks

References

Abel T., Bryan G.L., Norman M.L., ApJ **540**, 39 (2000)

Abel T., Anninos P., Norman M.L., Zhang Y., ApJ **508**, 518 (1998)

Adams F.C., Graff D.S., Richstone D.O., ApJ **551**, L31 (2001)

Armitage P.J., Natarajan P., ApJ **567**, L9 (2002)

Bahcall J.N., Wolf R.A., ApJ **209**, 214 (1976)

Bahcall N.A., Fan X., Cen R., ApJ Lett. **485**, 53 (1997)

Bahcall N.A., Fan X., ApJ **504**, 1 (1998)

Ball G.H., Narayan R., Quataert E., ApJ **552**, 221 (2001)

Bardeen J.M., Bond J.R., Kaiser N., Szalay A.S., ApJ **304**, 15 (1986)

Barkana R., Loeb A., Phys. Rep. **349**, 125 (2001)

Barnes J.E., MNRAS **333**, 481 (2002)

Barnes J.E., Hernquist L., ApJ **471**, 115 (1996)

Beasley M.A., Baugh C.M., Forbes D.A., Sharples R.M. et al. ,MNRAS **333**, 383 (2002)

Begelman M.C., Blandford R.D., Rees M.J., Nature **287**, 307 (1980)

Bennett C., Hill R.S., Hinshaw G., Nolta M.R. et al. , ApJ submitted, `astro-ph/0302208` (2003)

Binney J., Tremaine S., 1987, "Galactic Dynamics", Princeton University Press

Bloom J.S., PhD Thesis, CalTech 2002

Bondi H., Hoyle F., MNRAS **104**, 273 (1944)

Bondi H., MNRAS **112**, 195 (1952)

Bond J.R., Cole S., Efstathiou G. et al. , ApJ **379**, 440 (1991)

Bower R., MNRAS **248**, 332 (1991)

Broadhurst T.J., Ellis R.S., Koo D.C., Szalay A.S., Nature **343**, 726 (1990)

Bromm V., Ferrara A., Coppi P.S. et al. , MNRAS **328**, 969 (2001)

Bromm V., Kudritzki R.P., Loeb A., ApJ **552**, 464 (2001)

Bromm V., Coppi P.S., Larson R.B., ApJ **564**, 23 (2002)

Bromm V., Loeb A., ApJ **575**, 111 (2002)

Bryan G.L., Norman M.L., ApJ **495**, 80 (1998)

Bryan G.L., Abel T., Norman M.L., Proceedings 'Super Computing 2001, Denver, November 2001, `astro-ph/0112089`

Carr B.J., Bond J.R., Arnett W.D. ApJ **277**, 445 (1984)

Cavaliere A., Vittorini V., ApJ **570**, 114 (2002)

Chisholm J.R., Dodelson S., Kolb E.W., ApJ submitted, `astro-ph/0205138`

Chung D.J.H., Kolb E.W., Riotto A., Tkachev I.I., Phys. Rev. **D62**, 043508 (2000)

Ciotti L., Haiman Z., Ostriker J.P., The Mass of Galaxies at Low and High Redshift. Proceedings of the ESO Workshop held in Venice, Italy, 24-26 October 2001, p. 106., 106

Colbert E.J.M., Mushotzky R.F., ApJ **519**, 89 (1999)

Colbert E.J.M., Ptak A.F., ApJ **143**, S25 (2002)

Colpi M., Mayer L., Governato F., ApJ **525**, 720 (199)

Croft R.A.C., Weinberg D.H., Bolte M., Burles S. et al. , ApJ **581**, 20 (2002)

de Araujo J.C.N., Miranda O.D. & Aguiar O.D., MNRAS **330**, 651 (2002)

Dokuchaev V.I., Eroshenko Yu. N., A&A submitted

Dubinski J., Carlberg R.G., ApJ **378**, 496 (1991)

Ebisuzaki T., Makino J., Tsuru T.G., Yoko F. et al. , ApJ **562**, L19 (2001)

