The role of black holes in galaxy formation

Rowena Katherine Malbon

The Role of Black Holes in Galaxy Formation

by Rowena Katherine Malbon

PhD Thesis, September 2006

Abstract

We incorporate a model for black hole growth during galaxy mergers into the semi-analytical galaxy formation model based on ΛCDM proposed by Baugh et al. (2005). Our black hole model has one free parameter, which we set by matching the observed zeropoint of the local correlation between black hole mass and bulge luminosity. We present predictions for the evolution with redshift of the relationships between black hole mass and bulge properties. Our simulations reproduce the evolution of the optical luminosity function of quasars. We study the demographics of the black hole population and address the issue of how black holes acquire their mass. We find that the direct accretion of cold gas during starbursts is an important growth mechanism for lower mass black holes and at high redshift. The re-assembly of pre-existing black hole mass into larger units via merging dominates the growth of more massive black holes at low redshift. As redshift decreases, progressively less massive black holes have the highest fractional growth rates, in line with recent claims of 'downsizing' in quasar activity.

We are able to reproduce the observed population of rare, highly luminous quasars at $z \sim 6$ in the ΛCDM model, although we struggle when the model parameters are refined to those in the current best fit of Sanchez et al. (2006). which has been hinted at by recent observations. We find that the most massive black holes and the most luminous quasars at $z \sim 6$ are not hosted exclusively by the most massive dark matter haloes, as is often assumed.

Finally, we repeat our study of the assembly histories of black holes using the new galaxy and black hole formation model of Bower et al. (2006), which includes AGN feedback. Black hole growth in this model is dominated by the accretion required to fuel AGN feedback at low redshifts and by the accretion during starbursts triggered by disc instabilities at high redshifts. At no redshift is growth by BH-BH mergers dominant, although we still predict a high BH-BH merger rate. The Bower et al. model gives a somewhat better match to the $M_{\rm BH}$ – bulge relations, while the Baugh et al. model gives a better match to the quasar luminosity function. We suggest future directions to explore with these models.

The Role of Black Holes in Galaxy Formation

by Rowena Katherine Malbon

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

Department of Physics
University of Durham
September 2006

Dedication

To my mother and father.

Contents

List of Figures

List of Tables

Declaration

The work described in this thesis was undertaken between 2001 and 2006 while the author was a research student under the supervision of Prof. Carlos Frenk and Dr. Carlton Baugh in the Department of Physics at the University of Durham. This work has not been submitted for any other degree at the University of Durham or any other University.

Portions of this work have appeared in the following papers:

- Malbon, Rowena K., Baugh, Carlton M., Frenk, Carlos S., Lacey, Cedric L., MNRAS submitted, preprint (astro-ph/0607424) (Chapters 2, 3, 4 and 7)
- Bower, Richard G., Benson, Andrew J., Malbon, Rowena K., Helly, John C., Frenk, Carlos S., Baugh, Carlton M., Cole, Shaun, Lacey, Cedric G., 2006, MNRAS, 370, 645 (Chapter 6)

Acknowledgements

The work carried out in this thesis would not have been possible without the guidance and support of my supervisors, Dr Carlton Baugh and Professor Carlos Frenk. I would also like to acknowledge Richard Bower, Cedric Lacey, Shaun Cole, Andrew Benson and John Helly who, along with Carlos and Carlton, developed the `GALFORM` semi-analytic model used throughout this work.

Chapter 1 | *Introduction*

1.1 The Smooth Universe

The Universe is homogeneous and isotropic. These two assumptions, in conjunction with general relativity (Einstein 1916), lead to the Friedman equations for the dynamics of the Universe, a form of which is as follows:

$$\dot{R}^2 - 8\pi G\rho/3R^2 = -kc^2 \quad , \tag{1.1}$$

where R is the scale factor of the Universe, G is the gravitational constant, c is the speed of light and k is a constant (where $-1 \leq k \leq 1$; $k > 0$ if the Universe is spatially closed and $k < 0$ if the Universe is spatially open).

Galaxies in the Universe in general appear to be receding from us. We[1] observe that the velocity of galaxies with respect to us (as inferred by their redshift) is directly proportional to their distance from us. This is Hubble's law: *velocity* $= c \times z = H_0 \times$ *distance*, where H_0 is the Hubble constant (Hubble 1929). If we are not a 'special' observer, then this directly implies that the Universe as a whole is expanding: $\dot{R} = H_0 R$.

Given the expansion, extrapolating backwards in time from the present, the Universe must once have been much smaller and denser than it is today, and, assuming adiabatic expansion, also much hotter. When electrons and protons recombined at $z \sim 1100$, photons were emitted from the last scattering surface (Dicke et al. 1965). The photons have a black-body spectrum since the matter and radiation are strongly coupled, and in thermal equilibrium up until the last scattering surface. As the Universe expands adiabatically, the photons from the last scattering surface are doppler-shifted to longer wavelengths – the 'observer', (i.e. us) is moving away from the rest of the Universe, including the last scattering surface. The photons are now observed as an almost perfect black body spectrum (Penzias & Wilson 1965) with a temperature of 2.73 K (Mather et al. 1990). Most of the radiation is now at microwave wavelengths, so we call it the 'cosmic

[1]Throughout this Thesis I will use the conventional 'we' to refer to the first person.

microwave background' (CMB). The temperature is essentially constant in all directions, to approximately one part in 10^5, once the dipole due to our motion with respect to the CMB is accounted for (Smoot et al. 1992).

What is the Universe made of? We are able to make a reasonably accurate census of the density in the Universe which can be accounted for by baryons (Fukugita et al. 1998). Modelling of stars is very well understood, and can be used to make a very accurate inference of the mass of a star from its spectrum. Using these models, and our knowledge of the mass function of stars in galaxies, we are able to make a reasonably accurate estimate of the stellar mass in a distant galaxy from its integrated luminosity, in particular. By observing the galaxy K-band luminosity function in a large volume of the Universe, we are able to make a fairly accurate census of the total mass density which is accounted for by stars (e.g. Cole et al. 2001; Bell et al. 2003). We are also able to estimate the mass of gas in the Universe from observations such as X-ray emission from clusters, the Lyman-alpha forest and HI emission from galactic discs. We also make further corrections for the baryonic mass present in brown dwarfs and the warm-hot intergalactic medium, both of which are difficult to observe. We conclude that the Universe has a density in baryons of $\Omega_{\rm baryon}\,\rho_{\rm critical}$[2], where $\Omega_{\rm baryon} \approx 0.04$. Finally and importantly, there is an independent and perhaps more robust constraint on $\Omega_{\rm baryon}$ from nucleosynthesis arguments, which agrees well with the result of adding up all the contributions to the baryonic mass outlined above.

There is strong evidence that there is far more matter in the Universe than can be accounted for by baryons alone. For example, the velocities of galaxies within a cluster can be used to estimate the mass of the cluster. The resulting mass is much larger than that which can be accounted for by the luminous material. This observation and argument were first made by in 1933 by Fritz Zwicky, who described the shortfall as 'missing mass'. A second example is that the rotation curves of spiral galaxies have much higher velocities towards large radii than would be expected from the observed stars and gas alone. This is evidence that galaxies are embedded in a large halo of dark matter (Rubin et al. 1978). In the late 1970's and the early 1980's, neutrinos were commonly suggested to be a suitable dark matter candiate. However, we now believe that, whilst neutrinos have mass, they

[2]We express the densities of the components which make up the Universe in units of the critical density, $\rho_{\rm critical} = 3H_0^2/(8\pi G)$, where $\rho_{\rm critical}$ is the mass density required to give the Universe a flat geometry. In a Universe without a component of dark energy, $\rho_{\rm critical}$ is also the maximum density for which the Universe will never recollapse.

contribute a very small fraction of the mass in the Universe. Furthermore, computer simulations of the growth of the large scale structure of the Universe look quite different from the real Universe when hot dark matter is used (Frenk et al. 1983). The current paradigm is that the major component of the missing mass is 'cold dark matter' or CDM. This is a generic term describing dark matter which decouples from photons while it is non-relatavistic, and has an effectively zero thermal velocity. These properties of CDM mean that small-scale perturbations are not smeared out by free-streaming. There are many candidate particles for cold dark matter arising from extensions to the standard model of particle physics, but none has been detected to date in the laboratory. An alternative to dark matter is a 'modification of Newtonian gravity' (MOND; Milgrom 1983). Whilst this may explain the rotation curves of galaxies well, it is not so successful in explaining structure formation or observations of CMB fluctuations.

It turns out that, at least for $z \lesssim 1$, the dominant component of the Universe in terms of its dynamics is not in fact matter, but dark energy. The defining characteristic of dark energy is that, unlike matter, it has a negative pressure. Thus, it drives an acceleration of the expansion rate of the Universe. The most direct evidence we have for the presence of dark energy are supernova type Ia (SNIa) observations, which indicate that the expansion of the Universe is indeed accelerating (Riess et al. 1998; Perlmutter et al. 1999). Assuming that SNIa are 'standard candles' (i.e. that their absolute magnitude is constant), we can make the SNIa 'Hubble diagram', i.e. a plot of apparent magnitude against redshift. Since, for a standard candle, the apparent magnitude is a function of the distance away from us[3] and the redshift is a function of the recession velocity, From this, we are able to infer the scale-size and rate of expansion of the Universe as a function of redshift. The simplest form of dark energy is the cosmological constant, which has an equation of state $P = -\rho$, where P is pressure and ρ is energy density. There are also models of dark energy with more complicated equations of state, which may evolve with redshift ('quintessence'). From SNIa observations, we are not yet able to determine the equation of state to high accuracy, although we do have some weak constraints on its (lack of) evolution. The dark energy may be related to the energy density of the vacuum which is a consequence of quantum mechanics. While particle physics theories are able to produce a vacuum energy density with the correct sign, the magnitude of their predictions is $\sim 10^{40} - 10^{120}$ times too large. Thus the nature of dark energy is an even greater mystery than that of

[3]More accurately, apparent magnitude is a function of the luminosity distance, which is actually an integral over the scale-size of the Universe from $z = 0$ to the redshift being considered.

dark matter. There are some caveats to the inference of dark energy from the supernova data. Type Ia supernovae may turn out not to be standard candles – their properties may evolve with redshift. Furthermore, significant corrections for dust extinction must by made to the supernovae luminosities, and the properties of the dust may vary with redshift. However, there is independent evidence for the presence of dark energy, albeit in the context of a specific cosmological model, from a joint analysis of CMB temperature fluctuations and the power spectrum of galaxy clustering (Efstathiou et al. 2002).

1.2 The Clumpy Universe

When we observe the Universe, we see galaxies. In Fig. 1.1 (top, top-inset and left), we show the distribution of galaxies in three regions of the local Universe, as observed using galaxy redshift surveys. Galaxies tend to be clustered – we see clusters, filaments, walls and voids. When were these structures first imprinted? The highest redshift at which we can observe structure in the Universe is at $z \sim 1100$, where we can see fluctuations in the CMB. These fluctuations were first observed in 1992 (Smoot et al. 1992) using the COBE satellite. A number of experiments in subsequent years have improved the measurement of CMB fluctuations, in particular WMAP (Bennett et al. 2003); we show the pattern of CMB temperature fluctuations as observed by WMAP in Fig. 1.2. If structure in the galaxy distribution originates from the structure imprinted in the CMB, then we require a model of structure formation which links together the tiny, $\sim 10^{-5}$ pertubations which were in the CMB ~ 380000 years after the big bang to the complex structures we observe in the Universe $\sim$ 14 billion years later. How does this occur? The basic answer is gravitational instability, which, over $\sim$ 14 billion years, amplifies the small perturbations by many orders of magnitude, allowing the formation of galaxies, and structure in their distribution.

Gravity is an attractive force and the force it exerts becomes stronger as distance decreases and density increases. Small primordial fluctuations are amplified by gravity. Fluctuations in the density of the Universe can be approxinated as sperically symmetric perturbations – the so-called 'top-hat' model (Gunn & Gott 1972). Birkhoff's theorem allows us to model these overdensities as minature FRW universes, albeit with a slightly higher density than the Universe as a whole. At first, the expansion of the spherical overdensity slows down with respect to the background expansion of the Universe. At 'turnaround', the spherical region completely decouples from the expansion of the Uni-

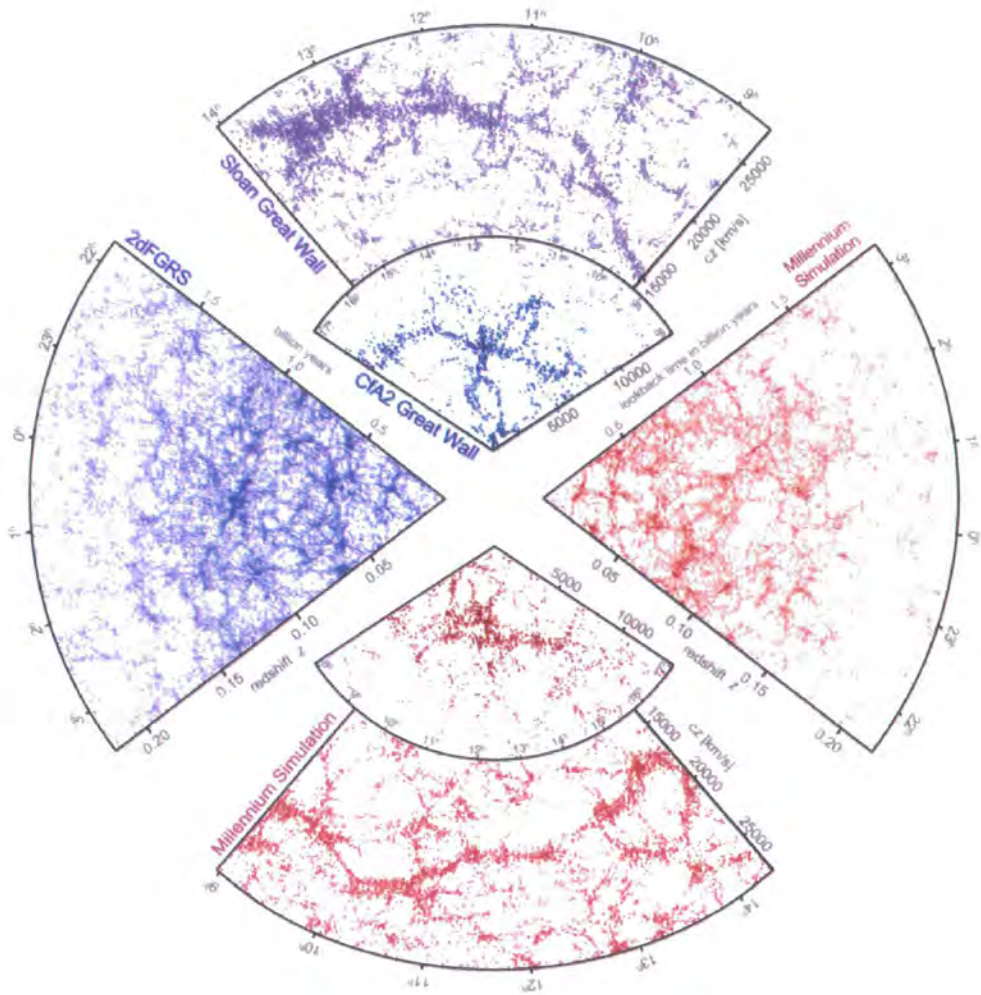

Figure 1.1: The galaxy distribution obtained from spectroscopic redshift surveys and from mock catalogues constructed from cosmological simulations. The small slice at the top shows the CfA2 'Great Wall' (Geller & Huchra 1989), with the Coma cluster at the centre. Drawn to the same scale is a small section of the SDSS, in which an even larger 'Sloan Great Wall' has been identified (Gott et al. 2005). This is one of the largest observed structures in the Universe, containing over 10,000 galaxies and stretching over more than 1.37 billion light years. The wedge on the left shows one-half of the 2dFGRS (Colless et al. 2001), which determined distances to more than 220,000 galaxies in the southern sky out to a depth of 2 billion light years. The SDSS has a similar depth but a larger solid angle and currently includes over 650,000 observed redshifts in the northern sky. At the bottom and on the right, mock galaxy surveys constructed using semi-analytic techniques to simulate the formation and evolution of galaxies within the evolving dark matter distribution of the 'Millennium' simulation (Springel et al. 2005b) are shown, selected with matching survey geometries and magnitude limits. This plot and its caption were taken from Springel, Frenk & White 2006).

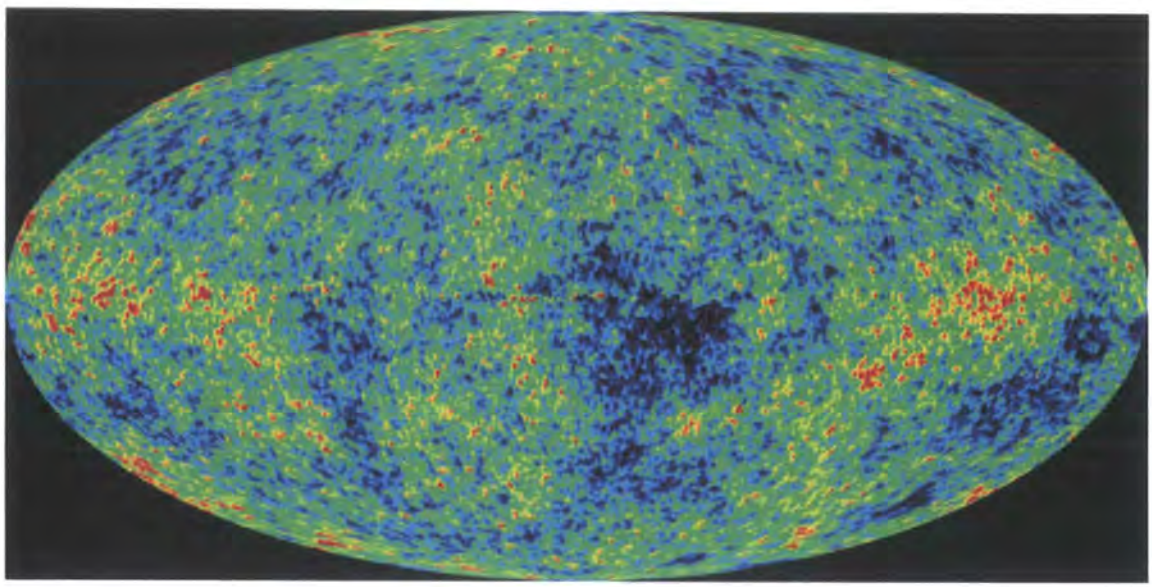

Figure 1.2: CMB fluctuations, as observed with WMAP. CMB temperatures are plotted using a red (warmest) → yellow → green → blue (coolest) colour scheme. This plot was taken from the NASA website

verse, and it begins to contract. A purely symmetric perturbation would collapse down to a singularity. We assume that in fact, due to inhomogeneities within the overdensity, the perturbation virializes, at the size where Potential Energy $= -2 \times$ Kinetic Energy, perhaps via a process of violent relaxation (Lynden-Bell 1967), to form a gravitationally bound structure in virial equilibrium, which we term a 'dark matter halo'.

Galaxy clusters at $z = 0$ have an overdensity, $\delta = (\rho - < \rho >)/ < \rho >)$, of ~ 1. How realistic is it that the perturbations in the CMB of order 10^{-5} can account for this? In linear theory, perturbations grow by a factor of $\sim a(0)/a(z)$ between redshift z and redshift zero. For perturbations to have grown by a factor of order 10^5, they must have been growing since at least $z \sim 10^5$. However, baryonic perturbations are unable to grow before recombination at $z \sim 10^3$, due to the pressure support of the baryons; the Jeans mass (i.e. the minimum mass unstable to gravitational collapse) for baryons before recombination is $\sim 10^{16} M_{\odot}$, but drops by a factor $\sim 10^{10}$ after recombination. This paradox is explained by the presence of dark matter, which, from the epoch of matter-radiation equality, at $z \sim 1 - 3 \times 10^5$ is gravitationally dominant. From the epoch of matter-radiation equality, fluctuations in the dark matter are able to grow between $z \sim 1 - 3 \times 10^5$ and recombination; dark matter is collisionless, so not pressure supported. After recombination, the baryons fall into the pre-existing dark matter potential wells. Indeed, this is a major motivation for the inclusion of dark matter in our cosmological

model.

We consider finally the primordial origin of the fluctuations in the density of the Universe. We believe that they arise from quantum mechanical fluctuations when the Universe was of a size comparable to the de Broglie wavelength of an electron. There was then a period of inflation when the Universe expanded exponentially, and the initial quantum mechanical fluctuations were stretched out onto classical spatial scales which would later be relevant for galaxy and cosmological structure formation (Guth 1981). Inflation also solves a number of other important cosmological fine-tuning problems e.g. the horizon problem and the flatness problem. The theory of inflation is highly attractive, but far from proven, although CMB observations provide substantial support for it (e.g. Spergel et al. 2006).

1.3 Modelling the Universe

It could be said that astronomy is a purely observational, taxonomic subject; there is presumably only one Universe which we can observe, and it is not possible to perform experiments upon it. However, the work of Newton (i.e. showing that the same gravitational force which operates on earth also governs objects in the celestial sphere) was highly influential in scientific studies of the cosmos. In principle, we assume that the same physical laws which operate locally, in experiments conducted upon the Earth, apply equally well to the Universe as a whole. It is therefore possible to perform computational experiments, using the best approximations of physics as we know it, to test how well our theories of the Universe agree with observations. We can create multiple universes in a computer, and compare their statistical properties to those of the observed Universe. One of these comparisons is shown in Fig 1.1 – the galaxy distribution obtained by applying a simple galaxy formation model to the 'Millennium' N-body simulation (Springel et al. 2005b) is shown opposite to that observed in a galaxy redshift survey.

Whilst the evolution of small gravitational perturbations can be treated using linear perturbation theory, their growth rapidly becomes non-linear, and we must resort to simulation. We discuss now the principle simulation techniques used to follow the growth of cosmic structures.

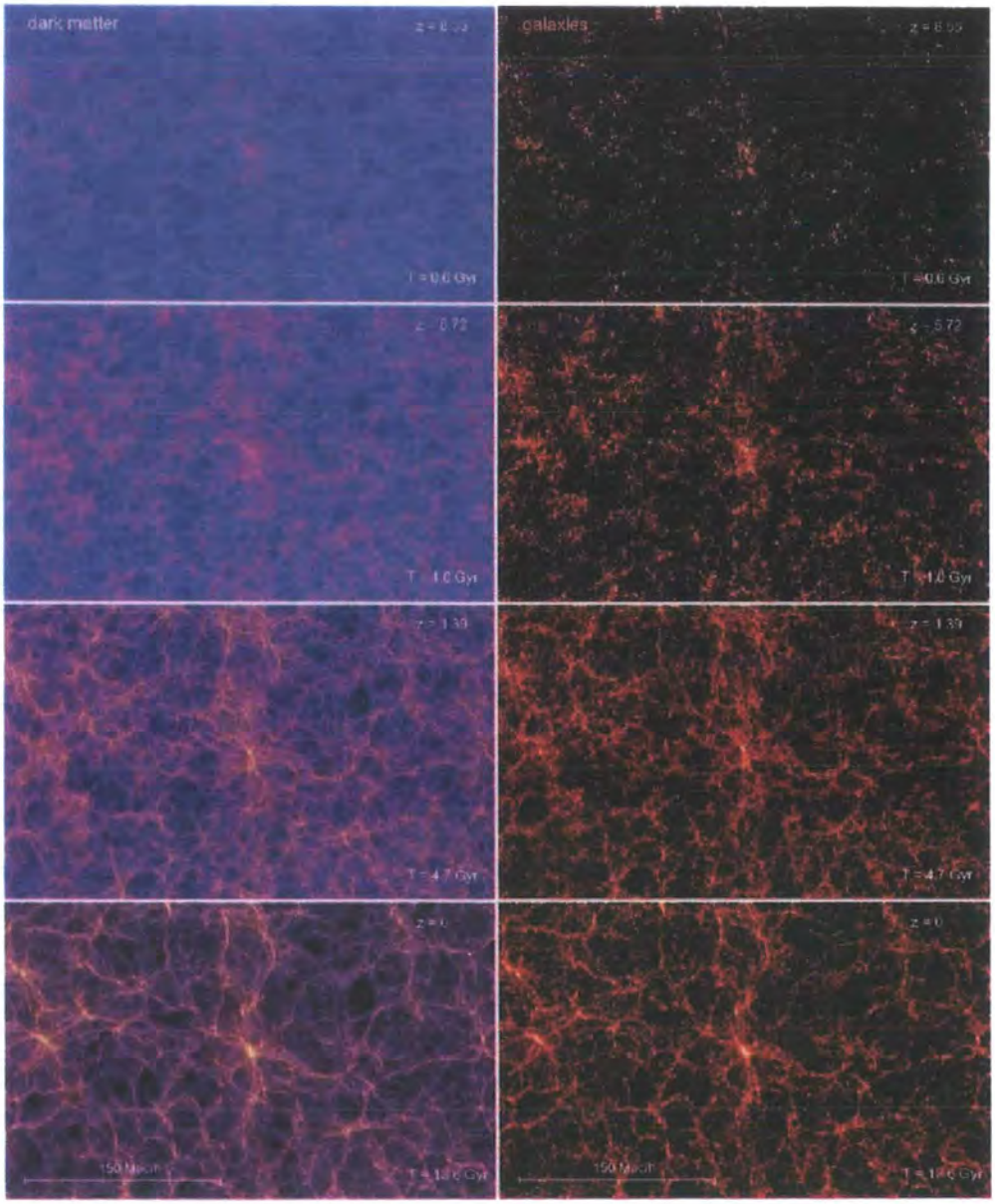

Figure 1.3: Evolution of the large-scale structure, obtained from cosmological simulations of the ΛCDM model. Left: The projected dark matter distribution in slices of thickness $15\,h^{-1}$Mpc, extracted at redshifts $z = 8.55$, $z = 5.72$, $z = 1.39$ and $z = 0$ from the Millennium N-body simulation of structure formation (Springel et al. 2005c). These epochs correspond to times of 600 million, 1 billion, 4.7 billion and 13.6 billion years after the Big Bang, respectively. The colour hue from blue to red encodes the local velocity dispersion in the dark matter, and the brightness of each pixel is a logarithmic measure of the projected density. Right: The predicted distribution of galaxies in the same region at the corresponding times obtained by applying semi-analytic techniques to simulate galaxy formation in the Millennium simulation (Springel et al. 2005c). Each galaxy is weighted by its stellar mass, and the colour scale of the images is proportional to the logarithm of the projected total stellar mass. This plot and its caption were taken from Springel, Frenk & White 2006).