Einasto J., Einasto M., Gottlöber S., Müller V. et al. , Nature **385**, 139 (1997)

Eisenstein D.J., Hu W., ApJ **515**, 5 (1999)

Eke V.R., Navarro J.F., Steinmetz M., ApJ **554**, 114 (2001)

El Eid M.F., Fricke K.J., Ober W.W., A&A **119**, 54 (1983)

Fabbiano G., Ann. Rev. A & A **27**, 87 (1989)

Fabian A.C., MNRAS **308**, L39 (1999)

Ferrarese L., ApJ **578**, 90 (2002)

Flanagan É.É., Hughes S.A., Phys. Rev. **D57**, 8 (1998)

Frank J., King A.R., Raine D.J., 1985, 'Acretion Power in Astrophysics'

Freudling W., Corbin M.R., Korista K.T. ApJ **587**, L67 (2003)

Fujita Y., Inoue S., Nakamura T., Manmoto T. et al. , ApJ **495**, L85 (1998)

Fukugita M., Hogan C.J., Peebles P.J.E., ApJ **320**, 518 (1998)

Fuller T.M., Couchman H.M.P., ApJ **544**, 6 (2000)

Gaztañaga E., Baugh C.M., MNRAS **294**, 229 (1998)

Genzel R., Pichon C., Eckart A., Gerhard O.E. et al. MNRAS **317**, 348 (2000)

Gerssen J., van der Marel R.P., Gebhardt K., Guhathakurta P. et al. , Astron. J. **124**, 3270 (2002)

Gould A. ApJ **392**, 442 (1992)

Gould A., Rix, H.-W., ApJ **532**, L29 (2000)

Graham A.W., Erwin P., Caon N., Trujilo I., ApJ **563**, L11 (2001)

Griffiths, L., Silk, J., Zaroubi, MNRAS **324**, 712 (2001)

Grimm L.-H., Gilfanov M., Sunyaev R., A&A **391**, 923 (2002)

Haardt F., Ripamonti E., Colpi M., Ferrara A., Astrophys. & Space Sc. **281**, 479 (2002)

Haiman Z., Loeb A., ApJ **552**, 549 (2001)

Haehnelt M., MNRAS **269**, 199 (1994)

Haehnelt M.G., Natarajan P. Rees M.J., (HNR98), MNRAS **300**, 817 (1998)

Haehnelt M.G., Kauffmann G., MNRAS **318**, L35 (2000)

Hamilton A., Tegmark M., MNRAS **330**, 506 (2002)

Hannestad S., Hansen S.H., Villante F.L., Astropart.Phys. **16**, 137 (2001)

Heger A., Woosley S. Baraffe I., Abel T. 2002, Lighthouses of the Universe: The Most Luminous Celestial Objects and Their Use for Cosmology Proceedings of the MPA/ESO/, p. 369, 369

Ho L.C. 1999, in ed. S. K. Chakrabarti, Observational Evidence for Black Holes in the Universe, Dodrecht: Kluwer 157

Holt S.S., Schlegel E.M., Hwang U., Petre R., ApJ accepted, `astro-ph/0301319`

Hut P., White S.D.M., Nature **310**, 637 (1984)

Hut P., Rees M.J., MNRAS **259**, 27 (1992)

Hutchings R.M., Santoro F., Thomas P.A., Couchman H.M.P., MNRAS **330**, 927 (2002)

Ibata R.A., Gilmore G., Irwin M.J., MNRAS **277**, 781 (1995)

Ichimaru S., ApJ **214**, 840 (1977)

Ipser J.R., Price R.H., ApJ **216**, 578 (1977)

Jeltema T.E., Canizares C.R., ApJ **585**, 756 (2003)

Kaaret P., ApJ **578**, 114 (2002)

Kahabka P., A&A **388**, 100 (2002)

Kauffmann G., Haehnelt M.G., MNRAS **311**, 576 (2000)

Kilgard R.E., Kaaret P., Krauss M.I., Prestwich A.H. et al. ApJ **573**, 138 (2002)

King I., Astron. J. **67**, 571 (1962)

King A.R., Davies M.B., Ward M.J., Fabbiano G. et al. ApJ **552**, L109 (2001)

Klessen R., Burkert, A. MNRAS **280**, 735 (1996)

Klypin A., Gottlöber S., Kravtsov A.V., Khokhlov A.M., ApJ **516**, 530 (1999)

Klypin A., Kravtsov A.V., Bullock J.S., Primack J.R., ApJ **554**, 903 (2001)

Knebe A., Islam R.R., Silk J., MNRAS **326**, 109 (2001)

Kogut A., Spergel D.N., Barnes C., Bennett C.L. et al. , ApJ submitted, `astro-ph/0302213` (2003)

Komossa S.,"Lighthouses of the Universe: The Most Luminous Celestial Objects and Their Use for Cosmology", Garching, August 2001, `astro-ph/0109441`

Kormendy J & Richstone D., A&A Rev. **33**, 581 (1995)

Kormendy J , 2000, in Funes, Corsini & Corsini, Eds., Galaxy Disks and Disk Galaxies ASP Conference Series

Kormendy J & Gebhardt K. (KG01), 2001, in eds. Martel & Wheeler, Porc.AIP Symp. 586, XX Texas Syposium on Relativistic Astrophysics, New York: Am. Inst. Phys., p. 363

Lacey C., Cole S., MNRAS **262**, 627 (1993)

Lamb D.Q., Reichart D.E., ApJ **536**, 1 (2000)

Larson R.B, MNRAS **194**, 809 (1981)

Lacey C., Cole S., 1993, MNRAS **262**, 627 (1993)

Lesgourgues J., Polarski D., Starobinsky A.A., MNRAS **297**, 769 (1998)

Liang E.W., Liu H.T., MNRAS **340**, 632 (2003)

Lidsey J.E., Liddle A.R., Kolb E.W, Copeland E.J. et al. , Rev. Mod. Phys. **69**, 373, (1997)

Lin D.N.C., Murray S.D., IAU Symposium, 187, 165 (2002)

Lynden-Bell D., MNRAS **136**, 101 (1967)

Mackey J., Bromm V., Hernquist L., ApJ **586**, 1 (2003)

MacMillan J.D., Henriksen R.N., Carnegie Observatories Astrophysics Series, Vol. 1: Coevolution of Black Holes and Galaxies, ed. L. C. Ho, Pasadena: Carnegie Observatories)

Madau P., Rees M.J., ApJ **551**, L27 (2001)

Magorrian J., Tremaine S., Richstone D., Bender R. et al. (Mag98), Astron. J. **115**, 2285 (1998)

Magorrian J., Tremaine S., MNRAS **309**, 447 (1999)

Mahadevan R., ApJ **477**, 585 (1997)

Makishima K., Kubota A., Mizuno T., Ohnishi T., ApJ **535**, 632 (2000)

Manmoto T., Mineshige S, Kusunose M., ApJ **489**, 791 (1997)

Maoz E, ApJ **494**, L181 (1998)

Marigo P., Girardi L., Chiosi C., Wood P.R., A&A **371**, 152 (2001)

Markowitz A., Edelson R., Vaughan S., Uttley P. et al. , ApJ accepted `astro-ph/0303273`

Martin J., Riazuelo A., Sakellariadou M., Phys. Rev. **D61**, 3518 (2002)

Menou K., Haiman Z., Narayanan V.K., ApJ **558**, 535 (2001)

Merritt D. & Ferrarese L. (MF01a), MNRAS **320**, L30 (2001)

Merritt D. & Ferrarese L. (MF01b), ApJ **547**, 140 (2001)

Merritt D., Cruz F., ApJ **551**, L41 (2001)

Merritt D., Poon M.Y., `astro-ph/0302296`, (2003)

Milgrom M., ApJ **270**, 365 (1983)