1.3.1 N-body simulations

On the largest scales, the evolution of the Universe is dominated by gravity. Dark matter interacts only gravitationally and via the weak interaction (which we are able to ignore, since it is short range and weak at low energies), but does not interact electomagnetically, and so is dissipationless. Furthermore, whilst general relativity is required to describe the dynamics of the Universe as a whole, it turns out that Newtonian dynamics are sufficiently accurate to follow the gravitational behaviour of perturbations in the FRW metric of an expanding universe. In sum, collisionless cold dark matter is very simple to model, requiring us just to model (Newtonian) gravity. All we require is a very large, fast computer; the physics is all well understood. The evolution of the dark matter power spectrum in the early Universe is straightforward to compute, and can be described in terms of a power spectrum of primordial fluctuations (typically a scale-free Harrison-Zeldovich spectrum, which is likely to be generated from inflation) and a 'transfer function' (Bardeen et al. 1986), which converts the primordial spectrum of fluctuations generated during inflation to the power-spectrum of fluctuations at the time of decoupling. The transfer function takes care of any physical processes which modify the primordial spectrum – principally the effect of baryonic pressure which opposes the force of gravity at scales below the 'Jeans length' (the minimum size of a perturbation unstable to gravitational collapse) and the stagnation of fluctuations which enter the horizon before matter-radiation equality (this is the primary reason for the turnover in the power spectrum). We then take the power-spectrum, and randomly generate an initial density field, which we evolve using an N-body code.

The development of structure in the dark matter in an N-body simulation is shown in the left-hand panels of Fig. 1.3. In this figure, it is seen that at high redshifts, the Universe is very smooth. As redshift decreases, the development of clusters, filaments and voids can be seen. We now have a very accurate model of the dark matter, but it is luminous objects such as galaxies and quasars which we observe. We must next consider models which allow us to infer the distribution and properties of the galaxies, in the context of the hierarchical clustering of the dark matter.

1.3.2 Gas dynamical simulations

As discussed, it is easy for us to simulate the formation of structures in the dark matter. But where are the galaxies and what properties do they have? To determine this, we

must consider the properties of the baryons. We must simulate the baryonic gas, following processes such as cooling, star formation and feedback. Cooling is a dissipative process – galaxies are much more compact than their host dark matter haloes. It is baryonic dissipation through cooling which allows the baryons to lose energy and fall to the centre of the dark matter halo. Cooling, star formation and feedback are non-linear processes which are poorly understood. Baryonic processes can be modelled using numerical or analytic techniques. In this Thesis, we use an analytic technique, which we describe later. First of all, we will discuss the numerical tequniques.

Perhaps the most commonly used computational technique is smoothed particle hydrodynamics (SPH; Gingold & Monaghan 1977; Monaghan 1992). SPH is a method for following fluid flows. The fluid is divided into a large number of particles. Each particle moves with the fluid (i.e. it is a Lagrangian technique), and carries the latest local values of physical parameters (e.g. energy, entropy, pressure) with it. The values of these physical quantities at any point in the fluid are found by smoothing over the values associated with nearby particles in the simulation.

A big advantage of SPH in cosmological simulations is that it can be very easily combined with N-body simulations, since both are Lagrangian, particle-based techniques. Furthermore, since the particles trace the density, regions of high density are simulated at correspondingly high resolution in a very natural fashion. However, SPH codes also have a number of problems. Since the particles trace the density, it is very difficult to increase the resolution as a function of any physical parameter except to increase it in proportion to the density. SPH resolves regions of low density poorly. SPH also has a poor treatment of shocks, since physical quantities are derived by smoothing over the narrow shock boundary where the physical state of the gas changes significantly over a short spatial scale. It is also difficult to make galaxies with realistic scale sizes and angular momenta using SPH. Gas cools and fragments into clumps, which rapidly lose angular momentum to the outer parts of the dark matter halo due to dynamical friction, and fall towards the centre of the halo to form a compact bulge. This occurs even in dark matter haloes with a very quiet merger history, which would be expected to host spiral galaxies. It appears that this problem is to some extent some one of numerics in SPH, requiring an improved numerical resolution (Governato et al. 2004) and an appropriate numerical treatment of the two-phase (hot gas and cold gas) insterstellar medium (Okamoto et al. 2003; Springel & Hernquist 2003; Scannapieco et al. 2006), and also, largely, one of including sufficiently strong feedback in the earlier stages of galaxy formation (Weil et al. 1998; Okamoto et al.

2005). Many of the older SPH codes explicitly conserved energy, but it turns out that codes which perform an explicit conservation of entropy have much better convergence properties in both energy and entropy (Springel & Hernquist 2002).

The disadvantages of Lagrangian tequniques, in particular the poor treatment of shocks and of low density regions, mean that there is also a demand for grid-based methods. Initially, these were done using fixed grids. To circumvent the problem of insufficient resolution in regions where high resolution is required (e.g. the central parts of dark matter haloes), there has been a lot of work done to produce codes which refine the mesh to higher spatial (or time) resolution where this is required (Bryan & Norman 1997; Teyssier 2002; Quilis 2004). These codes are referred to as 'adaptive mesh refinement' (AMR) codes. One particularly exciting application has been a simulation of the formation of one of the first stars in the Universe (Abel et al. 2002). In the future, AMR is likely to produce many exciting results. However, it has taken a long time for this method to flourish since there are a large number of technical difficulties with AMR codes, particularly when refining the meshes whilst conserving all quantities which should be conserved. There is also a difficulty in combining it (an Eulerian, mesh-based method) with N-body codes, which are based on particles. Hopefully these problems will be successfully overcome in the next few years.

Even at the fairly high resolution of a gas dynamical simulation, whether done by SPH or AMR, there are still processes occuring on scales smaller than are accesible to direct simulation – therefore 'sub-grid physics' is still required. For example, a sub-grid model of the multi-phase gas in star-forming regions allowed a far more realistic treatment of star formation and feedback than had been possible previously (Springel & Hernquist 2003). Much of this physics is in fact modeled in a similar way to the physics we include in our semi-analytic modelling. We go on to discuss our modelling of this physics in the next section.

1.4 An overview of semi-analytic galaxy formation

Semi-analytic models of galaxy formation attempt an ab initio calculation of the formation and merging of large populations of galaxies. We put together a wide range of ingredients in our semi-analytic model, which we will refer from now on as '`GALFORM`' (Cole et al. 2000) :

- A cosmological background model

- Dark matter halo merger trees – dark matter clumps are built up by mergers of smaller dark matter clumps.
- Shock-heating of gas, gas cooling and condensation.
- Star formation.
- Feedback from star formation and from accretion of gas onto supermassive black holes.
- The formation of galactic discs by the cooling and condensation of gas in dark matter haloes.
- The formation of galactic spheroids via mergers of discs or disc instabilities.
- Dark matter halo mergers, dynamics of galaxies within dark matter haloes and galaxy mergers.
- Chemical enrichment.
- Prediction of the observable radiation from galaxies using stellar spectral synthesis and a simple dust model.

In Fig. 1.4, we illustrate pictorially our semi-analytic model (left), and show for comparison the dark matter in the region of an N-body simulation on which our galaxy formation model is superposed (right). The region in Fig. 1.4 is in fact a highly zoomed-in region taken from the Millennium simulation we showed earlier (Fig. 1.3). In Fig. 1.4, the main dark matter halo is represented by a white fuzzy blob/sphere. Subhaloes are represented by smaller, brighter fuzzy blobs superimposed on the main halo. Galaxies form in dark matter haloes and subhaloes. We represent them in terms of two stellar components – a bulge and a disc, and one cold gas component – a disc. In the picture, galaxy bulges and discs are represented by solid spheres and discs respectively. In our semi-analytic modelling, we take simple, symmetric models to describe dark matter and galaxies. These are the kinds of simplifications which allow us to perform our calculations much faster, for many more galaxies and for a much wider range of physical models than would be possible using an N-body or gas dynamical simulation. The simplifications we make may seem crude compared to the high spatial resolution and great computing power used in gas-dynamical simulations. But we remind the reader that gas-dynamical simulations are highly dependent on uncertain sub-grid physics, and have a number of other problems and difficulties as discussed above. As far as possible, we match our prescriptions to the results of more detailed numerical simulations, for example N-body and gas-dynamical simulations, detailed models of stellar population synthesis, radiative transfer calculations and detailled modelling of atomic processes in gas cooling. This

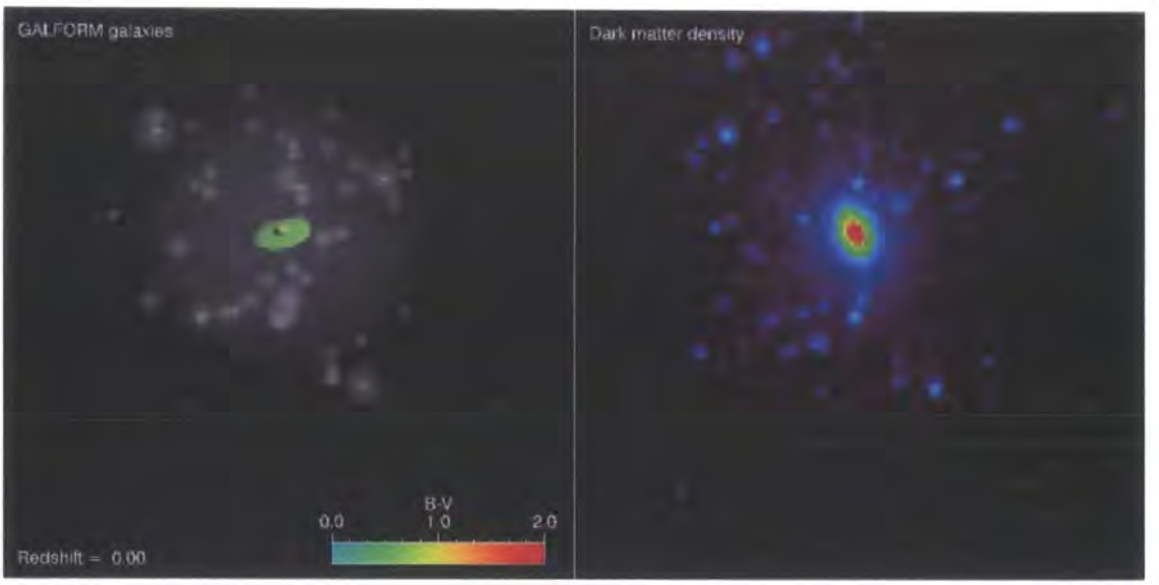

Figure 1.4: A region of dark matter from an N-body simulation compared with the same region as re-modelled by semi-analytics, using a dark matter merger tree taken from the corresponding region of the simulation. Left: The main dark matter halo is represented by a white fuzzy blob/sphere. Subhaloes are represented by smaller, brighter fuzzy blobs superimposed on the main halo. Galaxies hosted by the haloes and subhaloes are represented by solid discs (representing the disc component of the galaxy) and solid spheres (representing the bulge component of the galaxy). The linear scale sizes of the discs and bulges as plotted are proportional to their scale sizes as inferred from the semi-analytic modelling. The galaxies have been expanded significantly in the picture in comparison to the dark matter so that they can be seen. The colours of galaxy discs and bulges in the illustration represent their B-V colours, as indicated by the key. Right: The same region, showing just the dark matter density from the original N-body simulation. The density of a region is indicated by its colour: red (densest) → yellow → cyan → blue → black (least dense). This image was made by John Helly.

leaves a number of less well understood prescriptions which we adjust to reproduce, as far as possible, local observations of the Universe, as described in §1.5
We now consider the processes involved in galaxy formation in slightly more detail.

1.4.1 A cosmological background model

Although there has been a dramatic improvement over the past decade in observational constraints on the basic parameters of the cosmological model, uncertainties still remain. Some of these uncertainties may have significant implications for our modelling of galaxy

and quasar formation, particularly at high redshift. To some extent, we could take a physical model of galaxy formation and use it to provide constraints on the background cosmology. However, in practice, there are far more uncertainties due to our modelling of galaxy than due to the cosmological model. Therefore, in general, we take the best cosmological model as determined by the observations, and concentrate on altering the galaxy formation model (where there is more that one plausible cosmological model, we usually find that the same galaxy formation model can be used in each, with minor adjustments to the parameters). Specifically, the cosmological models we use may have be constrained by a combination of some or all of the following: CMB measurements (e.g. Spergel et al. 2003), galaxy clustering measurements (e.g. Cole et al. 2005), supernovae measurements (e.g. Riess et al. 1998; Perlmutter et al. 1999), determinations of the Hubble constant from the distance ladder (Freedman et al. 2001) or X-ray cluster abundances (e.g. Eke et al. 1996). This combination of data is typically done using maximum-likelihood statistical techniques (Sánchez et al. 2006; Spergel et al. 2006).

For most of this thesis, we use a ΛCDM cosmology based on the WMAP 1-year results and initial 2dFGRS $P(k)$ measurement (Spergel et al. 2003). For our first model used in Chapters 2-4 we use a 'Concordance Cosmology' with cosmological parameters $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_{\rm b} = 0.04$, $n_s = 1$ (fixed), $\sigma_8 = 0.9$ and $h = 0.7$. In Chapter 5, we consider a cosmology with a slight variation upon these parameters to take account of more recent results, and we also consider how our galaxy and black hole formation model must be varied (and the resultant variation in predictions at high redshift) when we use a cosmological model with an extra parameter to allow for a running spectral index (Sánchez et al. 2006). The parameters used in all of the cosmological models we consider are summarized in table 5.1 of Chapter 5

1.4.2 Dark matter halo merger trees

In cold dark matter models, the variance in density fluctuations is largest on small scales, so perturbations on small scales collapse before those on large scales. We refer to collapsed, virialized regions in the dark matter density field as dark matter haloes. Since small scales collapse first, this means that low mass haloes tend to form first, then merge to form higher mass haloes. Press & Schechter (1974) proposed a simple model of the mass function of dark matter haloes as a function of redshift by considering the overdensity in spherical regions, and assuming a gaussian density field. Bond et al. (1991) and Bower (1991) extended this work, and provided a simple method for predicting the mean mass function

of the small dark matter haloes which later merge to form any particular dark matter halo. Lacey & Cole (1993) developed a formula for producing merger trees using this 'extended Press-Schechter' (EPS) theory.

Lacey & Cole demonstrated that EPS merger trees provided a very good description of the evolution of the dark matter in N-body simulations (Lacey & Cole 1994), at least to the accuracy then available using N-body simulations. More recently, unpublished work by Cedric Lacey (priv. comm.) and work in preparation by Helly et al. (John Helly, priv. comm.) show that the merger trees in the Durham semi-analytic model agree well with the merger trees in N-body simulations, specifically the 'Millennium' simulation. EPS merger trees can have far superior mass resolution to N-body ones. The main drawback of EPS trees is that they tend to become less accurate as the time interval over which the trees are grown is increased (Somerville et al. 2000) – the EPS technique is consistent with maintaining a Press-Schecter mass function at earlier redshifts given a Press-Schecter at low redshifts, but is is known that the mass function at all redshifts is more consistent with that of Sheth et al. (2001). Furthermore, the EPS formalism does not treat mergers between two masses symetrically when the higher and lower masses are swapped, as it should if the theory is to be self-consistent. These limitations are discussed in more detail in Benson et al. (2005).

Using the method of Lacey & Cole, we can produce large numbers of realistic dark matter 'merger trees' very quickly. We produce the majority of our results using this method, but in Chapter 6, we use merger trees taken from N-body simulations.

1.4.3 Gas shock heating and cooling

We assume that when gas falls into a dark matter halo, it is shock-heated to the virial temperature of the halo. We assume that it then cools via various processes:

- Decay of collisionally excited molecular hydrogen. These excitations may be rotational or vibrational. This is most important for $T \sim 10^4 K$.
- Collisions between atoms or ions and electrons excite these atoms or ions to higher energy levels. There are then transitions to lower energy levels, causing emission of photons and cooling. Most of this atomic line cooling is through metal lines, which only occurs when the gas is ionized. Therefore cooling in the temperature range required to set up the plasma of ionized gas $10^4 K < T < 10^6 K$ occurs much more efficiently in gas which has been chemically enriched.
- Bremsstrahlung radiation as electrons are accelerated in an ionized plasma.

• We ignore inverse compton scattering of CMB photons – this is only significant at very high redshifts.

The first three of these processes are two-body processes which depend on the number-density of electrons squared. Therefore, it is important for us to know the density profile of the gas in the halo. We often assume it follows the density profile of the dark matter halo, which we take from a simple model fitted to N-body simulations (Navarro et al. 1997). Cooling also depends on density and temperature. Using these inputs, we are able to calculate the cooling rate of the gas, using the results of the detailed modelling of Sutherland & Dopita (1993). In its simplest form, the cooling model assumes that upon halo formation, gas is shock-heated to the virial temperature of the halo. Taking a simple assumption for the density profile of the gas in the halo, we calculate the cooling time of the gas. We then calculate the 'cooling radius', which is the radius within which the gas has had time to cool since halo formation, and we propagate this cooling radius outwards with time since halo 'formation'[4].

The method we use for finding cooling rates of gas in dark matter haloes has been shown to produce results which agree well with those found from SPH calculations of cooling (Yoshida et al. 2002; Helly et al. 2003). We should also test these models against the real Universe rather than only against each other of course. The models predict a flux of X-rays from the cooling gas in galactic sized haloes, which for many years could not be detected (e.g. Benson et al. 2000). However a detection has recently been reported (Pedersen et al. 2006), albeit at a lower flux than predicted by the simple models of gas cooling we use in our modelling. There is obviously much still to be understood.

1.4.4 Star formation

When gas cools, it loses its pressure support. This allows it to fall gravitationally to the centre of its dark matter halo, where it is able to form stars. In order to form stars, the gas must both be cool and dense. Therefore, we only allow star formation after the cold gas has had time to fall to the centre of the halo, which we assume occurs in one dynamical time.

When a mass of gas is converted into stars, we assume that it forms stars with a characteristic distribution of masses, the 'initial mass function' (IMF). This model explains the observed stellar luminosity function well (Salpeter 1955). The true IMF is still uncer-

[4]Halo 'formation' is defined to occur when a halo increases its mass (whether by merging with less massive haloes or smooth accretion of unresolved haloes) by a factor of two.

tain, and may or may not vary from region to region. We often use an IMF similar to the Salpeter IMF for star formation in discs, and an IMF with a particularly high fraction of high mass stars in bursts (see Chapter 2).

We assume that star formation has a characteristic timescale, $\tau_{\rm star}$. Presumably, the rate of star formation is dependent on the mass of cold gas available to form stars, $M_{\rm cold}$. By dimensional arguments then, we can propose a simple rule to calculate the rate of star formation, $\dot{M}_{\rm star}$ i.e.:

$$\dot{M}_{\rm star} = M_{\rm cold}/\tau_{\rm star} \tag{1.2}$$

Star formation is poorly understood, but it is reasonable to assume that the star formation timescale will scale with some characteristic timescale in the problem. We generally assume that $\tau_{\rm star}$ is either a constant, perhaps related to some intrinsic timescale of star-formation itself within molecular clouds, or that $\tau_{\rm star}$ scales with the dynamical timescale of the galaxy.

1.4.5 Feedback

A comparison of the galaxy luminosity function and the dark matter halo mass function shows that the efficiency of galaxy formation must vary quite strongly with halo mass. Galaxies form most efficiently in dark matter haloes of mass $\sim 10^{12} M_{\odot}$ (Eke et al. 2004). The efficiency of galaxy formation decreases as we consider haloes more or less massive than this mass. To some extent, this is because gas is virialized to high temperatures in high mass haloes, and gas at high temperatures can cool only by bremsstrahlung, which is less efficient than metal line or molecular hydrogen cooling. Therefore gas is less able to cool in high mass haloes (Rees & Ostriker 1977; White & Rees 1978). Feedback is necessary for a number of reasons, in particular to explain why the shape of the luminosity function is very different from the shape of the mass function of dark matter haloes (White & Frenk 1991). We include four types of feedback:

- Supernova feedback. A fraction of all stars formed will end their lives as high mass supernovae, which explode after a few million years, returning mass and energy to the inter-stellar medium or the intergalactic medium. We model this by assuming that cold gas is reheated and removed from the disc. We generally assume that the rate of gas reheating is a fixed fraction of the star formation rate. The gas is then returned to the hot gas halo[5].

[5]It could also be expelled from the halo altogether, but we distinguish this type of supernovae feedback in the next bullet point.

• Feedback from superwinds. There is evidence of massive outflows at very high velocities – up to 600km s^{-1} – coming from Lyman break galaxies with high rates of star formation (Adelberger et al. 2003; Wilman et al. 2005). We include a form of feedback where we give gas kinetic energy rather than thermal energy. This is done as follows:

$$\dot{M}_{\rm eject} = f_{\rm sw} \times \dot{M}_{\rm star\ formation} \tag{1.3}$$

where:

$$f_{\rm sw} = f_{\rm sw0} \qquad \text{for} \qquad V_{\rm circ} < V_{\rm sw}, \qquad \text{and :} \tag{1.4}$$

$$f_{\rm sw} = f_{\rm sw0} \times (V_{\rm sw}/V_{\rm circ})^2 \qquad \text{otherwise.} \tag{1.5}$$

where $\dot{M}_{\rm eject}$ is the rate of removal of gas mass from the halo by the superwind, $\dot{M}_{\rm star\ formation}$ is the rate of star formation (including high mass supernovae which do not leave long-lived stellar remnants), and $V_{\rm circ}$ is the circular velocity of the disc (for quiescent star formation) or the circular velocity of the bulge (for burst star formation). $f_{\rm sw}$ and $V_{\rm sw}$ are free parameters – these free parameters can be different depending on whether the star formation is in the quiescent or in the burst mode, but in our standard model we use the same values for both modes: $f_{\rm sw} = 2$ and $v_{\rm sw} = 200 kms^{-1}$.

In our fiducial model, the superwinds are driven by star formation, but given the large energy budget required, it may be that they are partly driven by AGN activity associated with star formation (Benson et al. 2003a). It is possible to allow gas expelled by a superwind to fall back into the halo depending on how its velocity compares with the escape velocity of the halo (this is considered in the Bower et al. (2006) model, which is used in Chapter 6 of this Thesis). However, in our fiducial model we generally assume that the gas expelled by a superwind is never recaptured, even when this halo becomes incorporated into a more massive halo further down the merger hierarchy.