Milosavljević M., Merritt D., ApJ **563**, 34 (2001)

Milosavljević M., Merritt D., Rest A., van den Bosch F.C., MNRAS **331**, L51 (2002)

Miller, M.C., ApJ **544**, 43 (2000)

Miller M.C., Hamilton D.P., MNRAS **330**, 232 (2002)

Misner C.W, Thorne K.S. & Wheeler J.A., "Gravitation", Freeman and Company, 1973, p.988

Moore B., Quinn T., Governato F., Stadel J. et al. , MNRAS **310**, 1147 (1999)

Moore B., Calcáneo-Roldán C., Stadel J., Quinn T. et al. , Phys. Rev. **D64**, 63508 (2001)

Mouri H., Taniguchi Y., ApJ **566**, L17 (2002)

Mukhanov V., Feldman H.A., Brandenberger R.H., Phys. Rep. **215**, 203 (1992)

Naab T., Burkert A., 2001, ASP Conf. Ser. 249: The Central Kiloparsec of Starbursts and AGN: The La Palma Connection, p. 735

Nakano T., Makino J., ApJ **510**, 155 (1999)

Nakano T., Makino J., ApJ **525**, L77 (1999)

Nakamura T., Sasaki M., Tanaka T., Thirne K.S., ApJ **487**, L139 (1997)

Nakamura T., Umeda H., Iwamoto K., Nomoto K. et al. , ApJ **555**, 880 (2001)

Narayan R., Yi I., ApJ **428**, L13 (1994)

Narayan R., Mahadevan R., Quataert E., in *The Theory of Black Hole Accretion Disks*, eds. M.A. Abramowicz, G. Bjornsson & J.E. Pringle, Cambridge University Press

Narlikar J.V., 1993, "Introduction to Cosmology", CUP

Navarro J.F., Frenk C.S., White S.D.M., ApJ **490**, 493 (1997)

Ong R.A., Phys. Rep. **305**, 93 (1998)

Omukai K., Palla F., ApJ **561**, L55 (2001)

Omukai K., Inutsuka S.,MNRAS **332**, 59 (2002)

Ostriker J.P., McKee C.F., Rev. Mod. Phys. **60**, 1 (1988)

Padmanabhan T., 1993, "Structure Formation in the Universe", Cambridge University Press

Padmanabhan T., Choudhury T.R., Phys. Rev. **D66**, 081301 (2002)

Peacock J.A., 1999, "Cosmological Physics", Cambridge University Press

Peebles P.J.E, 1980, "The Large Scale Structure of the Universe", Princeton University Press

Percival W.J., Sutherland W., Peacock J.A., Baugh C.M. et al. , MNRAS **337**, 1068 (2002)

Perlmutter S., Aldering G., Goldhaber G., Knop R.A. et al. , ApJ **517**, 565 (1999)

Peters P.C., Phys. Rev. **B136**, 1224 (1964)

Piran T., Phys. Rep. **333**, 529 (2000)

Press W.H., Schechter P., ApJ **187**, 425 (1974)

Quinlan G.D., NewA **1**, 35 (1996)

Ravindranath S., Ho L.C., Filippenko A.V., ApJ **566**, 801 (2002)

Rees M.J., Ostriker J.P., MNRAS **179**, 541 (1977)

Rees M.J., Begelman M.C., Blandford R.D., Phinney E.S., Nature **295**, 17 (1982)

Rees M.,Nature **333**, 523 (1988)

Richstone D., Ajhar E.A., Bender R., Bower G. et al. , Nature **395**, A14 (1998)

Riess P.G., Filippenko A.V., Challis P., Clocchiatti A. et al. , Astron. J. **116**, 1009 (1998)

Ripamonti E., Haardt F., Ferrara A., Colpi M., MNRAS **334**, 401 (2002)

Rix H.-W., Lake G., ApJ **417**, L1 (1993)

Roberts T.P., Warwick R.S., MNRAS **315**, 98 (2000)