• Photoionization feedback. Ultraviolet photons are emitted from galaxies and quasars. These are able to ionize gas in neighbouring haloes. This may heat the gas up, increasing its pressure support. This prevents some fraction of the baryons from cooling in low mass dark matter haloes as the temperature of the heated gas exceeds the virial temperature of the halo[6]. Furthermore, photoionization changes the ionization state of the hot gas,

[6]The virial temperature of a halo is the temperature of gas which has the pressure required to counteract its gravitational attraction towards the centre of the halo. Therefore, less pressure support is required to counteract the gravitational attraction of the gas towards the centre of haloes of lower mass. So if photoionization heating does not vary with halo mass, then it will cause a higher fraction of baryons on lower mass haloes to have the temperature and pressure required to resist collapse towards the centre.

removing channels for cooling.

• AGN feedback. The energy released by accretion of gas onto a central supermassive black hole balances the energy which would be released if hot gas in the halo were allowed to cool. This is only effective in haloes where the hot gas is in quasihydrostatic equilibrium. Since black holes are at galactic centres, AGN feedback is most effective in the central regions of haloes where the gas density is likely to be high and cooling times very short, and therefore where feedback most required. We discuss AGN feedback further in Chapter 6.

1.4.6 Morphology and sizes of galactic discs and spheroids

Why do disc galaxies rotate? Since dark matter haloes are not completely spherically symmetric, tidal forces in the large-scale density field surrounding these haloes (and even more significantly, the asymmetric overdensities which are in the process of collapsing into virialized haloes) lead to a tidal torque, which generates a rotation in these dark matter haloes (Hoyle 1949; Peebles 1969). Initially, the angular momentum profile of gas is likely to follow that of the halo. We assume that when baryonic gas within dark matter haloes cools, it falls to the centre of the halo due to its loss of pressure support, conserving its angular momentum, until it is rotationally supported (Fall & Efstathiou 1980). We assume that the resultant gas disc has an exponential profile.

We assume that spheroids and galactic bulges are produced by mergers of pre-existing galaxies. The resulting bulge has an $R^{1/4}$ law profile. Our model for bulge formation is described in much greater detail in §2.2.2 in the next Chapter.

1.4.7 Galaxies within haloes

When haloes merge, we assume that the most massive pre-existing galaxy will be at the centre of the new halo. We refer to this galaxy as the 'central' galaxy. Any other pre-existing galaxies we refer to as 'satellite' galaxies, and we assume that these satellite galaxies enter the newly formed halo on random orbits. We assume that satellite galaxy orbits decay through dynamical friction (Chandrasekhar 1943; Binney & Tremaine 1987; Lacey & Cole 1993). Dynamical friction occurs when a massive body (i.e. the satellite galaxy) moves through a sea of less massive particles (i.e. the halo) – there is a net loss of angular momentum from the galaxy to the halo the due to the summed contribution of the momentum changes of the satellite galaxy in its direction of motion from its multiple interactions with the sea of low mass particles. Alternatively, a more intuitive way of

viewing dynamical friction is that as the satellite galaxy moves through the halo, it induces a wake behind it. Thus there is a higher density of mass behind the galaxy than in front of it, and thus a net force on the galaxy in the opposite direction to its direction of motion, causing it to slow down and lose angular momentum.

When a satellite galaxy gets sufficiently close to a central galaxy such that their half-mass radii overlap, then we assume that the two galaxies merge. As discussed above and in Chapter 2, this may lead to formation of a galactic bulge.

1.4.8 Chemical enrichment

After nucleosynthesis, the Universe is $\sim$ 75% hydrogen and $\sim$ 25% helium by mass, with small traces of other light elements such as lithium and beryllium. However, on earth, we find that there are over 100 elements in the periodic table. For simplicity, astronomers often refer to all elements more massive than helium generically as 'metals'. Stars are fuelled by the conversion of mass into energy ($E = \epsilon Mc^2$, where approximately $\epsilon = 0.007$ of the mass of the star can be released by stellar nucleosysthesis). Initially, this is by the 'burning' of hydrogen to form helium. When the star begins to run out of hydrogen, conversion of hydrogen to helium slows down, the star loses some support from the radiative pressure of the photons produced, and it collapses somewhat. The resultant increase in temperature and density at the centre may allow production of other elements, particularly in more massive stars. When a massive star reaches the end of its life, it will explode (a supernova) releasing these elements. These 'metals' can be widely spread through the galaxy and its surroundings due to the energy from supernova explosions, by both the initial velocity imparted from the explosion, and the subsequent mixing of the gas from energy injection in general.

We model star formation and chemical enrichment with an instantaneous recycling approximation (Tinsley 1980). We assume that SnII explode immediately upon formation, and that the metals they release immediately are mixed uniformly in the cold gas of the galaxy, which may subsequently be reheated by star formation feedback. Thus the cold gas from which stars form will be chemically enriched, and we can then calculate the metallicity of stars formed from that gas. It is essential for us to follow chemical evolution for three reasons:

- The metallicity of gas affects its cooling rate – the principle cooling mechanism for gas in the temperature range $10^4 - 10^6 K$ is metal line cooling.
- The chemical composition of stars affects their spectra.

• Dust forms from metals, and the starlight which we observe is affected by dust extinction and re-radiation.

1.4.9 Stellar spectral synthesis and dust modelling

Equilibrium models of stellar structure are, in the most part, very well understood. It is possible to make very accurate models of the spectral energy distribution (SED) of any star if we know its mass at birth, its metallicity at birth and its age. This allows an integrated SED to be calculated as a function of IMF, initial metallicity and age. This calculation is called stellar spectral synthesis, and was pioneered by Beatrice Tinsley in the 1970's (Tinsley 1972). Using the star formation history of the galaxy, we add up the contributions of single stellar populations from all lookback times to form a composite stellar spectrum for the galaxy. We use the most up to date Bruzual & Charlot (1993; 2003) models for the spectral evolution of 'single stellar populations' (i.e. populations of stars with the same metallicity and formation time, and a given IMF). While stars in equlibrium are fairly well understood, the later stages of stellar evolution can be hard to model. The contribution of thermally pulsating stars on the asymptotic giant branch is very important for young stellar populations, and is particularly uncertain (Maraston 1998, 2005).

We also consider dust extinction. We take our modelling of the metallicity of the gas in a galaxy, and assume that a fraction of these metals form dust. The simplest models of dust extinction assume that there is a thin slab of dust between us and the galaxy. We assume, far more realistically, that the dust, gas and stars in the galaxy are all mixed up together. We also assume that a small fraction of the dust is in molecular cloud clumps surrounding the youngest stars. Thus from the dust mass in the galaxy, the galaxy scale-size and the galaxy geometry, we are able to calculate an optical depth to stars in the galaxy. We then calculate the dust extinction using a simple model based on more sophisticated radiative transfer calculations.

1.5 Fiducial model results at z=0

Our standard procedure is to set the parameters of the prescriptions outlined in §1.4 to create models of the Universe which agree well with a fiducial set of well-observed galaxy properties. We give full references to the large number of observations used in Tables 1.1 and 1.2. We show in Figs. 1.5 and 1.6 the match of the Baugh et al. (2005) model to

Table 1.1: Key and reference list for the observational data in Figs. 1.5, 5.3, 5.5 and 6.1.

Function plotted	Position of panel	
Text in key	Point type	Reference
b_J-band luminosity function	Top-left	
ESP	Red	Zucca et al. (1997)
2dF – Norberg	Green	Norberg et al. (2002)
Durham / UKST	Black	Ratcliffe et al. (1998)
Loveday	Magenta	Loveday et al. (1992)
K–band luminosity function	Top-right	
Glazebrook	Orange	Glazebrook et al. (1995)
Gardner	Black	Gardner et al. (1997)
Mobasher	Powder blue	Mobasher et al. (1993)
2dF-2MASS Kron	Green	Cole et al. (2001)
Star formation rate density	Middle-left	
–	Green triangle	Gallego et al. (1995)
–	Blue circle	Lilly et al. (1996)
–	Magenta square	Madau et al. (1996)
I-band Tully-Fisher relation	Middle-right	
–	Black open squares	de Jong & Lacey (2000)
Disc scale-length distribution	middle-left & middle-right)	
–	Black squares; open triangles	de Jong & Lacey (2000)

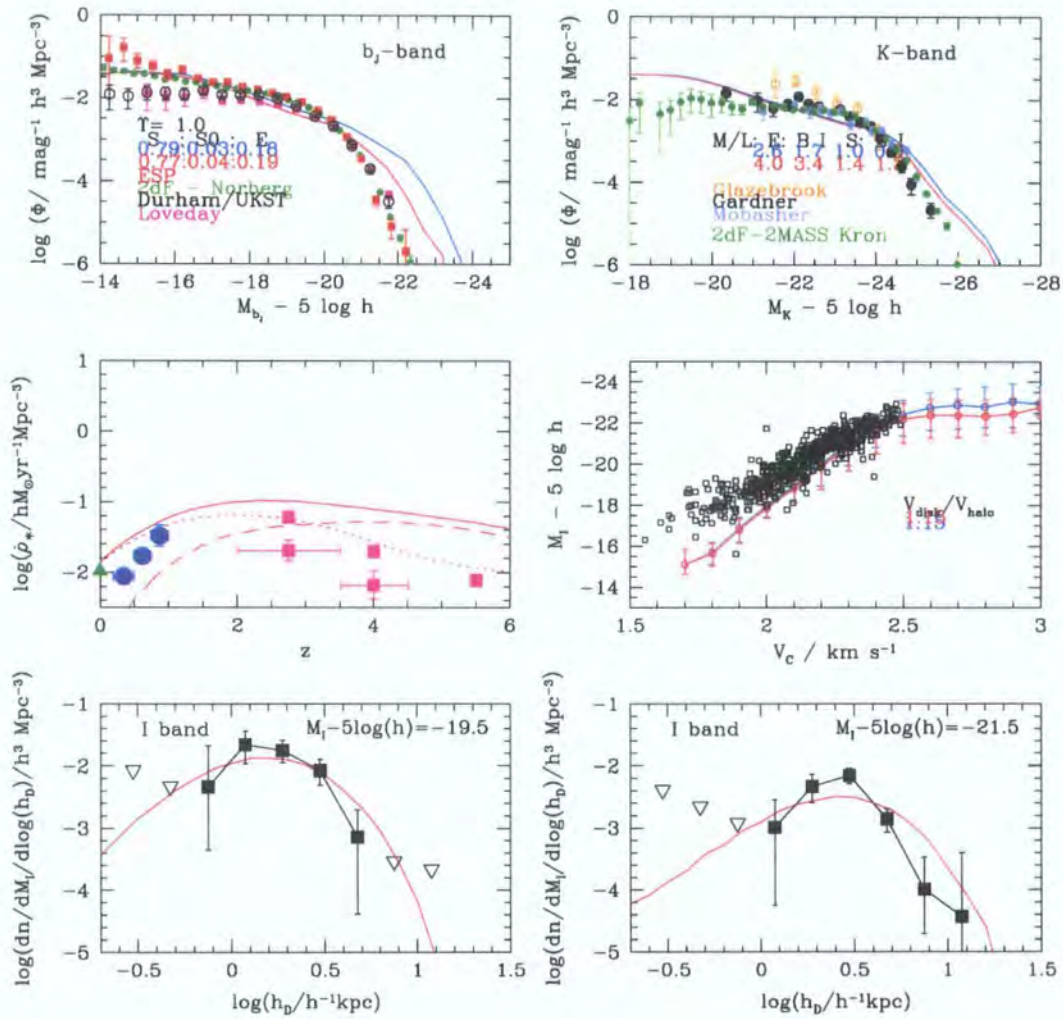

Figure 1.5: The predictions of the Baugh et al. (2005) GALFORM model using our standard dust extinction model (red lines) and switching off dust extinction (blue lines). Where the blue line is not visible, the two results are identical. We also plot a variety of observational data – the observational data in each plot are described in Table 1.1 and §1.5. Top-left: The $z = 0$ b_J-band luminosity function. Top-right: The $z = 0$ K-band luminosity function. Middle-left: The star formation rate per unit volume as a function of redshift (solid – total; dotted – quiescent; dashed – burst). Middle-right: The I-band Tully-Fisher relation at $z = 0$. Bottom-left: The distribution of disc scale-lengths (for galaxies with $-20 < M_I < -19$). Bottom-right: The distribution of disc scale-lengths (for galaxies with $-22 < M_I < -21$).

this standard set of data. We use the Baugh et al. galaxy formation model (sometimes with small variants) throughout this Thesis, except for our comparison in Chapter 6 to an alternative model which includes black hole feedback.

1.5.1 The galaxy luminosity function

We require that the model produce a good match to the local B-band and K-band luminosity functions. In particular, since K-band magnitudes have a reasonably tight correlation with stellar masses, the K-band luminosity function is a standard test of the feedback model (Benson et al. 2003a). The faint end of the luminosity function strongly constrains supernova feedback and the photoionization supression of cooling, whilst superwind and AGN feedback are required to fit the bright end of the luminosity function. The faint end of the luminosity function also depends on galaxy merging – without merging of galaxies, a faint end which is too steep may be produced (White & Frenk 1991; Cole et al. 1994).

We show this comparison in the top-left and top-right panels of Fig 1.5.

1.5.2 The Tully-Fisher relation

The Tully-Fisher relation (Tully & Fisher 1977) is a correlation between galaxy luminosity and circular velocity (measured at the peak of the rotation curve). Since circular velocity at any radius depends upon the mass enclosed within that radius, and since galaxy luminosity is a strong function of stellar mass, then the Tully-Fisher relation connects the dynamics of galaxies and their stellar content. We usually find that our model galaxies have a higher circular velocity than would be expected from their luminosity. This currently appears to be a problem with all semi-analytic models which calculate the circular velocity at the galaxy half-mass radius, and which make a realistic calculation of galaxy size. It seems that our galaxies may be too condensed. This may be a problem with our adiabatic contraction model, which allows the dark matter halo to contract in response to the gravitational pull of the baryons which have condensed in the central galaxy. The zero-point we find is much closer to the data if we consider the circular velocity of the halo at the virial radius, or if we neglect adiabatic contraction.

We show this comparison in the middle-right panel of Fig 1.5.

Table 1.2: Key and reference list for the observational data in Figs. 1.6, 5.4, 5.6 and 6.2.

Function plotted	Position of panel	
Point type	Reference	Additional notes
$M_{\rm gas}/L_B$ as $f^n(M_B)$	Top-left	
Black squares	Huchtmeier & Richter (1988)	for $M_B - 5\log h < -16$
Black squares	Sage (1993)	for $M_B - 5\log h > -16$
$M_{\rm gas}/L_B$ as $f^n(M_B)$	Top-right	
Arrows	Knapp et al. (1985)	
$Z_{\rm gas}$ as $f^n(M_B)$	Middle-left, middle-right	
Black squares	Zaritsky et al. (1994)	
$\rho_{\rm cold}/\rho_{\rm crit}$	Bottom-right	
Black open triangle	Zwaan et al. (1997)	
Black open circles	Turnshek (1997)	
Black squares	Storrie-Lombardi & Wolfe (2000)	

1.5.3 The distribution of disc scale-lengths

This is a useful test for our model of how disc galaxies acquire their angular momentum – i.e. of our model for the angular momentum profile of dark matter haloes, and of the assumption that as gas cools it conserves its angular momentum. Disc scale-lengths are also affected by feedback; as noted by Cole et al. (2000), if feedback is weak, discs form at high redshift in small haloes, and are therefore small, whilst if feedback is strong, then discs will form at lower redshift in larger dark matter haloes and be larger.

We show this comparison in the bottom-left and bottom-right panels of Fig 1.5.

1.5.4 The gas fraction as a function of luminosity

We constrain the overall scaling of the star formation timescale by comparing the gas fractions of L^* spirals to observational data. We constrain the scaling of the effective star formation timescale with the circular velocity of the galaxy by comparing gas fractions of faint spirals to observational data. Star formation is less efficient in shallower potential wells, since even with a constant timescale for the conversion of cold gas to stars, the

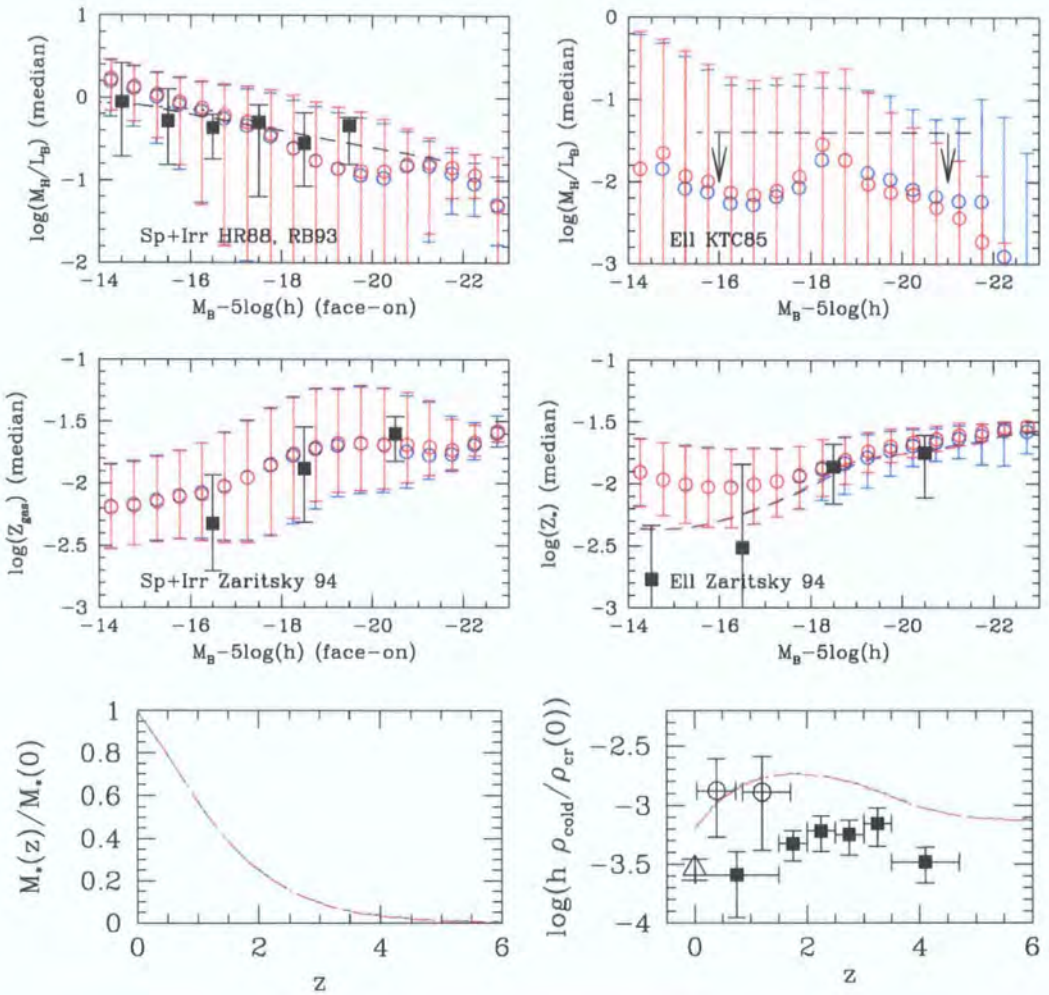

Figure 1.6: Further predictions of the Baugh et al. (2005) GALFORM model using our standard dust extinction model (red) and switching off dust extinction (blue). We also plot a variety of observational data – the observational data in each plot are described in Table 1.1 and §1.5. Top-left: Gas mass to B-band luminosity ratio as a function of M_B for spirals and irregulars. Top-right: Gas mass to B-band luminosity ratio as a function of M_B for ellipticals. Middle-left: Gas metallicity of spirals and irregulars as a function of M_B at $z = 0$. Middle-right: Stellar metallicity of ellipticals as a function of M_B at $z = 0$. Bottom-left: The fraction of the $z = 0$ stellar mass formed as a function of redshift. Bottom-right: The redshift evolution of the universal density of cold gas (as a fraction of the critical density).

effective timescale may be lengthen due to the greater effect of star formation feedback in shallower potential wells.

We show gas fractions as a function of luminosity and compare them to observational data in the top-left (spirals & irregulars) and top-right (ellipticals) panels of Fig 1.6.

1.5.5 Galaxy metallicity as a function of luminosity

This is useful for assessing whether our yield (the mass of metals produced per unit mass of star formation – this is a function of the IMF) is correct, and also whether the fraction of the metals produced which are ejected from the galaxy by feedback processes is correct.

In Fig 1.6, we compare the output of our model to the metallicity of the gas in spiral discs (middle-left) and the metallicity of stars in ellipticals (middle-right). We obtain a reasonable fit.

1.5.6 The global density of cold gas as a function of redshift

This is sensitive to whether we use a star formation timescale which is the same for every galaxy (and therefore constant with redshift) or a star formation timescale which scales with dynamical time (so that star formation is faster at higher redshifts, since characteristic timescales are shorter at higher redshifts). Star formation timescales which are invariant (decreasing) as redshift increases will lead to there being more (less) cold gas in the Universe at high redshifts.

We compare the results of our model to observational data in Fig 1.6 (bottom-right).

1.5.7 The star-formation history of the Universe

It is not so clear which particular parameters this constrains, since it depends on many. However, it is still an important prediction to make. It must be remembered that there are large uncertainties in observed star formation rates, particularly at high redshift. In particular, dust corrections and the extrapolation of the luminosity function from the observable bright end to galaxies fainter than the magnitude limit both introdude large uncertainties.

We show the star formation rate as a function of redshift and the comparison to data in Fig. 1.5 (middle-left). We show the integrated global stellar mass as a function of redshift in Fig. 1.6 (bottom-left).

1.5.8 A quick overview of recent Durham semi-analytic models, and selected other models

All of the current semi-analytic galaxy formation models being used at Durham are based on the work of Cole et al. (2000). Cole et al. (2000) obtained a very good match to the $z = 0$ galaxy luminosity functions, using a value of $\Omega_b = 0.02$. However, subsequent to this work, the observational estimates of the baryon density of the Universe (found using nucleosynthesis arguments) approximately doubled, to $\Omega_b = 0.04$. This caused all simple variants of the Cole et al. (2000) model to overproduce significantly the bright end of the luminosity function. Therefore, Benson et al. (2003) made a detailed study of additional feedback mechanisms, in an effort to eliminate this problem. The three additional feedback mechanisms which they tested were: (1) Superwind feedback – during star formation, cold gas is ejected mechanically from the halo. (2) Injection of heat into the hot gas from the halo, which will delay its cooling. (3) Conduction of heat from the inner parts of haloes to the outer parts of haloes.

Both the Baugh et al. (2005) and Bower et al. (2006) galaxy formation models are derived from the work of Benson et al. (2003). Baugh et al. took the version of the Benson et al. (2003) model which used superwind feedback (and made a few small simplifications). They tried to match observed number densities of sub-mm galaxies and UV galaxies, and found that to do so, they had to make three additional changes: (1) A constant star formation timescale in quiescent star formation (rather than a timescale which is proportional to the dynamical time). This means that there is more gas available at high redshift to fuel bursts. (2) A top-heavy IMF in starbursts – this increases the luminosity of bursts. (3) Bursts are triggered by minor mergers as well as major mergers.

The Bower et al. (2006) model is derived from the version of the Benson et al. (2003) model which has a feedback mode where the extra feedback energy is injected into the hot gas of the halo. The following are the main changes with respect to the Benson et al. (2003) model: (1) The cooling model is changed – we distinguish between cold accretion (when the cooling time is shorter than the freefall time) and cooling from a quasihydrostatic hot halo (when the cooling time is longer than the freefall time). (2) We track black hole growth, and add an extra mode of feedback – when cooling is in the quasi-hydrostatic regime, and the energy available from the black hole jet is sufficient to offset the cooling, then cooling is switched off. (3) We use N-body merger trees. (4) We allow both bulges and black holes to grow via disc instabilities.

We now briefly consider two examples of non-Durham models. The Croton et al. (2006) semi-analytic galaxy formation model is similar to the Bower et al. model in many respects. Both use merger trees made from the same N-body simulation (albeit using slightly different algorithms). The Croton et al. galaxy formation model uses the Munich semi-analytic galaxy formation code, which was written completely seperately from the Durham semi-analytic galaxy formation code. Therefore the computational algorithms are likely to be implemented rather differently, although they are based on very similar physical principles. The Bower et al. model allows both bulges and black holes to grow during disc instabilities, whereas the Croton et al. model only allows bulges to grow during disc instabilities. Both models distinguish between cold accretion and cooling from a quasihydrostatic hot halo, and both only allow black hole feedback to be associated with the quasihydrostatic regime. However, the black hole feedback is implemented slightly differently. Bower et al. calculate the power imput required to offset gas cooling, and compare this to the power available from the black hole (parameterized as a fraction of the Eddington luminosity). Croton et al. assume that the power input from the AGN scales with the black hole mass, and with the halo circular velocity cubed.

The Monaco et al. (2007) model was written completely seperately from the Durham and Munich models. It is based on quasi-Nbody merger trees made using the 'Pinnochio' dark matter code. They compute the linear density field numerically, and they assume that regions collapse into haloes after orbit crossing occurs. Thus large scale (non-collapsed) fluctuations are treated in an Nbody fashion, but the internal dynamics of dark matter haloes are treated in a semi-analytic manner. They use this code to make merger trees. Much of the physics is similar to that in the Durham and Munich models, but they implement a more detailed model of the multiphase interstellar medium – they track the thermal energy of the hot phase and the kinetic energy of the cold phase. While Bower et al. and Croton et al. only include feedback from radio jets, Monaco et al. include two forms of feedback from AGN – radio jets and quasar triggered winds.

1.6 Black holes

In general relativity, we predict that if matter becomes extremely dense, it will collapse to a singularity in space-time – a 'black hole'. The escape velocity for a black hole is the speed of light. Thus nothing can escape from a black hole, not even light.

1.6.1 Accretion onto black holes: some basic definitions

We now present a number of definitions related to black holes which will become useful later on in the Thesis. We take c to be the speed of light, G to be the gravitational constant, m_p to be the mass of the proton, $\sigma_{\rm T}$ to be the Thomson cross section of the electron, r to be the radial distance from the black hole and $M_{\rm BH}$ to be the mass of the black hole.

The accretion efficiency, ϵ

When matter is completely converted into energy, the resulting energy, E, is given by $E = Mc^2$. When matter accretes onto a black hole, some fraction will be added, as mass, to the mass of the black hole, and some fraction will be converted into electromagnetic radiation. The accretion efficiency, ϵ, is the fraction of accreted mass-energy which is radiated as light. Therefore, $L = \epsilon \dot{M} c^2$.

The Eddington luminosity, $L_{\rm Edd}$

We assume that ionized gas (75% hydrogen and 25% helium by mass) is being accreted onto the black hole. Due to the strong electromagnetic force between them, the electrons and ions cannot separate in bulk. The plasma has a gravitatonal force towards the black hole, mainly due the protons (and helium ions), which have the far greater mass than the electrons. The force per electron is $F_{\rm gravity} = GM_{\rm BH}\mu_e/r^2$ ($\mu_e = 1.143 m_p$ is the mass per electron, which takes into account the presence of 25 % helium by mass – if the accreting gas were purely hydrogen then the force would be given by $F_{\rm gravity} = GM_{\rm BH}m_p/r^2$). If the accretion disc is radiating, then the photons leaving it interact with the electrons, causing a force per electron away from the black hole, $F_{\rm radiation} = \sigma_{\rm T} L/(4\pi c r^2)$. The greater the luminosity being emitted, the greater is this force. If the luminosity increases, eventually the radiation force will become equal to the gravitational force which is attracting matter towards the black hole. Equating the two forces, $|F_{\rm gravity}| = |F_{\rm radiation}|$, gives the maximum luminosity, the 'Eddington luminosity' ($L_{\rm Edd}$), at which the black hole can radiate without terminating its own fuel supply. We find that:

$$L_{\rm Edd} = (4\pi c G \mu_e/\sigma_{\rm T}) \times M_{\rm BH} \quad .$$

Note that since the gravitational force and the radiative force vary as r^{-2}, the Eddington luminosity is independent of the distance from the black of the radiating, accreting matter.

The Eddington accretion rate, $\dot{M}_{\rm Edd}$

If we know that a black hole is radiating at its Eddington luminosity, then, from the accretion efficiency, ϵ, and the mass of the black hole, the accretion rate is fully determined. The accretion rate required to produce the Eddington luminosity is the Eddington accretion rate, $\dot{M}_{\rm Edd}$. Combining the equations for the above two definitions, we find that:

$$\dot{M}_{\rm Edd} = L_{\rm Edd}/(c^2\epsilon) = (4\pi G\mu_e/\sigma_{\rm T}c\epsilon) \times M_{\rm BH} \ .$$

The Salpeter timescale, $t_{\rm Salpeter}$

If a black hole grows at the Eddington rate, with a constant efficiency of photon emission, then, inserting the appropriate values for physical constants in the above equation for $\dot{M}_{\rm Edd}$, its rate of growth is directly proportional to its mass, i.e.: $\dot{M}_{\rm BH,Edd} = (4 \times 10^7\,{\rm yr})(0.1/\epsilon){\rm M_{BH}}$. Therefore, the black hole will grow in mass exponentially on a timescale called the Salpeter timescale – $t_{\rm Salpeter} = 4 \times 10^7(0.1/\epsilon)$yr (Salpeter 1964).

1.6.2 Are there really black holes in the Universe? The evidence for supermassive black holes in galactic centres

There is strong evidence that black holes really do exist in the Universe. In particular, when we consider the evidence for supermassive ($M_{\rm BH} \gtrsim 10^5 M_\odot$) black holes in the centres of galaxies, it appears that every galaxy with a bulge has a supermassive black hole at its centre (Kormendy & Richstone 1995). We are able to infer the mass profile of a galaxy from the dynamics of stars in the galaxy, from gas dynamics, or from maser dynamics (Kormendy & Gebhardt 2001). It is possible to infer the stellar mass from the galaxy luminosity and stellar spectral synthesis models. We find that, in the centre of the galaxy, there is a mass concentrated in a small region of space which is not luminous. Strictly speaking, for most galaxies, the evidence only points to there being a 'massive dark object'. However, for the Milky Way (and also for one other galaxy), the mass of the central dark object has been confined to a very small region from high accuracy observations of the dynamics of individual stars at the centre of the galaxy (Schödel et al. 2003; Ghez et al. 2005). The inferred density is so high that it can only be explained as a black hole, and nearly every other possible explanation (e.g. a dense cluster of dead stars, a fermion star) has been ruled out.

The discovery of quasars in the Universe (Schmidt 1963) was one of the first pieces of evidence for the existence of supermassive black holes in galactic centres, and is still

a powerful argument. Quasars are seen as extremely bright quasi-stellar objects in the centre of a galaxy, and often outshine the galaxy. Such a large output of energy from such a small region of space is very suggestive that quasars are fuelled by accretion onto a black hole at the centre (Lynden-Bell 1969). We assume that the massive dark objects in other galaxies are, in general, black holes, and that quasars are fuelled by accretion onto black holes.

1.6.3 The role of black holes in galaxy formation

We now summarize the relationships between black holes, quasars, bulges, starbursts and feedback which motivate the rest of this work. We discuss these relationships more thoroughly in §2.1, §6.1 and §6.2.

• There are strong relationships between the mass of the black hole in the centre of a galaxy and various galaxy properties, i.e. the bulge mass, the bulge magnitude and the bulge velocity dispersion (Kormendy & Richstone 1995; Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2000; Kormendy & Gebhardt 2001; Marconi & Hunt 2003; Häring & Rix 2004). This suggests that there may be an intimate link between black hole formation and galaxy formation.

• It is likely that a primary formation mechanism for the formation of galactic bulges is the merger of two disc galaxies. The reasonably ordered orbits in galactic discs become more random and thermalized due to the highly asymmetric and time-dependent gravitational potential well during the merger (Barnes & Hernquist 1991; Mihos & Hernquist 1994a,b).

• The global luminosity emitted from quasars peaks at similar redshifts to the global star formation rate (e.g. Boyle & Terlevich 1998). This suggests a link between black hole accretion and star formation.

• Galaxies with high star formation rates or which are undergoing starbursts (e.g. bright galaxies selected in the infrared or sub-mm, or via the Lyman break tequnique) often show some evidence of accretion, particularly when observed in the X-rays (e.g. Alexander et al. 2003, 2005a).

• The merger of gas-rich galaxies causes a loss of symmetry, and gas is able to lose angular momentum. There are consequently large gas flows to the centres of the merging galaxies and the final merger remnant. This gas is likely to undergo very rapid star formation (a 'starburst'), and also to fuel any black hole which may be in the galaxy centre (e.g. Di Matteo et al. 2005; Springel et al. 2005a).

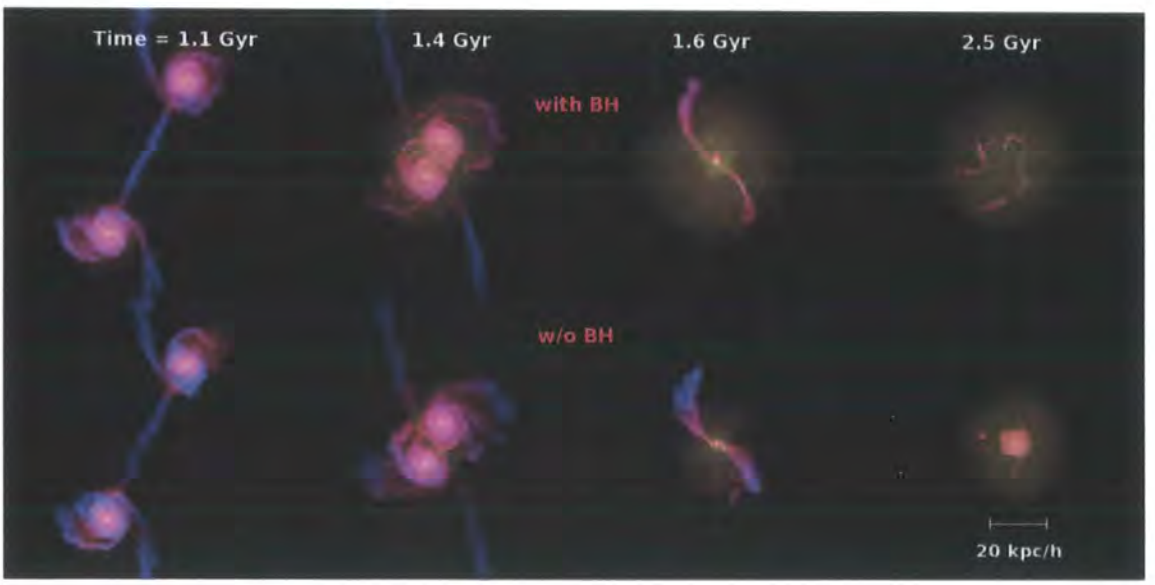

Figure 1.7: A galaxy merger both with (top) and without (bottom) a black hole. The images visualise the projected gas distribution in the two galaxies, with colours representing temperature – blue (cold) → red (hot). The first snapshot ($t = 1.1$ Gyr) shows the systems after the first passage of the two galaxies. The second snapshot ($t = 1.4$ Gyr) depicts the galaxies distorted by their mutual tidal interaction, just before they merge. The peak in the star formation and black hole accretion is reached at the time of the third snapshot ($t = 1.6$ Gyr), when the galaxies finally coalesce. At this time, a strong wind driven by feedback energy from the accretion expels much of the gas from the inner regions in the simulation with black holes. Finally, the last snapshots show the systems after the galaxies have merged ($t = 2.5$ Gyr), leaving behind quasi-static spheroidal galaxies. In the simulation with black holes, the remnant is very gas poor and has little gas left dense enough to support ongoing star formation. This plot and caption were taken from Di Matteo, Springel & Hernquist (2005).

• It is difficult to match the galaxy luminosity function and to prevent overcooling in clusters and groups with only stellar feedback (e.g. Benson et al. 2003a). Feedback from accretion onto a central supermassive black hole could solve this problem.

We now illustrate some of these points. In Fig. 1.7, we show a numerical simulation by Di Matteo et al. (2005) which illustrates a merger-triggered starburst. The simulation begins with two disc galaxies. As the galaxies approach, gas in each loses angular momentum and falls to the centres, rapidly undergoing star formation. When the galaxies merge, gas falls into the centre of the merger remnant, and undergoes a second burst of star formation. The simulation is done twice, both without and with black holes in the centres of the two galaxies. Without a black hole, a fair amount of the gas is blown out due to supernova feedback, and the final merger remnant is a fairly compact bulge. When black holes are included, the black holes accrete gas which has been driven to the centres of the two progenitor galaxies, and in a second burst of activity, from gas which has been driven to the centre of the merger remnant. The black holes drive much stronger feedback than was possible due to supernova feedback alone, heating up and blowing out nearly all the cold gas in the galaxy which does not manage to form stars.

In summary, there is very strong motivation for including black holes in our galaxy formation model. In Chapter 2, we describe how we do this, and demonstrate that we are able to match a set of the most important data relating galaxies to black holes. In Chapter 3 we describe how black holes grow in the model, and in Chapter 4, we demonstrate that we are able to reproduce the observations which indicate that black hole accretion 'downsized' to lower mass objects at lower redshift. In Chapter 5, we explore the implications of our model for quasars at $z \sim 6$. We introduce an alternative model which includes black hole feedback in Chapter 6, which we compare it to our fiducial model. We summarize, briefly outline further directions for our research and conclude in Chapter 7.

Chapter 2 | *The model*

In this Chapter we first discuss the motivation for including black holes in galaxy formation (§2.1.1 and §2.1.2), and in particular, the association between black hole growth and bursty star formation (§2.1.3). We will briefly mention previous theoretical approaches to the link between black holes and galaxy formation (§2.1.4) and overview the approach we plan to take (§2.1.5).

We will describe our model for galaxy formation and black hole growth in detail in §2.2. We will begin with an overview of the Baugh et al. (2005) galaxy formation model upon which it is based (§2.2.1), paying particular attention to the modelling of bulge formation and starbursts during galaxy mergers (§2.2.2). We explain how the model has been extended to follow the formation of black holes in §2.2.3, and discuss the sensitivity of our model predictions to the mass resolution of the dark matter merger trees in §2.2.4.

We set the main model parameters in §2.3 by comparison to observations. Our black hole model has one free parameter, $f_{\rm BH}$, which determines the mass accreted onto the black hole during a starburst (see §2.2.3). In §2.3.1, we set $f_{\rm BH}$ by requiring that the model should reproduce the local observed relationship between black hole mass ($M_{\rm BH}$) and the stellar mass of the bulge ($M_{\rm bulge}$) in which it resides. We also show the model predictions for how black hole mass scales with other properties of the bulge. Any viable model of black hole growth should also be consistent with the observed quasar population. In §2.3.2, we briefly describe how a quasar luminosity can be assigned to accreting black holes, and present some illustrative results for the quasar luminosity function at selected redshifts.

2.1 Motivation

2.1.1 The association of quasars with luminous, dusty starbursts

In the local Universe, luminous, dusty, merger-driven starburst activity has long been suspected to have quasar activity associated with it (Sanders & Mirabel 1996). Some authors find that the most powerful Seyfert II active galactic nuclei are usually found in

galaxies which have had a starburst in the past 1-2 Gyr and use this observation to argue that the brightest quasars are associated with galaxy mergers (e.g. Kauffmann et al. 2003), whilst others claim that the brightest quasars are hosted in elliptical galaxies which are indistinguishable from the general elliptical population (e.g. Dunlop et al. 2003). At high redshift, sources detected in the submillimeter are thought to be starbursts (Chapman et al. 2004), many are associated with galaxy mergers (Swinbank et al. 2004) and many show evidence of active nuclei when probed deeply in the X-rays, although it appears that the AGN makes a much smaller contribution to the powerful submm flux than does the starburst (Alexander et al. 2003).

2.1.2 Connections between black holes and their host galaxies

Black holes (BH) display strong correlations with the properties of their host galaxy, particularly those of the galactic bulge (Kormendy & Richstone 1995; Magorrian et al. 1998). Black hole mass is observed to scale with the bulge's B-band luminosity (Magorrian et al. 1998; Kormendy & Gebhardt 2001), K-band luminosity (Marconi & Hunt 2003; Häring & Rix 2004), stellar mass (Marconi & Hunt 2003; Häring & Rix 2004) and velocity dispersion (Ferrarese & Merritt 2000; Gebhardt et al. 2000). We refer to these collectively as the '$M_{\rm BH}-$bulge' relations. It has long been theorized that galactic bulges form through galaxy mergers (Toomre & Toomre 1972), so it is natural to speculate that these events also drive the strong correlation between the properties of the bulge and the mass of the black hole.

2.1.3 Links between black hole accretion and starbursts

There is strong evidence for a link between galactic star formation and accretion onto central black holes. The evolution with redshift of the global star formation rate and the luminosity density of optical quasars are strongly correlated (Boyle & Terlevich 1998). At low redshift, the ratio of the global star formation rate to the global black hole accretion rate for bulge dominated galaxies, $SFR/\dot{M}_{\rm BH}$ is $\sim$ 1000, which is remarkably similar to the ratio of $M_{\rm BH}/M_{\rm bulge}$ (Heckman et al. 2004). However, it is still an open question whether black hole growth is correlated with all star formation equivalently, or whether its strongest relationship is with star formation in bursts.

The physical conditions in mergers and starbursts are amenable to fuelling the accretion of material onto a central supermassive black hole. Numerical simulations of galaxy mergers have shown that the asymmetrical gravitational potential present during the

merger is responsible for driving gas to the centres of the merging galaxies and of their remnant in both major mergers (Mihos & Hernquist 1994b) and minor mergers (Mihos & Hernquist 1994a). The enhanced supply of gas to the centre of the galaxy leads to rapid star formation and is also available to fuel an AGN (Norman & Scoville 1988; Di Matteo et al. 2005; Springel et al. 2005a). Furthermore, the formation of a dense stellar system with a steep $R^{1/4}$-law potential well during a gas-rich merger may help to funnel gas to the AGN at the very centre. Starbursts appear to be required in the high redshift Universe to explain observations of various galaxy populations such as Lyman break galaxies and sub-mm galaxies (Somerville et al. 2001; Baugh et al. 2005). The increased prevalence of starbursts at early epochs may be responsible for the accelerated growth of the most massive black holes towards high redshift (e.g. Granato et al. 2004, 2006).

2.1.4 Previous theoretical work

Theoretical calculations of the growth of black holes in the Cold Dark Matter (CDM) cosmology in which structures grow through gravitational instability have tended to fall into one of three classes: (i) calculations based on the rate at which dark matter haloes are assembled, either without any treatment of galaxy formation (e.g. Efstathiou & Rees 1988; Haehnelt & Rees 1993; Haiman & Loeb 1998; Percival & Miller 1999; Wyithe & Loeb 2003; Haiman 2004; Koushiappas et al. 2004; Yoo & Miralda-Escudé 2004; Mahmood et al. 2005) or with very simple estimates of the supply of gas accreted onto the BH (Volonteri et al. 2003; Islam et al. 2003; Bromley et al. 2004; Libeskind et al. 2006); (ii) numerical simulations of galaxy mergers, which use a mixture of smooth particle hydrodynamics and simple recipes to follow the fuelling of a supermassive black hole (Cattaneo et al. 2005; Di Matteo et al. 2005; Hopkins et al. 2005a; Springel et al. 2005a; Robertson et al. 2006; Hopkins et al. 2006); (iii) semi-analytical modelling of the formation of galaxies and black holes (Kauffmann & Haehnelt 2000; Cattaneo 2001; Enoki et al. 2003; Menci et al. 2004; Granato et al. 2004; Cattaneo et al. 2006a; Kang et al. 2006; Monaco et al. 2007). Recently, the semi-analytical approach has been extended to produce models in which the evolution of galaxies and black holes are coupled, with energy released by accretion onto the black hole either truncating ongoing star formation or suppressing the rate at which gas can cool in more massive haloes (Granato et al. 2004; Monaco & Fontanot 2005; Bower et al. 2006; Croton et al. 2006).

2.1.5 Our theoretical approach and our objective

In this Thesis, we incorporate a model for the growth of black holes into the Durham semi-analytical galaxy formation code GALFORM (Cole et al. 2000; Benson et al. 2003a). Our prescription for growing black holes is tied to galaxy mergers and is similar to the first implementation of black hole growth in semi-analytical models by Kauffmann & Haehnelt (2000). Our starting point is the galaxy formation model introduced by Baugh et al. (2005). This was the first model to match the observed properties of galaxies in both the low and high redshift Universe, following the whole of the galaxy population and incorporating a self-consistent calculation of the reprocessing of starlight by dust. In particular, the model reproduced the number counts of Lyman break galaxies and sub-mm sources, which are both dominated by starbursts. The success of the model was primarily due to an increased level of star formation in bursts at high redshift compared with previous models, and the adoption of a flat initial mass function (IMF) for stars produced in starbursts. The same model also accounts for the metal content of the hot gas in clusters and stars in ellipticals (Nagashima et al. 2005a,b) and the numbers of Lyman-alpha emitters (Le Delliou et al. 2005, 2006).

The use of a semi-analytical model allows us to follow a much wider population of objects than is accessible by direct numerical simulation. This means that we can follow the demographics of the black hole population and explore how black holes acquire their mass. The latter is of great importance in view of the recent observational evidence suggesting that the most massive black holes acquired the bulk of their mass at early epochs and that it is the lower mass black holes which are being built up most rapidly today. This phenomenon has been termed 'downsizing' (Cowie et al. 2003; Steffen et al. 2003; Ueda et al. 2003; Barger et al. 2005; Hasinger et al. 2005). At first sight, downsizing appears to imply that the growth of black hole mass is 'anti-hierarchical' and thus incompatible with the CDM cosmological framework (Marconi et al. 2004; Merloni 2004; Shankar et al. 2004). We will examine here whether or not such downsizing is really a problem for hierarchical models of galaxy formation.

2.2 Description of the model

2.2.1 The semi-analytical galaxy formation model

Our starting point is the model for galaxy formation in the CDM cosmology described by Baugh et al. (2005). As we have already pointed out in the §2.1.5, in addition to giving a reasonable match to the properties of galaxies in the local Universe, this model also reproduces the counts of sub-millimetre sources and the luminosity function of Lyman-break galaxies at high redshift. In both cases, the model associates these high redshift objects with galaxies which are undergoing merger-driven starbursts. The success of the Baugh et al. (2005) model in reproducing observations linked with vigorous starbursts and the formation of spheroids is important for the current analysis. Here we will follow the proposal of Kauffmann & Haehnelt (2000) and assume that black hole growth is driven by galaxy mergers. For an exhaustive description of the physics and methodology behind the semi-analytical model, we refer the reader to Cole et al. (2000) and Benson et al. (2003). A gentler introduction to hierarchical galaxy formation may be found in Baugh (2006).

We will review the aspects of the model which control the outcome of galaxy mergers in the next subsection, and will limit ourselves here to more general aspects of the cosmological and galaxy formation models. We assume a standard ΛCDM cosmology, with a flat geometry, a matter density $\Omega_0 = 0.3$, a baryon density $\Omega_b = 0.04$, a Hubble constant of $H_0 = 70\,\mathrm{km\,s^{-1}Mpc^{-1}}$ and a fluctuation amplitude specified by $\sigma_8 = 0.9$. The break in the local galaxy luminosity function is reproduced by invoking a superwind which drives cold gas out of galaxies (Benson et al. 2003a; Nagashima et al. 2005a); an alternative physical mechanism to produce this break is AGN feedback in quasi-hydrostatically cooling haloes (Bower et al. 2006; Croton et al. 2006). Gas cooling is prevented below $z = 6$ in low circular velocity haloes ($v_c = 60\,\mathrm{km\,s^{-1}}$), to mimic the impact of the presence of a photoionizing background on the intergalactic medium (Benson et al. 2002). Baugh et al. (2005) adopt a timescale for quiescent star formation in galactic discs which is independent of the dynamical time, which results in gas rich mergers at high redshift (see their fig. 1). They assume that stars which form in bursts are produced with a top-heavy initial mass function; this choice has no impact on the predictions presented in this Thesis, except that a large fraction of the cold gas forming stars is recycled into the IGM.

The parameters of the Baugh et al. (2005) galaxy formation model are held fixed in Chapters 2, 3 and 4 of this Thesis; we do not adjust these parameters in any way

when generating predictions for black holes and quasars. This is a clear strength of our approach and choice of galaxy formation model. Thus, our results are to be viewed as genuine predictions of the model. The properties of the quasars and active nuclei in our model can easily be related to the properties of their host galaxies; such comparisons are deferred to future work.

2.2.2 Galaxy mergers

Mergers between galaxies play an important role in building up the mass and determining the morphology of galaxies. When dark matter haloes merge in our model, the galaxies they contain are ranked in mass. The most massive one is designated as the 'central' or 'primary' galaxy in the new dark halo and the remaining galaxies become its satellites. The satellites lose any hot gas reservoir that they may have had prior to the merger and any subsequent accretion of cooling gas is funnelled into the central galaxy. The orbits of the satellite galaxies decay through dynamical friction. If the timescale for a satellite to sink to the centre of the halo is shorter than the lifetime of the halo, then the satellite is merged with the central galaxy at the appropriate time (see e.g. Cole et al. 2000).