Roy Choudhury T., Srianand R., MNRAS **336**, L27 (2002)

Scalo J.M., Fund. of Cosm.Phys., **11**, 1, (1986)

Schlegel D.J., Finkbeiner D.P., Davis M. ApJ **500**, 525 (1998)

Schneider R., Ferrara A., Ciardi B., Ferrari V. et al. , MNRAS **317**, 385 (2000)

Schödel R., Ott T., Genzel R., Hofmann R. et al. , Nature **419**, 694 (2002)

Shakura N., Sunyaev R.A., A&A **24**, 377 (1973)

Shapiro S.L. & Teukolsky S.A., "Black Holes, White Dwarfs and Neutron Stars", Wiley 1983

Shapiro S.L., Teukolsky S.A., ApJ **292**, L41 (1985)

Silk J., ApJ **211**, 638 (1977)

Silk J., Rees M.J., A&A **331**, L1 (1998)

Smith D.A., Wilson A.S., to appear in ApJ July 2003, `astro-ph/0303387`

Soltan A, MNRAS **200**, 115 (1982)

Somerville R.S., Kolatt T.S., MNRAS **305**, 1 (1999)

Songaila A., ApJ **561**, L153 (2001)

Spergel D.N., Verde L., Peiris H.V., Komatsu E., ApJ submitted, `astro-ph/0302209` (2003)

Swartz D.A., Ghosh K.K., McCollough M.L., Pannuti T.G., ApJ **144**, 213 (2003)

Swartz D.A., Ghosh K.K., Tennant A.F., AAS meeting 201 (2003), `astro-ph/0302203`

Syer D., Ulmer A., MNRAS **306**, 35 (1999)

Taylor J.E., Ph.D. thesis, University of Victoria 2001

Taylor J.E., Babul A., ApJ **559**, 716 (2001)

Taylor J.E., Babul A., `astro-ph/0301612` (2003)

Taylor J.E., Navarro J.F., ApJ **563**, 483 (2001)

Taylor J.E., Navarro J.F., "The Evolution of Galaxies III: From Simple Approaches to Self-Consistent Models", proceedings of the 3rd EuroConference on the evolution of galaxies, held in Kiel, Germany, July 16-20, 2002, `astro-ph/0210074`

Tennant A.F., Wu K., Ghosh K.K., Kolodziejczak J.J. et al. ApJ **549**, L43 (2001)

Tegmark M., Silk J., Rees M.J., Blanchard A. et al. , ApJ **474**, 1 (1997)

Tegmark M., Zaldarriaga M., Hamilton A.J., Phys. Rev. **D63**, 43007 (2001)

Thorne K. S.,"Three Hundred Years of Gravitation", eds. S.W. Hawking and W. Israel, CUP, 1989, p.330

Tremaine S., Gebhardt K., Bender R., Bower G. et al. , ApJ **574**, 740 (2002)

Turner T.J., Mushotzky R.F., Yaqoob T., George I.M. et al. , ApJ **574**, L123 (2002)

van der Marel R.P., Gerssen J., Guhathakurta P., Peterson R.C. et al. , in press Astron. J. **124**, 3255 (2002)

Voit G.M., ApJ **465**, 548 (1996)

Volonteri M., Haardt F., Madau. P, ApJ **582**, 559 (2003)

Wang Q.D., MNRAS **332**, 764 (2002)

Weinberg M.D., ApJ **478**, 435 (1997)

Wyithe J.S.B., Loeb A., ApJ **588**, L69 (2003)

Zang Z., Meurs E.J.A., ApJ **556**, 24 (2001)

Zezas A., Fabbiano G., ApJ **577**, 726 (2002)

Zhao H., Haehnelt M.G., Rees M.J., NewA **7**, 385 (2002)

Portegies Zwart S.F., McMillan S.L.W., ApJ **576**, 899 (2002)

Yu Q., Tremaine S., MNRAS **335**, 965 (2002)

Yu Q., MNRAS **331**, 935 (2002)