The result of a galaxy merger is determined by two principal quantities:
(i) $f_{\rm merge}$, the ratio of the mass of the accreted satellite to the mass of the primary, where:

$$f_{\rm merge} = M_{\rm smaller}/M_{\rm larger} \ . \tag{2.1}$$

(ii) $f_{\rm gas}$, the fraction of the mass of the primary disc which is cold gas, where:

$$f_{\rm gas} = M_{\rm cold,primary}/M_{\rm disc,primary} \ , \tag{2.2}$$

and $$M_{\rm disc} = M_{\rm stars} + M_{\rm cold} \ .$$

If the mass ratio, $f_{\rm merge}$, exceeds a threshold $f_{\rm ellip}$, the merger is termed 'violent' or 'major'. In this case, all stars present are rearranged into a spheroid, with a radius determined by arguments based on the conservation of energy and the virial theorem (see Cole et al. 2000; Almeida et al. 2006). In addition, any cold gas in the merging galaxies is assumed to undergo a star formation burst and the stars thus produced are added to the new spheroid. Here, following Baugh et al. (2005), we set $f_{\rm ellip} = 0.3$.

When $f_{\rm merge} < f_{\rm ellip}$, the merger is termed 'minor'. In this case, the stars in the accreted satellite are added to the spheroid of the primary, leaving intact any stellar disc present in the primary. In minor mergers, the fate of the gas in the merging galaxies depends upon the gas fraction in the primary disc. If the primary disc is gas rich (in practice

if $f_{\rm gas} > f_{\rm gas,burst}$, where we take $f_{\rm gas,burst} = 0.75$), we assume that the perturbation introduced by the merging satellite is sufficient to drive all the cold gas, from both the primary and the satellite, into the spheroid, where it takes part in the burst. Otherwise, if in a minor merger the gas fraction in the primary disc is small, no burst occurs. Furthermore, if the secondary galaxy is very much less massive than the primary (i.e. if $f_{\rm merge} < f_{\rm burst}$, where we take $f_{\rm burst} = 0.05$) then the primary disc remains unchanged, the accreted stars are added to the spheroid and there is no burst, irrespective of $f_{\rm gas}$. These refinements of the Cole et al. (2000) model relating to minor mergers were described by Baugh et al. (2005).

2.2.3 The growth of black holes in galaxy centres

The observed correlation between the inferred mass of galactic central black holes and the properties of their host spheroids suggests a common origin for these two classes of object (e.g. Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2000).

We assume that any contribution to the black hole mass from processes other than galaxy mergers (e.g. the end products of population III stars, primordial black holes or accretion onto black holes from galactic discs or from the hot gas within a halo) is small compared with the increase in mass during merger-driven starbursts. The physics of black hole seeding is very uncertain, and many mechanisms of black hole seeding have been proposed, with widely varying associated seed masses (see Volonteri 2006 and references therein). The largest mass of seed black holes suggested by models in the current literature is $\sim 10^5 - 10^6 M_\odot$ (e.g. Bromm & Loeb 2003; Volonteri & Rees 2005a; Lodato & Natarajan 2006). These models typically apply only to metal-free and highly biased regions at high redshift and probably therefore only to the seeding of the the most massive black holes (they are motivated by the difficulty in producing $\sim 10^9 h^{-1} M_\odot$ black holes at $z \sim 6$ from smaller seeds, since at high redshifts the age of the Universe is not long compared to the Salpeter time). Other models for seed black holes predict seeds which are less massive than $10^5 h^{-1} M_\odot$, which is our black hole mass resolution limit. Therefore, it is reasonable for us to neglect the role of seed black holes in our calculations. In practice, we assume that if a pre-existing seed is indeed required for supermassive black hole formation, then the mass is small enough that it only makes a negligible contribution to the mass of the final black hole. Futhermore, theoretical considerations (e.g. Begelman 1978, 2002; King 2002) and observations (e.g. Collin & Kawaguchi 2004) suggest that super-Eddington accretion of mass is possible, and we assume that this occurs during the

early stages of black hole growth, so that the mass of any seed does not affect the final mass by Eddington limiting of mass accretion.

The mass of black holes is assumed to grow during galaxy mergers via two channels: accretion of gas during merger-driven starbursts and mergers with other black holes. (Note that in the recent model by Bower et al. 2006, additional modes of black hole growth are considered: accretion during starbursts triggered by disc instabilities, and accretion of cooling gas from quasi-hydrostatically cooling haloes.) As discussed in §2.2, we allow starbursts, and thus accretion, in both major and minor mergers. (In contrast, Kauffmann & Haehnelt 2000 only allowed starbursts and black hole accretion during major galaxy mergers.) The two channels for black hole growth are as follows:

Firstly, in starbursts triggered by galaxy mergers, we assume that a fraction, $f_{\rm BH}$, of the gas mass which is turned into stars is accreted onto the black hole. The mass of stars produced in the burst is the mass *after* taking into account feedback processes that may expel gas from the galaxy and the recycling of mass from stars. Typically, we use $f_{\rm BH} < 0.03$ (this is explained in §2.3.1), and so for simplicity we ignore the depletion of the cold gas reservoir by black hole growth when we calculate star formation. We assume that the growth of black hole mass is not limited to the Eddington accretion rate appropriate to our chosen radiative efficiency.

Secondly, if the merging galaxies already host black holes, then we assume that these black holes merge when the host galaxies merge. In reality, black holes do not merge instantaneously, but gas-dynamical processes are likely to speed-up black hole coalescence in gas-rich mergers (Armitage & Natarajan 2002) and circumstantial observational evidence exists to suggest that most binary black holes do merge efficiently, even in gas-poor mergers (Merritt & Milosavljević 2005). Since we only consider binary galaxy mergers with instantanous central black hole merging, all BH-BH mergers in our model are binary, and we ignore slingshot ejection of black holes from the galactic centre (Saslaw et al. 1974). We also ignore the recoil of the merger products of unequal mass black holes due to the anisotropic emission of gravitational waves, which may lead to the resultant black hole being ejected from the galaxy nucleus (Fitchett 1983; Volonteri et al. 2003; Libeskind et al. 2006). Older calculations suggest that most recoil velocities are likely to be in the range $10 - 100\,{\rm km\,s^{-1}}$ (Favata et al. 2004), and thus unimportant except in very low mass galaxies.

However, following great progress in numerical relativity, estimates of typical kick velocities have been significantly updated. The new estimates are significantly higher,

suggesting that we may need to include a better treatment in future work. Baker et al. (2006) find that for a mass ratio $M_1/M_2 = 0.67$, the kick velocity after the merger of two non-spinning black holes is $101 \pm 15\,\mathrm{km\,s^{-1}}$. Gonzalez et al. (2006) studied mergers of non-spinning black holes over a range of mass ratios $M_1/M_2 = 0.25 - 1.0$. They calculate the magnitude of the kick velocities to an accuracy of 2%, and found that they ranged from $0 - 175\,\mathrm{km\,s^{-1}}$ depending on the mass ratio, and that the maximum kick velocity was $175.2 \pm 11\,\mathrm{km\,s^{-1}}$ for a mass ratio $M_1/M_2 = 0.36 \pm 0.03$. Their estimated kick velocities were above $150\,\mathrm{kms}^{-1}$ for mass ratios in the range $M_1/M_2 = 0.25 - 0.5$. Their results are in agreement with those of Baker et al. (2006).

Even more recently, a number of authors (Baker et al. 2007; Campanelli et al. 2007a,b; Gonzalez et al. 2007; Koppitz et al. 2007; Herrmann et al. 2007) have studied kicks from mergers of spinning black holes. When one or both initial black holes are spinning, this initial spin is often the dominant component to the final kick velocity. Firstly and more simply (due to the symmetry about the orbital plane – there can be no component of the kick velocity perpendicular to the orbital plane), this has been considered where any initial black hole spins are orthogonal to the orbital plane (Baker et al. 2007; Koppitz et al. 2007; Herrmann et al. 2007). Baker et al. (2007) considerd initial BH of mass ratio $M_1/M_2 = 0.66$, each with an initial dimensionless spin parameter of $a = 0.2$, with the initial spins aligned or anti-aligned with the orbital angular momentum. They obtained kick velocities ranging from $\sim 60\,\mathrm{km\,s^{-1}}$ (when the spin of the larger BH is aligned and that of the smaller BH anti-aligned with the orbital angular momentum) to $\sim 190\,\mathrm{km\,s^{-1}}$ (when the spin of the larger BH is anti-aligned and that of the smaller BH aligned with the orbital angular momentum). Koppitz et al. (2007) considered an equal mass binary with initial spins equal, antialigned, orthogonal to the orbital plane, and with a dimensionless spin parameter of $a = 0.58$. They obtained a kick velocity of $257 \pm 15\,\mathrm{km\,s^{-1}}$. They extrapolate to a kick velocity of $\sim 440\mathrm{km\,s^{-1}}$ for maximally rotating ($a = 1$) initial BH. Herrmann et al. (2007) considered two initial BH of equal mass with equal, oppositely directed spins which are aligned/anti-aligned with the orbital angular momentum. They varied the dimensionless spin parameter a between 0 and 1. For zero spin, there is no kick. They found that the kick velocity appears to increase in proportion to the dimensionless Kerr spin parameter a, with a value of $392 \pm 33\,\mathrm{km\,s^{-1}}$ for $a = 0.8$.

Secondly, some authors have considered initial BH configurations where one or both initial BH have a component of spin in the orbital plane. A component of the kick velocity orthogonal to the orbital plane of the black holes is allowed when one or both black holes

initially have a spin with a component orthogonal to their orbital plane. This component of the kick velocity orthogonal to the orbital plane may often be far more significant than that in the orbital plane, depending upon the initial spin configuration (Campanelli et al. 2007a,b; Gonzalez et al. 2007). Campanelli et al. (2007a) study a non-spinning BH orbiting a spinning BH of twice the mass with the spin oriented at -45^o relative to the orbital plane. They find a kick velocity of $454\,\mathrm{km\,s^{-1}}$. They suggest that the most favorable configuration for producing a maximal kick velocity would be an equal mass binary with initial spins anti-aligned in the orbital plane. They also study two BH with initial spins (both of $a = 0.5$) anti-aligned and in the orbital plane, finding a kick velocity of $\sim 1830\,\mathrm{km\,s^{-1}}$, and suggest that the kick velocity could scale up to $\sim 4000\,\mathrm{km\,s^{-1}}$ if the initial spins were maximal ($a = 1$). Campanelli et al. (2007b) simulate equal mass initial BH, with initial spins equal, anti-aligned, and in the orbital plane. They find that the recoil velocity is perpendicular to the orbital plane, and varies sinusoidally with the angle between the initial linear angular momenta and initial spin of the initial BH, up to a maximum of $\sim 4000\,\mathrm{km\,s^{-1}}$ for maximally rotating initial BH. Gonzalez et al. (2007) simulate the merger of two equal mass BH with equal anti-aligned spins (both with dimensionless spin parameter $a \sim 0.8$) in the orbital plane. They obtain a kick velocity of $2500 \pm 250\,\mathrm{km\,s^{-1}}$.

In summary, while the kick velocity due simply to the mass-asymmetry between initial merging BH may be important, the contribution from their spins may be even more significant, and the contribution due to the component of their spins perpendicular to their orbital angular momentum may be even more significant still, depending upon the geometry. We must properly consider these effects in future work.

Note that we neglect any loss of mass arising from the radiation of gravitational waves during the merger of two black holes. Such radiation could result in the mass of the merger product being less than the sum of the masses of the black holes from which it formed (Yu & Tremaine 2002). This effect is very uncertain, but is maximal for equal mass black holes, and even then it is likely to be small – approximately 3% or less of the initial mass energy for equal-mass non-spinning or Kerr black holes (Baker et al. 2002, 2004). Since most BH–BH mergers in the Universe have unequal mass ratios, the cumulative mass loss by gravitational radiation is unlikely to be more than the figure of 20% predicted using the most extreme models for gravitational wave loss in individual BH–BH mergers (Menou & Haiman 2004). Therefore, we assume that the final black hole mass is the sum of the mass accreted plus the mass of the two progenitors.

2.2.4 Resolution tests

The black hole mass down to which our predictions for the properties of black holes can be trusted depends upon two factors: the accuracy of our prescriptions for handling the physical ingredients of our galaxy formation model and the mass resolution of the dark matter halo merger trees. The semi-analytical galaxy formation model gives a reasonable match to the field galaxy luminosity function, including its faint end (Benson et al. 2003a). Further tests of the modelling of the phenomena operating in low mass systems are deferred to future work. This leaves the mass resolution of the halo merger trees as a numerical parameter that directly influences the properties of low-mass black holes.

In Chapters 2, 3, 4 and 5, we use dark matter halo merger trees generated using the Monte Carlo scheme described by Cole et al. (2000), whilst in Chapter 6, we use merger trees extracted from the 'Millennium' N-body simulation (Springel et al. 2005b). Merger trees extracted from N-body simulations are, in some respects, more accurate (e.g. Kang et al. 2005; Nagashima et al. 2005c). However, a major limitation of the trees extracted from simulations is their finite mass resolution. Unpublished work by Cedric Lacey (priv. comm.) and work in preparation by Helly et al. (John Helly, priv. comm.) show that the merger trees in the Durham semi-analytic model agree well with the merger trees in N-body simulations. Monte-Carlo generated trees can have far superior mass resolution, because the whole of the computer memory is devoted to one tree at a time, rather than to a large ensemble of haloes within a cosmological volume. This also means we are able to extend our merger trees to high redshifts (we use $z_{\rm start} = 20$). Also, Monte-Carlo trees typically have superior time resolution to those taken from N-body simulations. On the other hand, Monte Carlo trees tend to become less accurate as the time interval over which the trees are grown is increased (Somerville et al. 2000).

Putting this caveat aside, we have performed extensive tests of the impact of the mass resolution of the dark matter merger trees on our predictions for the mass function of black holes. The results of this convergence study are presented in Fig. 2.1 for $z = 0$ and $z = 6$. With improved mass resolution in the merger tree, we are able to trace more of the gas which cools in low mass haloes before reionization. This is the reason for the odd-looking 'bumps' at low BH masses in the $z = 0$ panel. Our fiducial choice of halo mass resolution is $5 \times 10^8 h^{-1} M_\odot$. This is an order of magnitude better than the resolution used in our standard galaxy formation calculations, and thirty times better than the resolution of the best N-body merger trees currently available within a cosmological

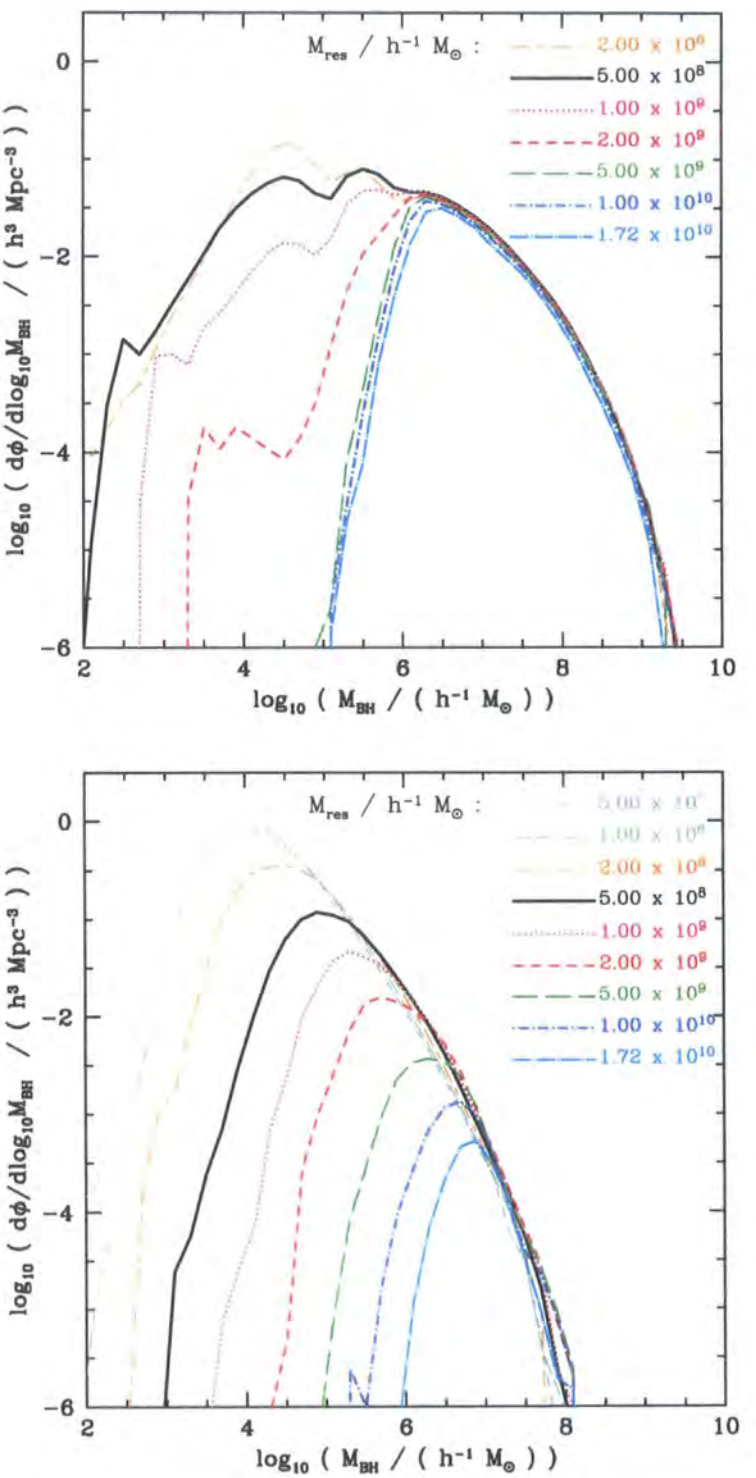

Figure 2.1: The mass function of black holes, computed using different resolutions for the merger trees of dark matter haloes, as indicated by the legend in each panel. Our fiducial resolution is shown by the thick solid line. The top panel shows the results for $z = 0$ and the bottom panel shows $z = 6$.

volume (the 'Millennium' simulation of Springel et al. 2005c which can resolve haloes of mass $1.72 \times 10^{10} h^{-1} M_{\odot}$). With our fiducial halo mass resolution, our predictions for the mass function of black holes have converged for masses of $10^5 h^{-1} M_{\odot}$ and above.

2.3 Comparisons with observational data

2.3.1 Setting the model parameter: predictions for the present day bulge–black hole relation

The main parameter of our black hole model is the fraction, $f_{\rm BH}$, of the mass of stars formed in a starburst which is accreted onto the central black hole (after taking into account gas ejected from the galaxy by feedback processes and the recycling of mass in supernova explosions and stellar winds). We fit the value of $f_{\rm BH}$ by comparing the model predictions to the observed correlation between the mass of galactic central black holes and the stellar mass of the bulge component, $M_{\rm BH} - M_{\rm bulge}$, as inferred by Häring & Rix (2004). Häring & Rix make a dynamical estimate of the stellar mass of the bulge. They compile from the literature black hole mass estimates made using a variety of techniques (stellar, gas or maser dynamics). A review of these techniques and their uncertainties can be found in Kormendy & Gebhardt (2001).

We find that a value of $f_{\rm BH} = 0.022$ is required for the model to match the zeropoint of the observed $M_{\rm BH} - M_{\rm bulge}$ relationship (Fig. 2.2a). It is important to remember that the normalization of this relationship is set by the choice of $f_{\rm BH}$. However, the slope and scatter are genuine model predictions, and as Fig. 2.2 shows, these predictions are in good agreement with the observations.

Naïvely one might argue that, since we have assumed that a fixed fraction of the mass of stars formed in a burst is added to the mass of the black hole, it is hardly surprising that a tight $M_{\rm BH} - M_{\rm bulge}$ relationship results. However, in the model, bursts play a minor role in the formation of bulge stars. The dominant channel responsible for building up the mass of present day spheroids is the re-assembly of stellar fragments (discs and bulges) during mergers, not the burst accompanying the last major merger experienced by the galaxy (Baugh et al. 1996). Baugh et al. (2005) found that around 30% of all (bulge + disc) star formation in their model takes place in the burst mode, with the rest occuring in galactic discs. Once the high recycled fraction resulting from the use of a top-heavy IMF in starbursts is taken into account, less than 10% of the total mass (bulge + disc) locked up in stars at redshift zero is produced in bursts. Thus, the slope and scatter of

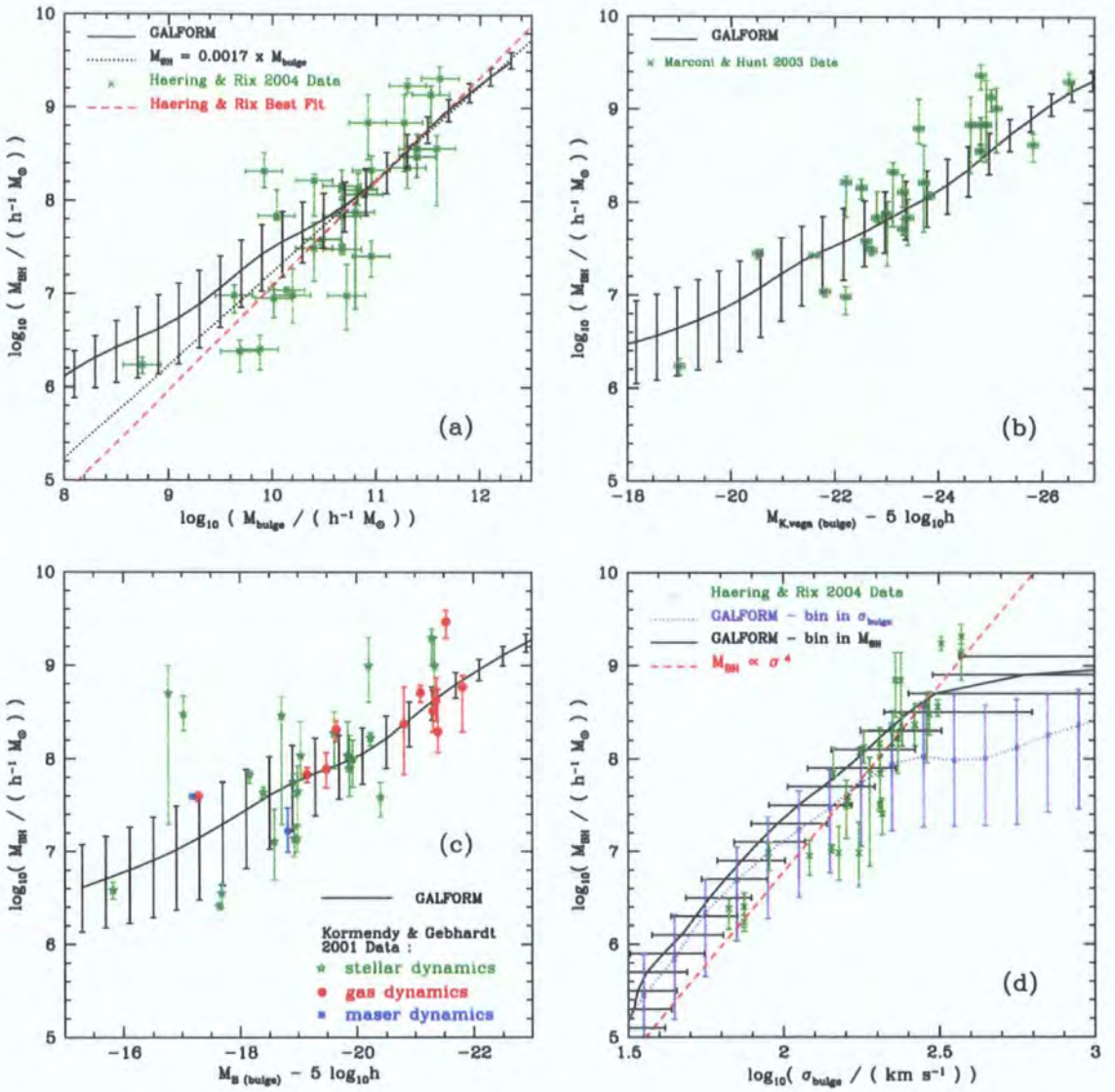

Figure 2.2: The relation between black hole mass, $M_{\rm BH}$, and a selection of properties of the spheroid of the host galaxy. Each panel shows the correlation with a different bulge property: (a) the stellar mass of the bulge; (b) the bulge rest-frame K-band magnitude; (c) the rest-frame B-band bulge magnitude; (d) the velocity dispersion of the bulge. The model predictions are shown by the line with errorbars: the line shows the median and the errorbars the 10–90 percentile spread of the distribution. The observational measurements are shown by symbols, with sources indicated in each panel.

the $M_{\rm BH} - M_{\rm bulge}$ relation are non-trivial predictions of the model. The scatter is due to the variation in the fraction of the stellar mass of a bulge which originally formed stars quiescently, before being rearranged into the spheroid.

Further support for both the galaxy formation model and our new model for the growth of black holes comes from examining the other relationships between black hole mass and observable properties of the galactic spheroid, as shown in Fig. 2.2 (b)-(d).

In Fig. 2.2 (b), we compare our model predictions for black hole mass as a function of the K-band magnitude of the bulge with the measurements by Marconi & Hunt (2003). Again, the match is very good. K-band magnitudes correlate well with stellar mass. In the model, the K-band magnitude depends upon the star formation and merger history of the galaxy, taking into account all of the progenitors of the galaxy, its dust content and its linear size. Observationally, this property is completely independent of the bulge stellar mass estimates based on the velocity profile fitting method used by Häring & Rix (2004).

It is notable that the scatter in both the $M_{\rm BH} - M_{\rm bulge}$ and $M_{\rm BH} - M_{\rm K}$ relations decreases significantly as the bulge magnitude gets brighter. A number of factors may contribute to this result. As we will show in Chapter 3, less massive black holes vary far more in their formation histories than do larger black holes. Therefore, for bulges hosting less massive black holes, there is more scatter in the time available for stars to form in progenitor discs before starbursts and black hole accretion occur. Stars in larger ellipticals and bulges tend to be formed earlier. Once stellar populations exceed a certain age, scatter in their ages has only a small impact on colour and luminosity.

In Fig. 2.2 (c) we plot black hole mass against the B-band magnitude of the bulge and compare the model with a compilation of data by Kormendy & Gebhardt (2001). The scatter in this relationship is the greatest of all four variations on the $M_{\rm BH}$ – bulge relations shown in Fig. 2.2. This is due to the sensitivity of the B-band magnitude to the details of the recent star formation history of the bulge which can vary considerably between galaxies with similar mass black holes.

Finally, in Fig. 2.2 (d), we compare the model prediction for the $M_{\rm BH} - \sigma_{\rm bulge}$ relation to data from Häring & Rix (2004). We calculate the velocity dispersion directly from the circular velocity of the bulge, assuming $\sigma_{\rm bulge} = V_{\rm circ,bulge}/\sqrt{3}$ (Almeida et al. 2006). The full details of the calculation of $V_{\rm circ,bulge}$ are given in Cole et al. (2000), but we note that this is a particularly uncertain property in our model which depends on poorly understood processes such as adiabatic contraction and the effect of mergers on the structure of the

bulge. Despite this crude model, we obtain a reasonable match to the data, reproducing the tightness of the relationship. For the less massive black holes, our model gives $M_{\rm BH} \propto \sigma^4_{\rm bulge}$, which compares well with the Tremaine et al. (2002) estimate of the slope of 4.02 ± 0.32. However, for black holes more massive than $M_{\rm BH} = 10^{7.5} h^{-1} M_\odot$, the slope is shallower than observed, closer to $M_{\rm BH} \propto \sigma^3_{\rm bulge}$. The problem is more obvious when we choose to bin our data in $\sigma_{\rm bulge}$ (purple line) than when we bin in $M_{\rm BH}$, suggesting that it is at least as much a problem with the calculation of $\sigma_{\rm bulge}$ as it is a problem with that of $M_{\rm BH}$. To some extent, direct accretion of cooling gas from a hot reservoir may help to bring the slope of the $M_{\rm BH} - \sigma_{\rm bulge}$ relation closer to that observed (Bower et al. 2006; Chapter 6 of this Thesis). However, we do believe that, due to calculational errors, we have a tail of bulges with unrealistically high velocity dispersions.

We have identified two possible primary problems which may contribute to this, both of which are likely to propagate through the hierarchy (since during mergers, the circular velocity and circular radius of the progenitor galaxies are always an input when we calculate the circular velocity and circular radius of the final galaxy). Firstly, in our model, we assume that after baryonic dissipaton and star formation, the baryonic mass in the centre of a dark matter halo acts gravitationally upon the halo, leading to a gravitational contration of the halo which occurs adiabatically. However, some recent simulations (Navarro et al. in prep.) suggest that it may in fact be more accurate to neglect this adiabatic contraction of the dark matter halo. Secondly, in our model, we assume that when gas in a halo cools, contract and forms a disc, it conserves its angular momentum. We calculate the disc scale-size to be the size where the disc is rotationally self-supported. Thus discs formed from low angular momentum material (i.e. gas from haloes with low spin) may contract to very small radii, and the velocity dispersion of any bulge later formed from this disc material will be correspondingly high. However, a disc with low angular momentum is likely to become dense enough to be unstable to its own self-gravity before it becomes rotationally self-supported – the resulting bulge will have a longer scale length and a lower velocity dispersion than it would if we neglected this disc instability. In more recent work, Bower et al. do not allow any galactic discs to contract further than the point at which they become unstable to a disc instability – a galactic disc contracts via baryonic dissipation, but stops contrating as soon as it reaches the point where *either:* it becomes rotationally self-supported and forms a stable disc *or:* it becomes unstable to its own self-gravity before becoming rotationally self-supported, undergoing a disc instability and starburst to form a bulge.

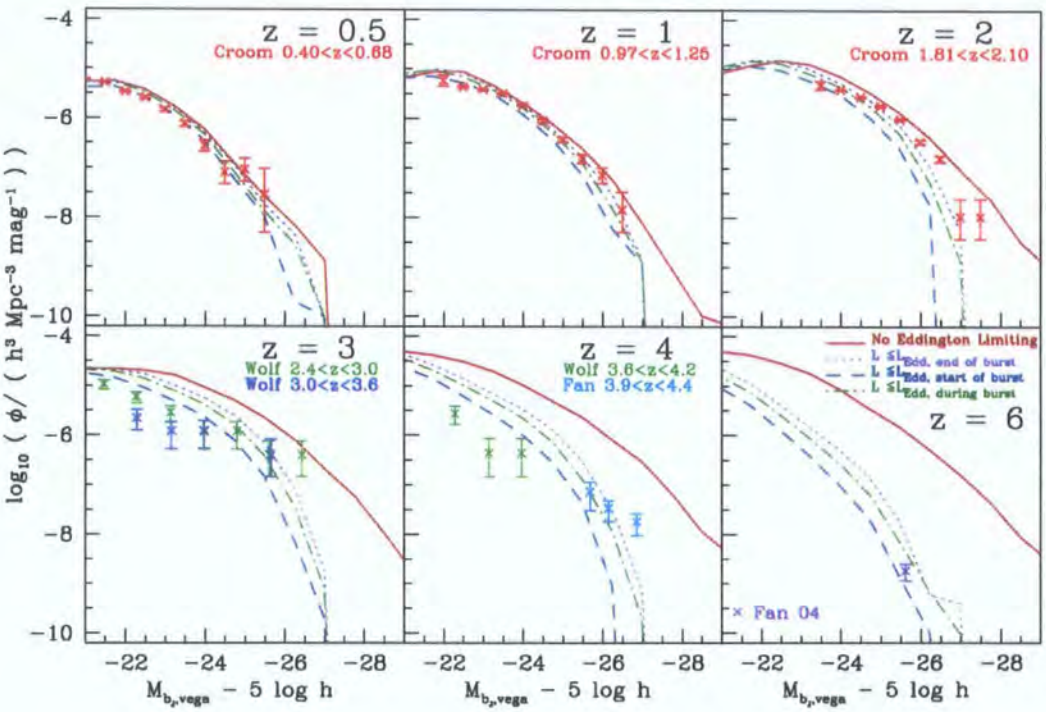

Figure 2.3: The quasar luminosity function at selected redshifts, as indicated in each panel. The model predictions are shown by lines and the data by symbols, with the source indicated in each panel. The different line styles correspond to different assumptions for how the quasar luminosity depends upon the Eddington luminosity of the black hole, as indicated by the legend in the bottom right panel. The data are taken from the following papers: Croom = Croom et al. 2004; Fan = Fan et al. 2001; Wolf = Wolf et al. 2003; Fan 04 = Fan et al. 2004.

The slope of the $M_{\rm BH} - \sigma_{\rm bulge}$ relation we obtain here should not be taken as a robust prediction of our model. Our calculation of $\sigma_{\rm bulge}$ could become progressively less accurate with increasing bulge mass. An independent test of the prescription used to compute the size of the spheroidal component of galaxies is presented in Almeida et al. (2006).

The level of agreement with observations that we find between different bulge properties and black hole mass is encouraging and suggests that, overall, our model model of galaxy and black hole formation is on a firm footing.

2.3.2 Predictions for the luminosity function of quasars

In order to establish further the credentials of our model, we present some illustrative predictions for the evolution of the quasar luminosity function. Further assumptions and model parameters are required to assign a luminosity to the quasar phase which occurs when the black hole accretes gas during a galaxy merger. In this section, we give a brief outline of our model for calculating the luminosity of the quasar and present some results for the quasar luminosity function at different redshifts. These predictions are included here for completeness and to allow comparison with previous work (e.g. Kauffmann & Haehnelt 2000). We will explore the form and evolution of the quasar luminosity function in more detail in future work.

There are two basic parameters in our model for quasar luminosity: the lifetime of the quasar, $t_{\rm Q}$, and the fraction of the accreted mass-energy that is turned into the bolometric luminosity of the quasar, $\epsilon_{\rm Q}$. We assume that the gas available for accretion onto the black hole in a galaxy merger is accreted at a constant rate, $\dot{M}(t)$, over the quasar lifetime:

$$\dot{M}(t) = \Delta M / t_{\rm Q} \qquad \text{for} \qquad \mathrm{t} < \mathrm{t_Q}. \tag{2.3}$$

The quasar lifetime, $t_{\rm Q}$, is assumed to be directly proportional to the dynamical time of the bulge, $t_{\rm bulge}$. In the simplest case, without imposing any further conditions on the luminosity of the quasar, this assumption results in a top-hat light curve:

$$L_{\rm Q}(t) = \epsilon_{\rm Q} \dot{M}(t)\, c^2 \qquad \text{for} \qquad \mathrm{t} < \mathrm{t_Q}. \tag{2.4}$$

When computing the luminosity of quasars, the Eddington limit may play an important role. A quasar is said to be radiating at its Eddington limit when the pressure of the radiation emitted following accretion onto the black hole balances the gravitational force exerted by the black hole on new material that is being accreted. The Eddington limit depends upon the mass of the black hole. Physical mechanisms have been proposed which permit mass to be accreted at rates which exceed the Eddington limit (see, for example, Begelman 1978). Here, we show the impact of the Eddington limit on the luminosity of quasars by considering four different cases:

Case (1) No Eddington limit is applied to the bolometric luminosity of the quasar.

Case (2) The bolometric luminosity is limited by the Eddington luminosity corresponding to the black hole mass at the *end* of the accretion episode:

$$L_{\rm Q}(t) = \max(\, \epsilon_{\rm Q} \dot{\rm M}({\rm t})\, {\rm c}^2,\, {\rm L_{Edd}(M_{final})}\,) \tag{2.5}$$

Case (3) The bolometric luminosity is limited by the Eddington luminosity corresponding to the black hole mass at the *start* of the accretion episode:

$$L_{\rm Q}(t) = \max(\,\epsilon_{\rm Q}\dot{\rm M}({\rm t})\,{\rm c}^2,\,{\rm L_{Edd}(M_{start})}\,) \tag{2.6}$$

Case (4) The bolometric luminosity is limited by the Eddington luminosity corresponding to the black hole mass calculated *during* the accretion episode:

$$L_{\rm Q}(t) = \max(\,\epsilon_{\rm Q}\dot{\rm M}({\rm t})\,{\rm c}^2,\,{\rm L_{Edd}(M(t))}\,) \tag{2.7}$$

We assume that all visible quasars have identical, flat spectra over the range of wavelengths of interest, and that a fraction, $f_{\rm b_J}$, of the bolometric luminosity is emitted in the B and $\rm b_J$-bands. We adopt the often used value, $f_{\rm b_J} = 0.1$, to facilitate comparison with previous work. Taking into account the $\rm b_J$-band filter profile, we can calculate a magnitude for a quasar from its bolometric luminosity:

$$M_{\rm b_J,vega} = 13.2 - 2.5 \times \log_{10}(f_{\rm b_J} \times L_{\rm bol}/10^{40}{\rm erg\,s^{-1}}). \tag{2.8}$$

We set the parameters t_Q and ϵ_Q in order to produce a reasonable match to the Croom et al. (2004) measurement of the $\rm b_J$-band luminosity function over the redshift interval $0.5 \leq z \leq 2$, as shown in Fig. 2.3. At higher redshifts, we show a comparison between our predicted luminosity functions and the 1450 Å rest frame quasar luminosity functions from the SDSS survey (Fan et al. 2001a) and the combo-17 survey (Wolf et al. 2003), using the corrections given in the respective papers to convert to the B-band, and applying a further minor correction to the $b_{\rm J}$ band. To achieve the best fit, we require $t_Q = 0.3\,t_{\rm bulge}$ and $\epsilon_Q = 0.01$. (We note that an exponentially decaying mass accretion rate, with timescale $t_{\rm Q} = 0.1 t_{\rm bulge}$ (i.e. $\dot{M}(t) = \Delta M {\rm exp}(-{\rm t}/0.1\,{\rm t_{bulge}})$) gives very similar results.) In Fig. 2.4, we show the distribution of bulge dynamical times at a number of redshifts. A typical bulge has $t_{\rm bulge}$ of 2×10^7yr at $z = 1$, 8×10^6yr at $z = 3$ and 2.5×10^6yr at $z = 6$, although within each redshift bin, the distribution in $t_{\rm bulge}$ is very broad. As noted by Kauffmann & Haehnelt (2000), this redshift evolution in $t_{\rm bulge}$ helps us to reproduce the evolution in the quasar luminosity function.

Our quasar timescales may seem rather short and our quasar efficiencies rather low compared to other estimates in the literature estimates (e.g. $\epsilon_Q \sim 0.1$ is a canonical choice, and Martini & Weinberg (2001) estimate $t_{\rm Q} = 4 \times 10^7$yr at $z = 2$ which is a typical estimate). However, we remind readers that we have assumed that there is no population of obscured quasars, and our $t_{\rm Q}$ should be taken as the timescale for which

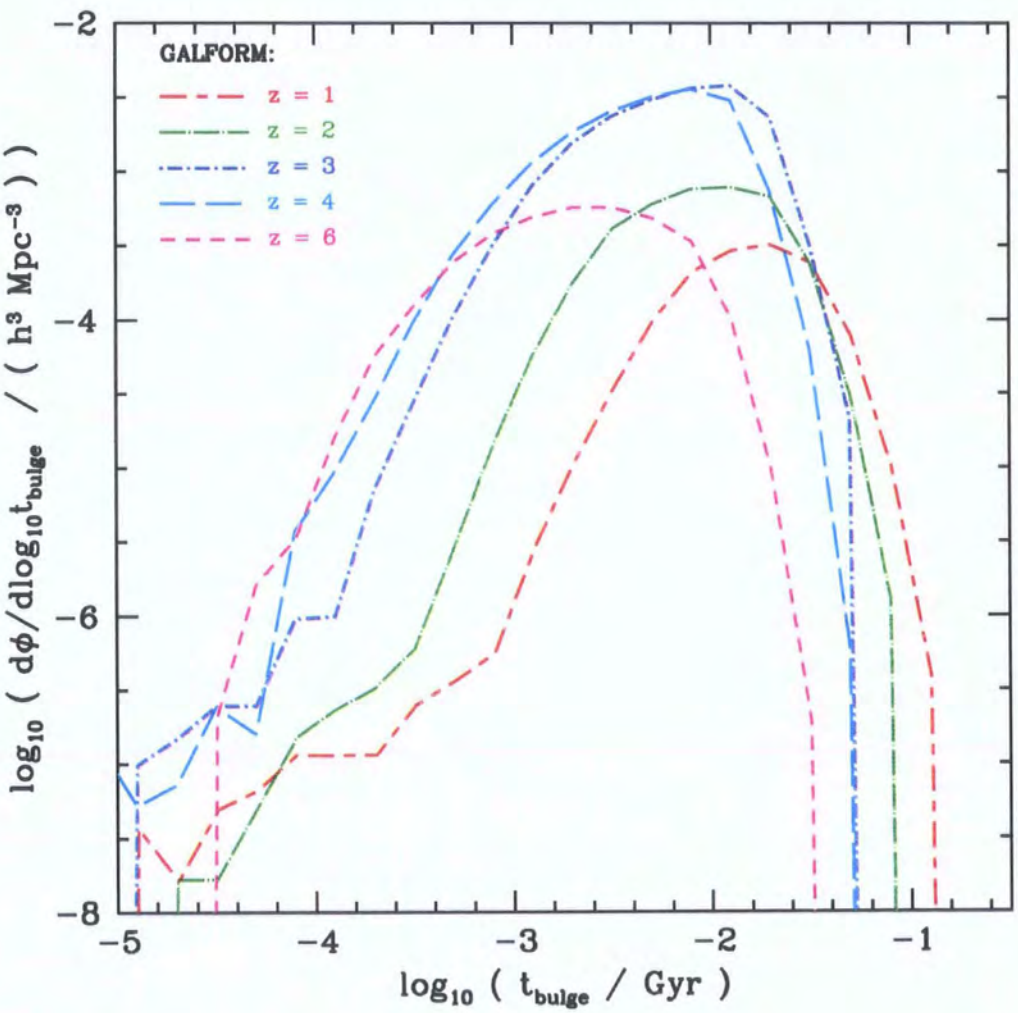

Figure 2.4: The distribution of bulge dynamical times. This is shown at a number of redshifts, as indicated by the key.

a quasar is visible, averaged over the entire population, rather than the time over which a black hole typically accretes mass. Assuming that only a fraction $f_{\rm visible}$ of quasars is unobscured, there is a degeneracy in our model. We would largely reproduce the quasar luminosity functions we have now if a timescale $t_{\rm Q}/f_{\rm visible}$ and an efficiency $\epsilon_Q/f_{\rm visible}$ were chosen, irrespective of our choice of Eddington limiting.

Our simple model does a reasonable job of reproducing the observed quasar luminosity function at $z \leq 2$, but over-predicts the luminosity function at higher redshifts. Our basic prediction for the quasar luminosity function (shown by the solid lines in Fig 2.3) shows strong evolution with redshift which cannot be described as pure luminosity evolution. If the Eddington limit is taken into account, then the form of the model predictions changes, particularly at bright luminosities, where the abundance of objects is strongly suppressed, with the result that the predictions match the data better. The suppression affects more objects at higher redshifts – the gas supply then is greater for any given mass of black hole, and the dynamical timescales are shorter, leading to higher rates of supply for any given mass of available gas. The predicted luminosity functions are relatively insensitive to the precise details of how the Eddington limit is allowed to influence the quasar luminosity.

Now that we have a model which reproduces the principal observables of the population of supermassive black holes in galaxies, we will proceed to study the growth and assembly histories of these black holes in the next Chapter.

Chapter 3

Black hole growth by mergers and accretion

In this Chapter, we move beyond the observable properties of black holes, and attempt to map out how they grow over time in the hierarchical cosmogony. In particular, do they grow in many small pieces, which later merge together, or by accretion onto a single object?

Firstly, in §3.1, we give illustrative examples of how black holes acquire mass, tracing the mass assembly history of two black holes. We then explore the demographics of the black hole population: in §3.2, we present the predictions for the evolution of the mass function of black holes and in §3.3, we investigate how the black holes are distributed between dark haloes of different mass. The next few sections deal with how black holes build up their mass. In §3.4, we show the distribution of progenitor masses of black holes, and, in §3.5, we address the issue of whether the accretion of gas or mergers is the main mechanism for accumulating black hole mass. We present results for the formation redshift of black holes in §3.6 and for their merger rates in §3.7. Finally, in §3.8, we compare the amount of baryons locked up in black holes with other phases, such as cold gas and stars.

3.1 Illustrations of black hole growth

Before concentrating on statistical descriptions, it is instructive to show some illustrations of how individual black holes grow in our simulations. These examples serve to provide a qualitative picture of the model, and to make clear certain definitions and results on black hole formation histories that will be of use later on. Note that, although space limitations restrict us to only two examples, there is, in fact, a rich diversity in black hole formation histories in the model.

The mass assembly history of two black holes is shown in Fig. 3.1 and Fig. 3.2. Fig. 3.1 shows the central galaxy in a halo of mass $2.9 \times 10^{11} h^{-1} M_{\odot}$ and Fig 3.2 shows the central

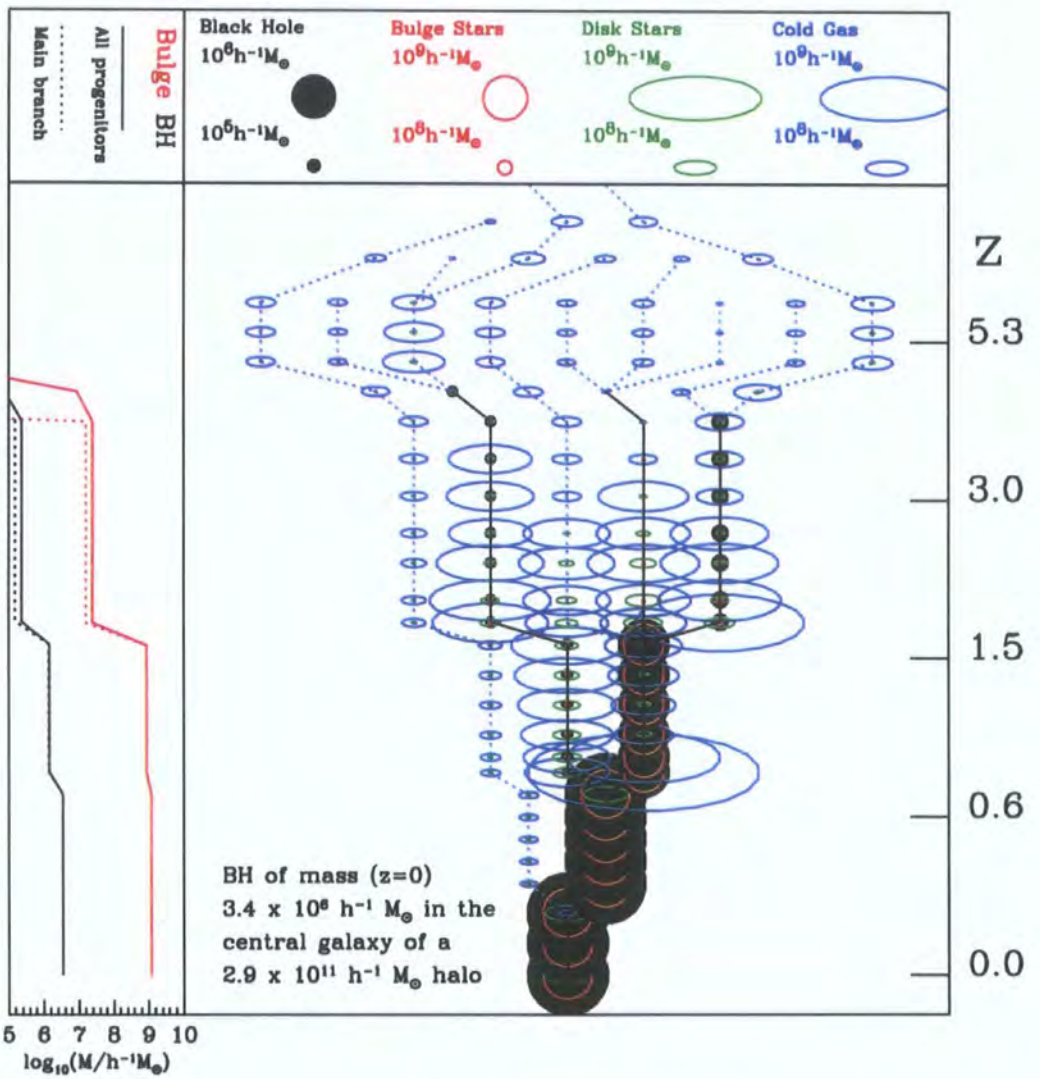

Figure 3.1: An example of a mass assembly tree of a black hole and its host galaxy. Progenitor galaxies without black holes are connected by dotted lines. The trees show the relative amounts of cold gas, disc stars, bulge stars and black hole mass, as indicated by the key. The area of the symbols is proportional to mass. The left-hand side panel shows the assembly of black hole mass and bulge stars, adding all progenitors (solid line) and tracing back the main branch, which is usually the most massive progenitor (dotted line). The left–right positioning in the plot is purely schematic and has no relevance to the spatial positions of galaxies. The final galaxy is the central galaxy of a halo of mass $2.9 \times 10^{11} h^{-1} M_\odot$.

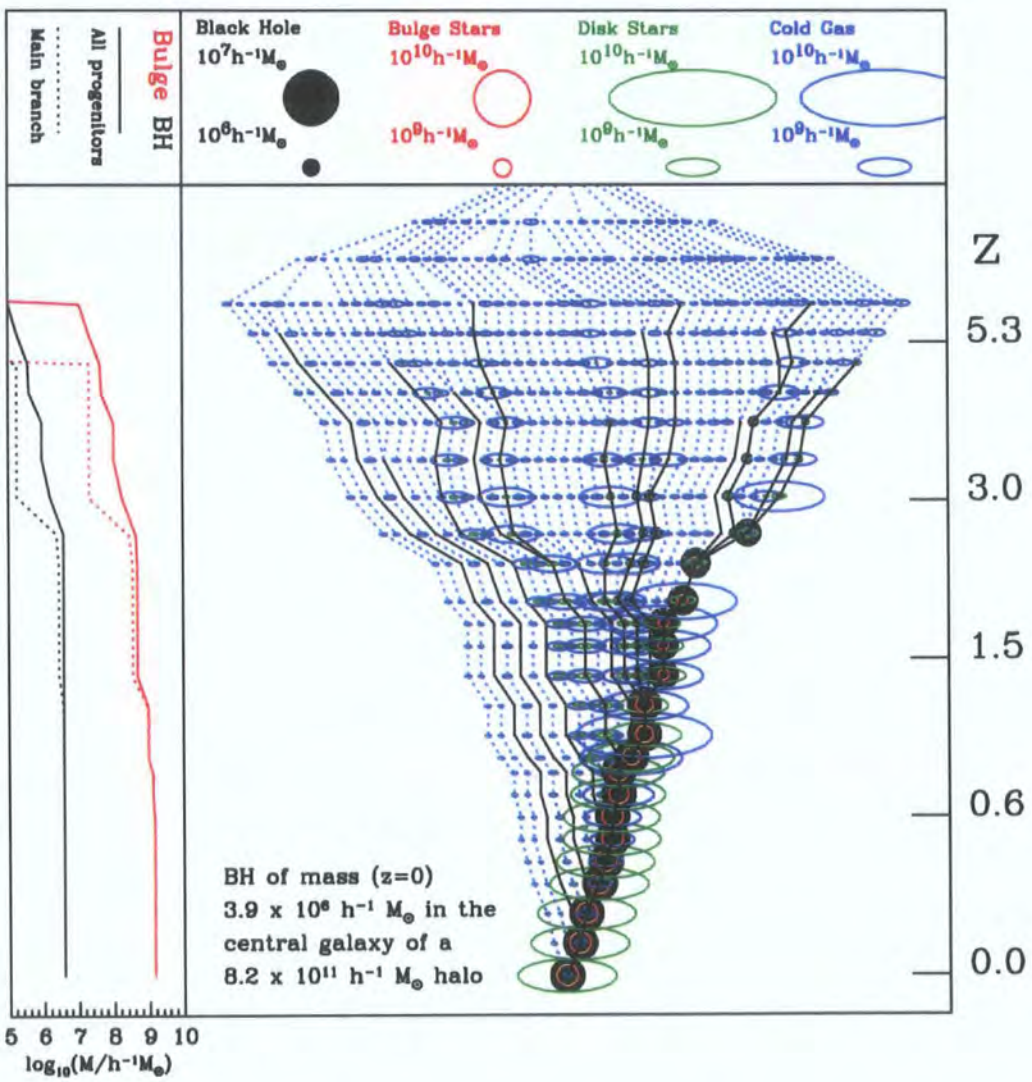

Figure 3.2: A second example of a mass assembly tree of a BH and its host. The final galaxy is the central galaxy of a halo of mass $8.2 \times 10^{11} h^{-1} M_\odot$.

galaxy in a halo of mass $8.2 \times 10^{11} h^{-1} M_{\odot}$. The main part of each panel follows the mass assembly tree. Various components are plotted, as indicated by the key at the top of each plot: black hole mass, bulge stars, disc stars and cold gas. The area of the symbols is proportional to the mass in a given component, with reference areas/masses provided at the top of each plot. Galaxies containing black holes are linked by solid lines, while galaxies not containing black holes are linked to their descendents with dotted lines. The redshifts plotted are the output redshifts of the simulation. The left–right positioning in the plot is schematic, and has no relevance to spatial positions of galaxies within the dark matter halo; the 'main branch' (i.e. the most massive progenitor at each merger) is always the right-most branch of the merger tree.

In some cases, the black hole in the 'main branch', which we denote the 'main progenitor', may not actually be *the* most massive of *all* the progenitor black holes at a given epoch, particularly at higher redshifts. We have chosen to avoid jumping from one branch of the black hole merger tree to another when following the 'main progenitor' backwards in time. Instead, we start from the present day black hole, find its most massive progenitor, and then build up a continuous branch by tracking the most massive progenitor at each of $\sim 25 - 30$ output redshifts.

In the side panel of each figure, we plot the cumulative masses of the black hole and bulge stars as a function of time. The solid line shows the total mass in these components, adding together all of the progenitors. The dotted line shows the mass in the branch tracing back the most massive progenitor of the present day black hole (the 'main branch').

In general, we find a very wide variety of black hole formation histories, and we have chosen the ones we plotted to be illustrative. The formation trees of the most massive black holes tend to be too large and complicated to plot effectively. Meanwhile, there is a high abundance of black hole trees with just one burst in their history, which were not very interesting to plot. All merger trees are included, however, in the quantitative results we present later.

Inspection of the mass assembly trees, particularly the one for the more massive galaxy, reveals that there can be many branches to the black hole merger trees at high redshifts. However, most of the black hole mass is contained in one or two main branches, as shown by the closeness of the solid and dotted lines in the side panels. In the Baugh et al. (2005) model, the quiescent star formation timescale is independent of the dynamical time. This results in discs which are gas rich at early epochs (blue discs), with significant quantities

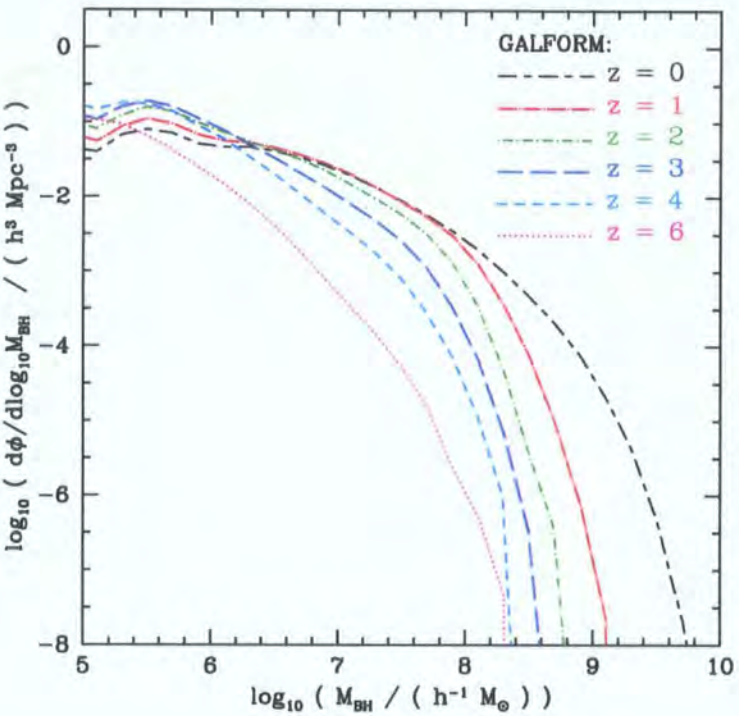

Figure 3.3: The evolution of the black hole mass function with redshift, as indicated by the key.

of stars only forming at relatively recent epochs (green discs). At later times, it is also apparent that the ratio of the stellar mass of the bulge to the mass of the black hole is increasing. We will present predictions for the evolution of the $M_{\rm BH}$ – bulge relations in a later section.

3.2 Black hole mass function

The black hole mass function at various redshifts is shown in Fig. 3.3. The high mass end advances to higher masses at lower redshifts. This is unsurprising in a hierarchical galaxy formation model, and reflects the corresponding evolution of the dark matter halo mass function. The predicted evolution in the black hole mass function is quite strong. This is in contrast with observational claims that the abundance of large black holes does not vary with redshift (e.g. McLure & Dunlop 2004). Such studies, however, typically include only optically selected quasars, and so can only probe accreting black holes. We examine the relationship of accreting black holes to the general black hole population in a later section on downsizing (§4.2) and consider the implications of the rare, massive black holes at high redshift inferred from the observations of Fan et al. (2001a) in the

discussion (§7.1).

At redshift zero, the break in the black hole mass function occurs around $10^8 h^{-1} M_\odot$. This corresponds to the scale at which there is a transition between accretion-dominated growth and merger-dominated growth (as we demonstrate specifically in §3.5). In larger galaxies hosting more massive black holes, the cold gas has been substantially depleted, so the black hole mass can only increase significantly through mergers. Gas depletion and suppression of further cooling by feedback processes is also the likely mechanism by which a break in the luminosity function of galaxies is produced (Benson et al. 2003a; Bower et al. 2006; Croton et al. 2006).

From the observed $M_{\rm BH} - M_{\rm K,bulge}$ relation, $M_{\rm BH} = 10^8 h^{-1} M_\odot$ corresponds to a K-band magnitude $M_{\rm K,vega(bulge)} - 5\log_{10} h \sim -23.5$. This is very close to the break in the K-band luminosity function, $M^*_{\rm K} - 5\log_{10} h = -23.44 \pm 0.03$ (Cole et al. 2001). Similarly, from the observed $M_{\rm BH} - M_{\rm bulge}$ relation, $M_{\rm BH} = 10^8 h^{-1} M_\odot$ corresponds to $M_{\rm bulge} \sim 3 \times 10^{10} M_\odot$. This is the stellar mass at which Kauffmann et al. (2003) find a transition in galaxy properties. Although black hole mass is related to bulge properties only, the identification of the knee in the black hole mass function with a transition in the global properties of galaxies is reasonable since galaxies brighter and more massive than the transition mass tend to be bulge-dominated. The conclusion is that galaxies (particularly galactic bulges) and black holes grow together (as demonstrated graphically in the side panels of Fig. 3.1 and Fig. 3.2).

As time advances, the black hole mass function becomes progressively flatter at the low-mass end. To a large extent, this is a generic feature of a hierarchical mass assembly model in which small objects merge into larger objects (at least when this effect is not exceeded by the production of new low mass objects). A further contribution to the change in slope comes from less massive black holes accreting larger amounts of gas as a fraction of their mass than larger ones (i.e. downsizing, §4.2). The combination of these effects is greater than the effect of the formation of new, lower mass black holes; most black holes are seeded at high redshift, as discussed in §3.6.

3.3 Black hole demography: the conditional mass function

Black holes of a given mass form in haloes with a broad range of masses. The contribution to the black hole mass function from different ranges of dark matter halo mass are shown in Fig. 3.4. At the high mass end of each of these conditional mass functions, there is

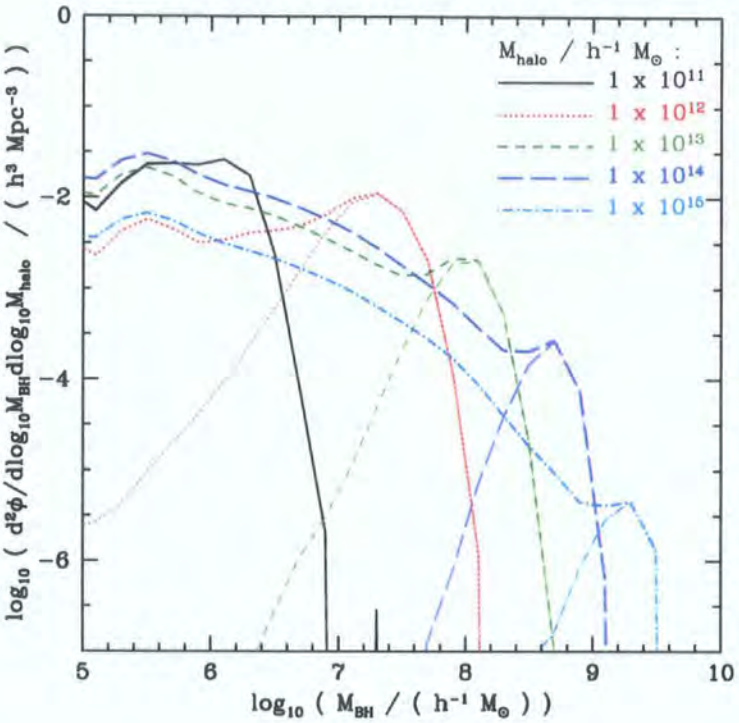

Figure 3.4: In thick lines, the contribution to the black hole mass function from dark matter haloes of various masses, as indicated by the key. The black hole mass function for each halo mass is normalized to be the mass function per dex in dark matter halo mass. Overplotted in thin lines are the corresponding contributions from the central galaxies only.

a peak and a cut-off. The peak corresponds to the mass of the black hole in the central galaxy, which increases as the mass of the central galaxy bulge which is strongly correlated with the mass of its host halo. This phenomenon is not restricted to black holes. Eke et al. (2004) find tentative evidence for a similar bump in the galaxy luminosity function of groups and clusters which they attribute to central galaxies, although this remains controversial (Yang et al. 2005).

In galaxy formation models, a bump is sometimes present in luminosity functions where only galaxies in a limited range of halo masses are selected – this is because of the contribution of central galaxies (Benson et al. 2003b). This reflects the different physical processes relevant to central and satellite galaxies: in the model, satellite-satellite mergers are not allowed, while all cooling gas is funnelled into the central galaxies. These simple assumptions, common in semi-analytic models, have been validated in gasdynamic simulations (Zheng et al. 2005). Models with intense star formation in bursts, such as the

Baugh et al. (2005) model used here, smear out the bumps somewhat (Eke et al. 2004), since the bursts introduce additional scatter in the properties of galaxies that form in haloes of a given mass.

3.4 Mass function of progenitor black holes

The mass distributions of the progenitors of present day black holes are shown in Fig. 3.5, at a range of redshifts. In the top panel, we show results for black holes of present day masses $10^7 - 10^{7.5}h^{-1}M_\odot$ and in the bottom panel we show results for present day masses of $10^9 - 10^{9.5}h^{-1}M_\odot$. The evolution of these progenitor mass functions from $z = 6$ to $z = 1$ looks remarkably similar to that of the universal black hole mass function, albeit truncated at the final $z = 0$ black hole mass, and with an overall normalization which increases with increasing final black hole mass.

The similarity of the form and evolution of the progenitor mass functions with those of the universal mass function is remarkable. Only at the lowest progenitor redshift plotted ($z = 1$) for black holes with present day masses in the range $10^7 - 10^{7.5}h^{-1}M_\odot$ do we see a significant deviation from the form of the overall mass function. The progenitor mass function in this case is rather flat, with fewer low mass black holes and more high mass black holes close in mass to the final black hole. The amplitude of the progenitor mass functions is substantially larger for the black holes with present day mass of $10^9 - 10^{9.5}h^{-1}M_\odot$: larger black holes have a significantly larger number of progenitor black holes. This fits in well with our later result (§3.5) that less massive black holes grow primarily by accretion onto a single main branch, whereas black holes larger than $5 \times 10^7 h^{-1}M_\odot$ grow primarily by mergers of pre-existing black holes.

3.5 Black hole growth by mergers and accretion

We come now to one of the principal results of this Thesis. In Fig. 3.6, we plot the fraction of the mass in the 'main branch' which is assembled by mergers or gas accretion as a function of redshift. We show results for black holes in two mass ranges at redshift zero: $10^7 - 10^{7.5}h^{-1}M_\odot$ (top) and $10^9 - 10^{9.5}h^{-1}M_\odot$ (bottom). Fig. 3.6 shows that at high redshifts, growth by accretion dominates over growth through mergers. Mergers become increasingly important as redshift decreases, but for the $10^7 - 10^{7.5}h^{-1}M_\odot$ black holes, the cumulative growth by mergers never exceeds the cumulative growth by accretion, even at

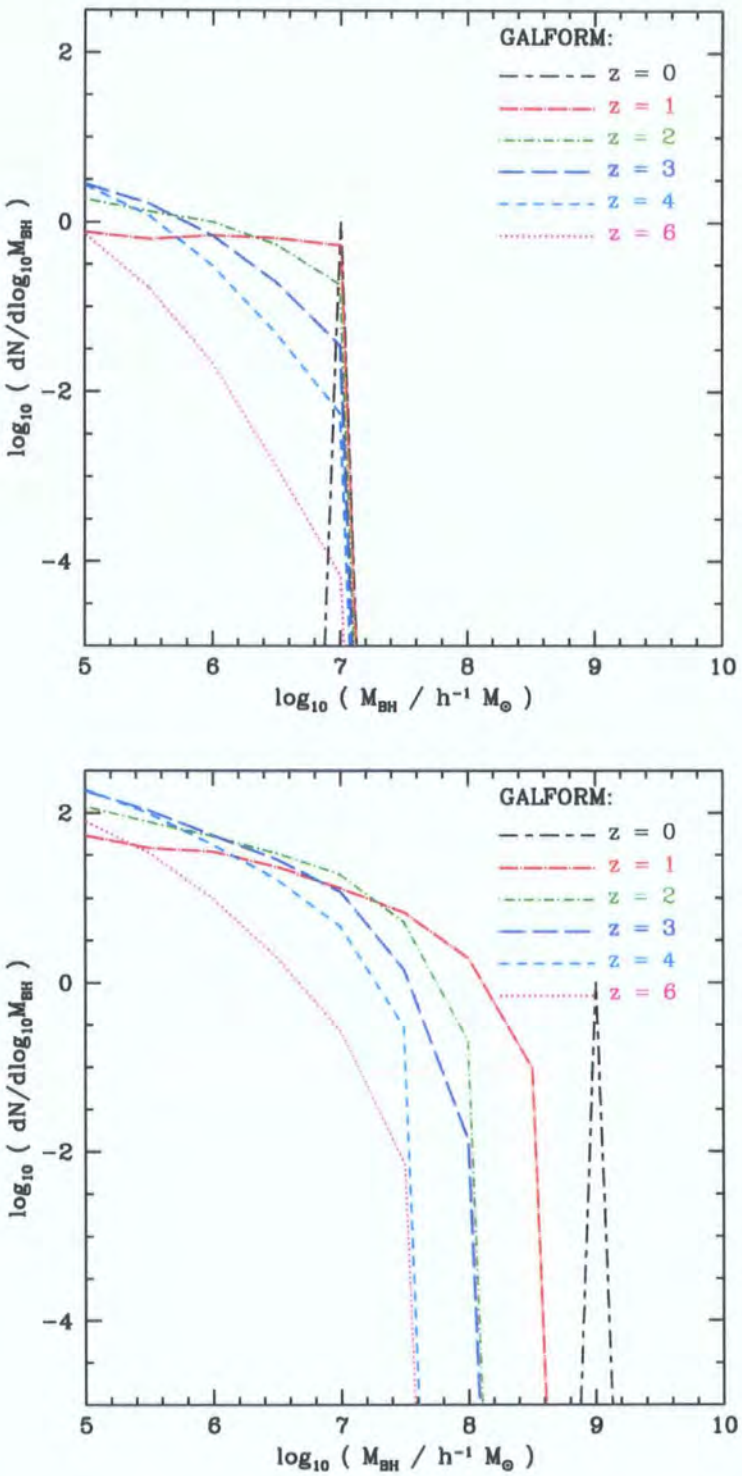

Figure 3.5: The mass functions of progenitor black holes. Black holes are selected at $z = 0$ to be of mass $10^7 - 10^{7.5} h^{-1} M_\odot$ (top) and $10^9 - 10^{9.5} h^{-1} M_\odot$ (bottom). The distribution of progenitor masses is plotted at different redshifts, as indicated by the key. The mass functions are generated for a large sample of black holes at $z = 0$, and the normalization chosen so that each progenitor mass function is the mass function *per black hole at* $z = 0$.

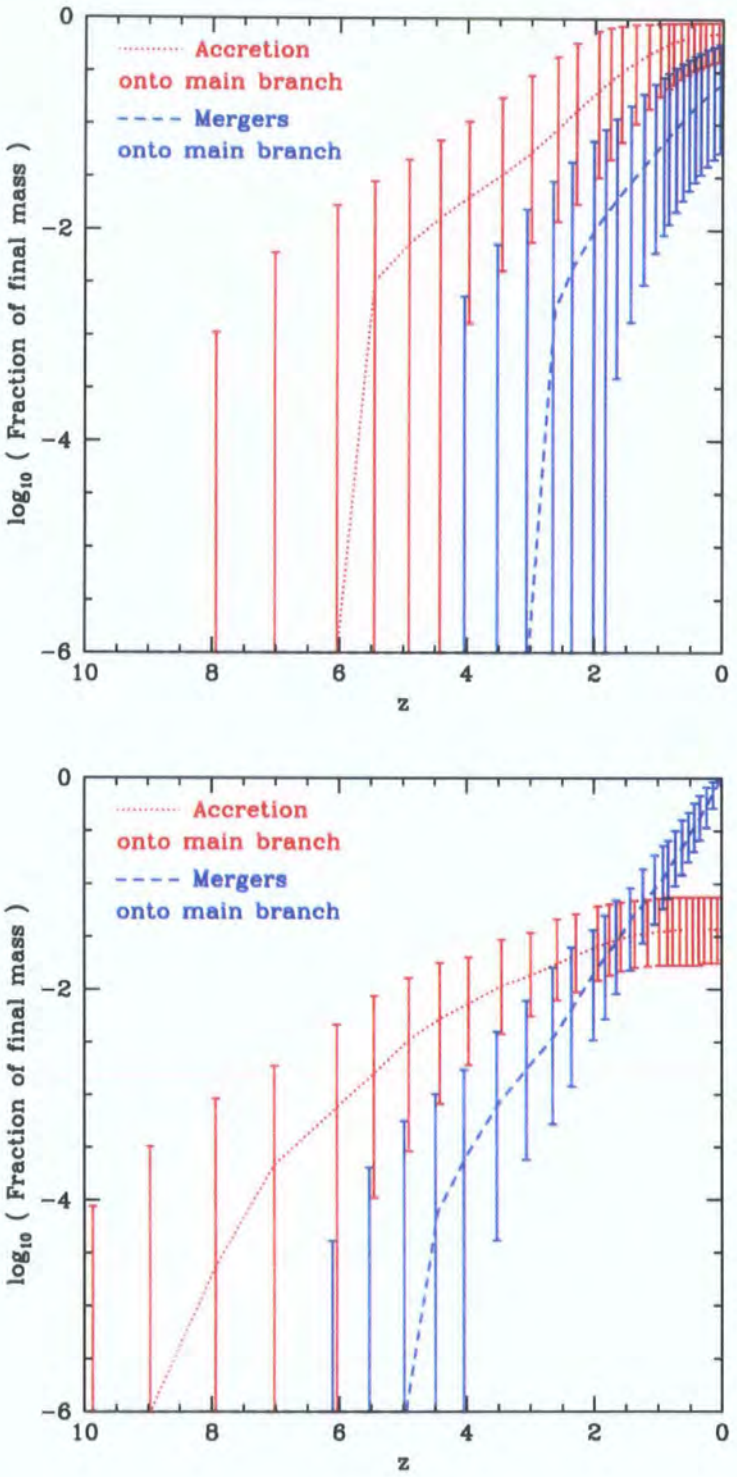

Figure 3.6: The cumulative growth with redshift of the black hole mass in the 'main branch' divided into the contribution from mergers and accretion. We consider large samples of black holes with $z = 0$ masses in the range $10^7 - 10^{7.5}h^{-1}M_\odot$ (top) and $10^9 - 10^{9.5}h^{-1}M_\odot$ (bottom). The dotted and dashed lines connect the medians of the distribution, while the 10–90 percentiles of the distribution are shown as errorbars.

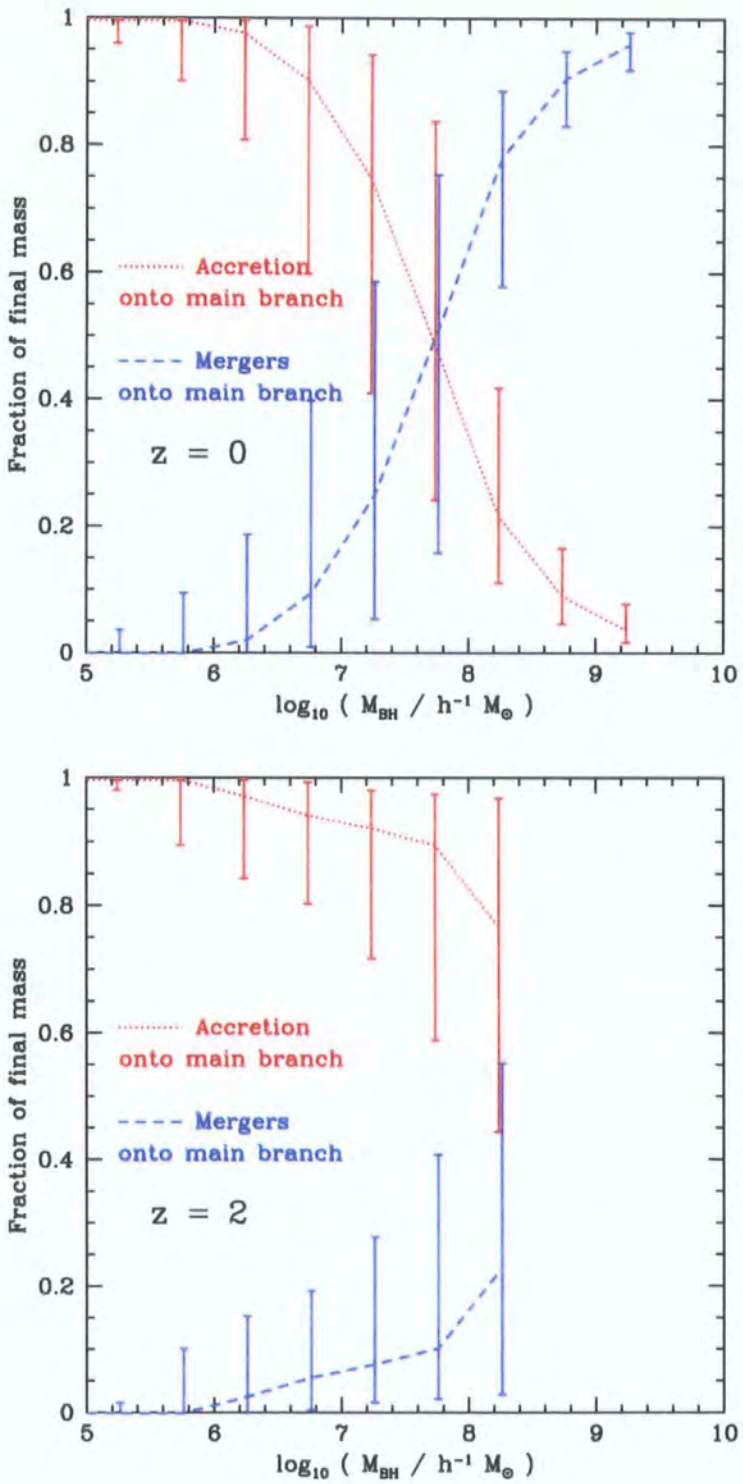

Figure 3.7: The cumulative fraction of the mass assembled by mergers and accretion, as a function of final black hole mass at $z = 0$ (top) and $z = 2$ (bottom). The medians are connected by lines, and the 10–90 percentile spread of the distribution is shown as an errorbar for each black hole mass.

redshift zero. However, for black holes of mass $10^9 - 10^{9.5}h^{-1}M_\odot$ at $z = 0$, the cumulative growth of their main progenitors by mergers exceeds that by accretion around a redshift of 1.7, and growth by accretion almost halts after this.[1] By redshift zero, the cumulative mass assembled by mergers greatly exceeds that assembled by accretion. The declining importance of growth by accretion for black holes of mass $> 10^{7.5}h^{-1}M_\odot$ reflects the decline in the amount of gas available in mergers as more and more of the gas in collapsed haloes is consumed into stars.

In Fig. 3.7, we plot the fraction of the mass of black holes which, by $z = 0$ (top) and $z = 2$ (bottom), has been accumulated by mergers or accretion onto the 'main branch'. This is shown as a function of black hole mass. Fig. 3.7 (top) shows that, by redshift zero, low mass black holes have accumulated nearly all of their mass by direct accretion onto a single 'main branch', while the most massive black holes accumulate 80–90% of their mass by mergers of less massive black holes onto the 'main branch'. The transition from accretion-dominated growth to merger-dominated growth occurs at a $z = 0$ mass of just over $10^8h^{-1}M_\odot$. Fig. 3.7 (bottom) shows that at $z = 2$, all black holes, even those more massive than $10^8h^{-1}M_\odot$, grow predominantly by accretion, although there is a contribution from mergers which increases with black hole mass. Comparison of the results in Fig. 3.7 for $z = 0$ and $z = 2$ shows that for any given black hole mass, growth by accretion is more significant for a black hole at $z = 2$ than at $z = 0$, and that this difference is greater for the more massive black holes. This is consistent with the idea that the luminous growth (i.e. growth by direct accretion of gas) of higher mass black holes switches off towards lower redshifts (see §4.2).

We note that in our model, a substantial fraction of the mass of the highest mass black holes at low redshift assembles through repeated merging of black holes – is this realistic? For most galaxies with black hole mass estimates, it is not possible to determine observationally whether the central 'massive dark object' is a single black hole or a black hole binary. Thus, to some extent, it does not matter too much whether a black hole binary in the centre of a galaxy merges, provided that the central mass is not assembled from more than two black holes. However, if a third black hole falls to the centre of a galaxy hosting a black hole binary, one of the three black holes will be ejected by sling-shot ejection. Thus if the central dark objects in our model are made by mergers of many black holes, it becomes very relevant how long black holes take to merge. Black

[1] This redshift can vary from black hole to black hole – $z = 1.7$ is the redshift where the median growth by mergers exceeds the median growth by accretion.

hole binaries only generally harden to an orbital radius of $\sim 1pc$ by stellar dynamical interactions – many gas-poor elliptical galaxies may in fact host binary black holes. We must consider in detail how long it takes black hole binaries to merge, and how much black hole mass is lost through 3-body gravitational slingshot processes. It may be that we are over-estimating the masses of the most massive black holes, and it is not clear how significant this is until we do the calculation. To some extent, the problem may be mitigated since in these three-body interactions, it is generally the lowest mass black hole which will be ejected.

3.6 The redshift of black hole formation

In Fig. 3.8 we show the formation redshifts of black holes, binned by $z = 0$ mass. Each of the six panels corresponds to a different definition of formation redshift. 'Formation' is defined as the time when either the main progenitor (right three panels) or the sum of all existing progenitor black holes (left three panels) first exceeds a given fraction of the final black hole mass. Where the formation redshift is defined as the time when the *main progenitor* first exceeds a given fraction of the final mass, we refer to this as the *mass assembly* redshift, since this is the redshift where the stated fraction of the mass has been assembled into a single object. Where the formation redshift is defined as the time when the *sum of all progenitors* first exceeds a given fraction of the final mass, we refer to this as the *mass transformation* redshift. This distinction between the *mass transformation* time and *mass assembly* time for black holes is analagous to that between the star formation time and stellar mass assembly time for the stars in a galaxy. We consider three different mass fraction thresholds to define formation times: 0.01 (top), 0.5 (middle) and 0.95 (bottom).

When we consider the assembly of 50% or 95% of the final black hole mass (Fig. 3.8 – middle-right & bottom-right), we see clear hierarchical behaviour; the more massive black holes at redshift zero peak in their formation times at lower redshift than the less massive ones. This is evidence for the hierarchical *assembly* of black hole mass into a single final object. However, when we consider the redshift at which 50% of the final black hole mass has *accreted* onto *any* black hole in the merger tree (Fig. 3.8 – middle-left), we see the opposite trend; in the mass range $M_{\rm BH} = 10^7 - 10^{9.5}h^{-1}M_{\odot}$, the more massive black holes display a distribution of formation redshifts which peaks at higher redshift. When we consider the redshift at which 95% of the final black hole mass is *accreted* onto

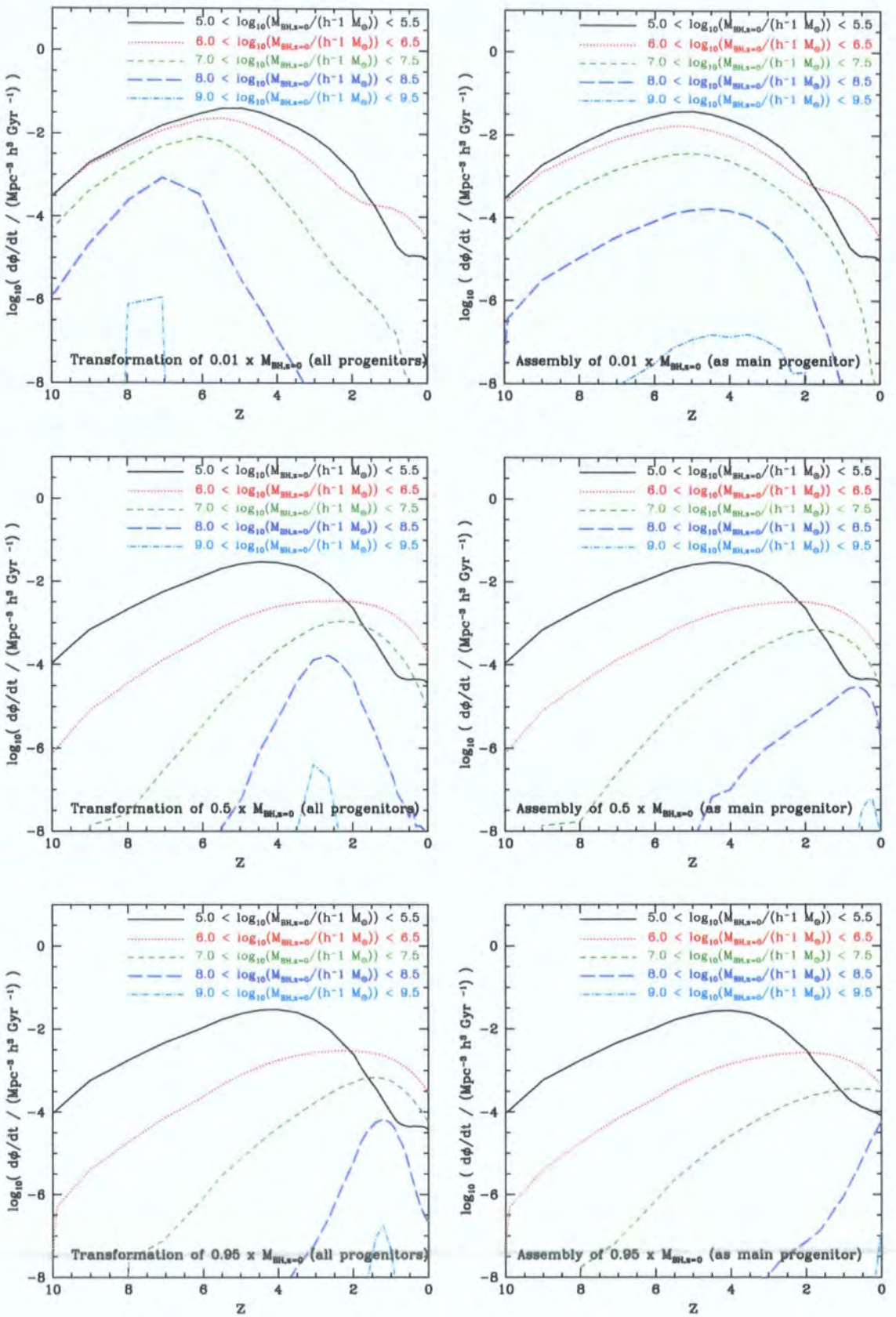

Figure 3.8: The distribution of formation redshifts of black holes in 5 different bins of $z = 0$ mass, as indicated by the key. The differing definitions of formation redshift used in each of the six plots are noted briefly on each plot and explained more fully in §3.6 of the text.

any black hole in the merger tree (Fig. 3.8 – bottom-left), we see further evidence of downsizing. Although formation peaks at a similar redshift ($z \sim 1.2$) for all black holes in the mass range $M_{\rm BH} = 10^7 - 10^{9.5}h^{-1}M_{\odot}$, the decline in the fraction forming as redshift approaches zero is far steeper for more massive black holes within this mass range. The observational evidence for "downsizing" refers to the accretion of mass onto a particular progenitor, which is accompanied by the release of energy. Hence, it is the latter trend which is relevant – as black hole mass increases, the redshift when mass is *accreted* onto *any* progenitor increases. We return to this point in §4.2.

We consider now the early growth of the black holes. Fig. 3.8 (top-left) shows the redshift when the first 1 per cent of the final black hole mass has collapsed into any of the branches of the merger tree. There is a clear trend for larger black holes to be seeded earlier. This is also a form of downsizing. All of the black holes in our largest mass bin ($M_{\rm BH} = 10^9 - 10^{9.5}h^{-1}M_{\odot}$) and many of those in the next mass bin, $M_{\rm BH} = 10^8 - 10^{8.5}h^{-1}M_{\odot}$, are seeded before reionization occurs in the model at $z = 6$. Another interesting feature of Fig. 3.8 is that, for almost any definition of formation time, less massive black holes have a much wider spread in formation times than more massive black holes.

There is little difference in the distribution of formation times of black holes of mass $M_{\rm BH} = 10^5 - 10^{6.5}h^{-1}M_{\odot}$ regardless of whether we use a definition which relates to the 'main branch' or to 'all progenitors'. This follows from our earlier result that black holes in this mass range formed almost exclusively by accretion onto a single object, with little contribution from mergers between black holes (§3.5, Fig. 3.7). The differentiation between the different definitions of formation time begins to become apparent for black holes of mass $M_{\rm BH} = 10^7 - 10^{7.5}h^{-1}M_{\odot}$ and is increasingly more significant as black hole mass increases further. This relates to our earlier result that the contribution to the final black hole mass from mergers of pre-formed black holes compared to the contribution from direct cold gas accretion onto the main branch increases strongly with increasing black hole mass (§3.5, Fig. 3.7).

3.7 Black hole merger rates

We show the merger rate per unit time of black holes as a function of redshift in Fig. 3.9. We show this for a number of mass thresholds which must be exceeded by both of the black holes that take place in the merger. The merger rates peak at lower redshift for

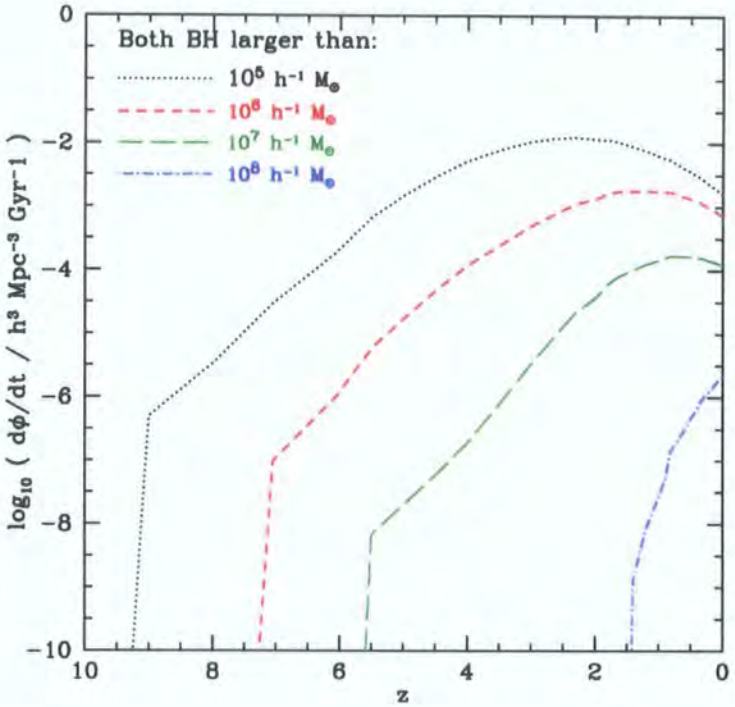

Figure 3.9: Black hole merger rate per unit time as a function of redshift. The merger rate is plotted for 4 different mass thresholds, as shown by the key, which the (pre-starburst) masses of both black holes must exceed.

more massive black holes, with the merger rate for the most massive bin still rising at $z = 0$. This is consistent with the trend seen at $z = 0$ that larger mass black holes grow primarily by mergers, while less massive black holes grow primarily by accretion (§3.5).

This behaviour in the growth and merging of black holes of varying mass is largely a reflection of the general hierarchical growth of structure, moderated in the case of galaxies and black holes by baryonic processes. The results presented in this section concern only mergers of black holes, not necessarily their total growth which can also involve accretion. There is no evidence for 'anti-hierarchical' behaviour in the evolution of black hole mergers. However, as we show in §6, this is perfectly compatible with quasar downsizing – black hole merging can be a 'dark' process in which no gas is present, whereas the observational evidence for downsizing refers to processes involving star formation or gas accretion.

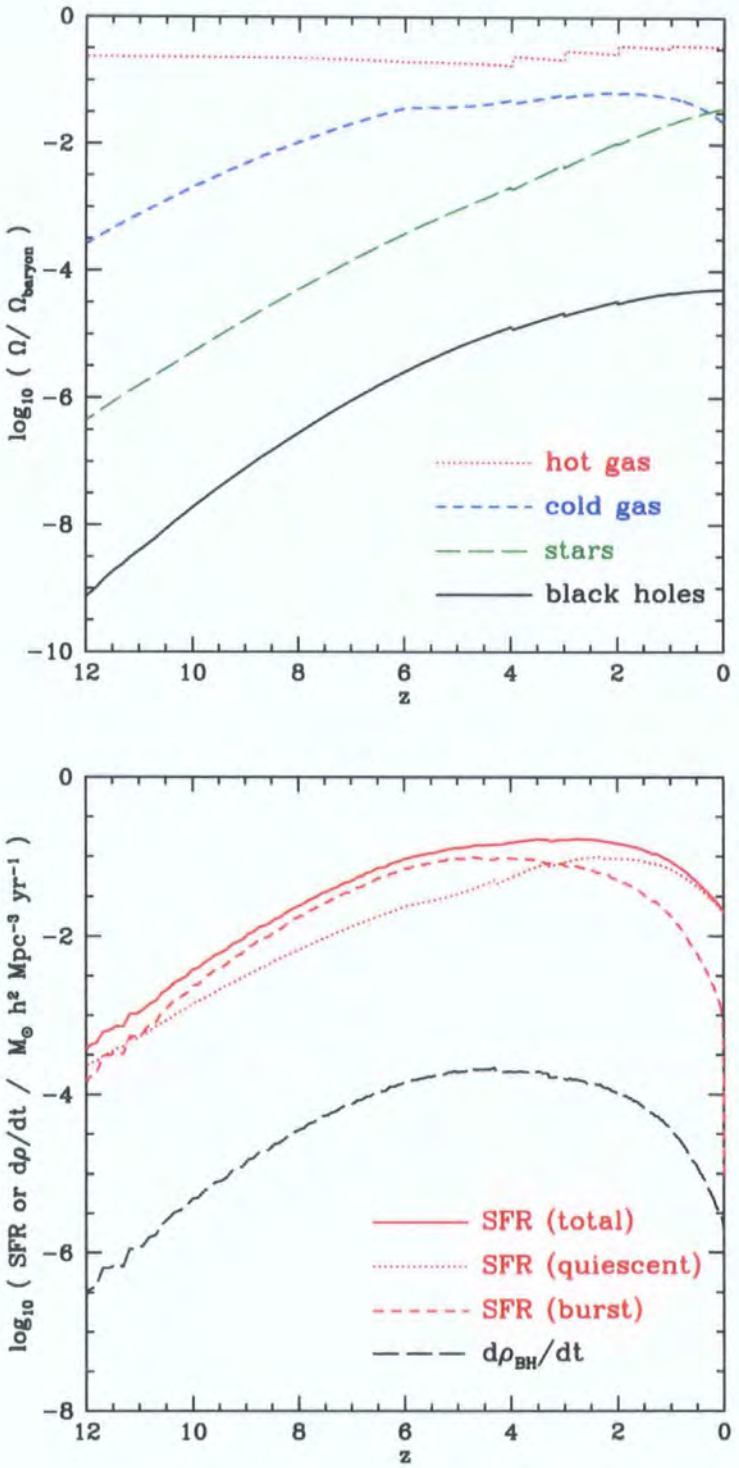

Figure 3.10: Top: The evolution of the fraction of the baryons in the universe in hot gas, cold gas, stars (disc plus bulge) and black holes. The small abrupt changes in some of the lines are due to simulation runs that finish at different times. Bottom: The variation with redshift of the global star formation rate (starbursts, quiescent and total) and the global rate of black hole growth.

3.8 The fraction of baryons in black holes

Having considered the formation of individual black holes, we now look at the global picture. In Fig. 3.10 (top), we show the integrated cosmic density of all the baryonic components of the universe; hot gas, cold disc gas, stars and black holes. After $z \sim 4$, the growth of black hole mass in the universe slows down in comparison to that of stars, as quiescent star formation begins to dominate over star formation in bursts. The decline in cold gas from redshift 2 to 0 goes a small way towards explaining the decline in quasar activity over this redshift interval. The decline in the galaxy merger rate and the transition from burst-dominated star formation to quiescent star formation also play a role. In Fig. 3.10 (bottom), we show the star formation rate, divided into burst and quiescent modes, and the rate of black hole growth. By construction in our model, black hole growth is more strongly correlated with the star formation rate in bursts than with star formation in general. Very broadly, although perhaps less so at low redshifts, black hole accretion tracks the overall star formation over cosmic time, as observed (Boyle & Terlevich 1998).

The cosmological mass density of black holes at $z = 0$ is a quantity of interest. In our model, we find that $\rho_{\rm BH} = 2.83 \times 10^5 M_\odot {\rm Mpc}^{-3}$. Observationally, $\rho_{\rm BH}$ is determined by integrating the black hole mass function which, in turn, is inferred from a combination of the velocity dispersion distribution of galaxies, the K-band luminosity function or the bulge stellar mass function, and the appropriate $M_{\rm BH}$ – bulge relation. Observed values of $\rho_{\rm BH}/(10^5 M_\odot {\rm Mpc}^{-3})$, converted to $H_0 = 70 {\rm kms}^{-1}{\rm Mpc}^{-1}$, are : 2.9 ± 0.5 (Yu & Tremaine 2002), 2.4 ± 0.8 (Aller & Richstone 2002), 2.8 ± 0.4 (McLure & Dunlop 2004), 4.2 ± 1.1 (Shankar et al. 2004) and $4.6^{+1.9}_{-1.4}$ (Marconi et al. 2004). Our estimate is towards the lower end of the broad range spanned by the observational estimates. We do not include any measurement errors in our estimate. A detailed comparison would need to take into account galaxy type (some estimates are based only on ellipticals), the flux limits of the observational samples and the treatment of the dispersion in the $M_{\rm BH}$ – bulge relations when converting from bulge properties to black hole masses. Many observational estimates assume that the scatter in $\log(M_{\rm BH})$, at a given value of the bulge property under consideration, is symmetrical. This assumption then leads to larger values of $\rho_{\rm BH}$ for larger assumed values of the scatter (McLure & Dunlop 2004). However, it is not at all clear that the scatter in these relations is symmetrical.

Chapter 4

Evolution and downsizing of the black hole population

In this Chapter, we consider the redshift evolution of the black hole population and of its growth. In §4.1, we make predictions of the evolution of the $M_{\rm BH}$ – bulge properties, and compare these to the indications in the available observational data. There appears to be an observational trend for black hole growth to 'downsize' to lower mass black holes as redshift decreases. In §4.2 we examine whether this occurs in our model. Finally, in §4.3, we discuss improvements to our model which may help us to produce higher black hole masses at high redshifts, as the observational data suggests may be necessary.

4.1 The evolution of the relation between black hole mass and bulge properties.

In this section, we discuss the evolution of the relationship between black hole mass and various galaxy bulge properties: K-band and B-band bulge magnitude, bulge stellar mass and bulge velocity dispersion. We show these relationships in Fig. 4.1, in each case plotting the model predictions for the $M_{\rm BH}$ – bulge relationships at $z = 0, 1, 2, 3, 4$ and 6. We discuss each of these in turn, briefly referring to any relevant observational data. However, it is difficult to make rigorous comparisons to current data. While relationships between black hole mass and bulge properties are fairly well determined at $z = 0$, this is currently not the case for $z > 0$, where observational samples are small and subject to selection effects. In particular, different surveys sample the population of galaxies and, where relevant, the AGN subpopulation, in ways that are not always straightforward to replicate in the models.

We show our model predictions for the $M_{\rm BH} - M_{\rm B,bulge}$ relation in Fig. 4.1(a) and the $M_{\rm BH} - M_{\rm K,bulge}$ relation in Fig. 4.1(b). In the model, these relationships shift to-

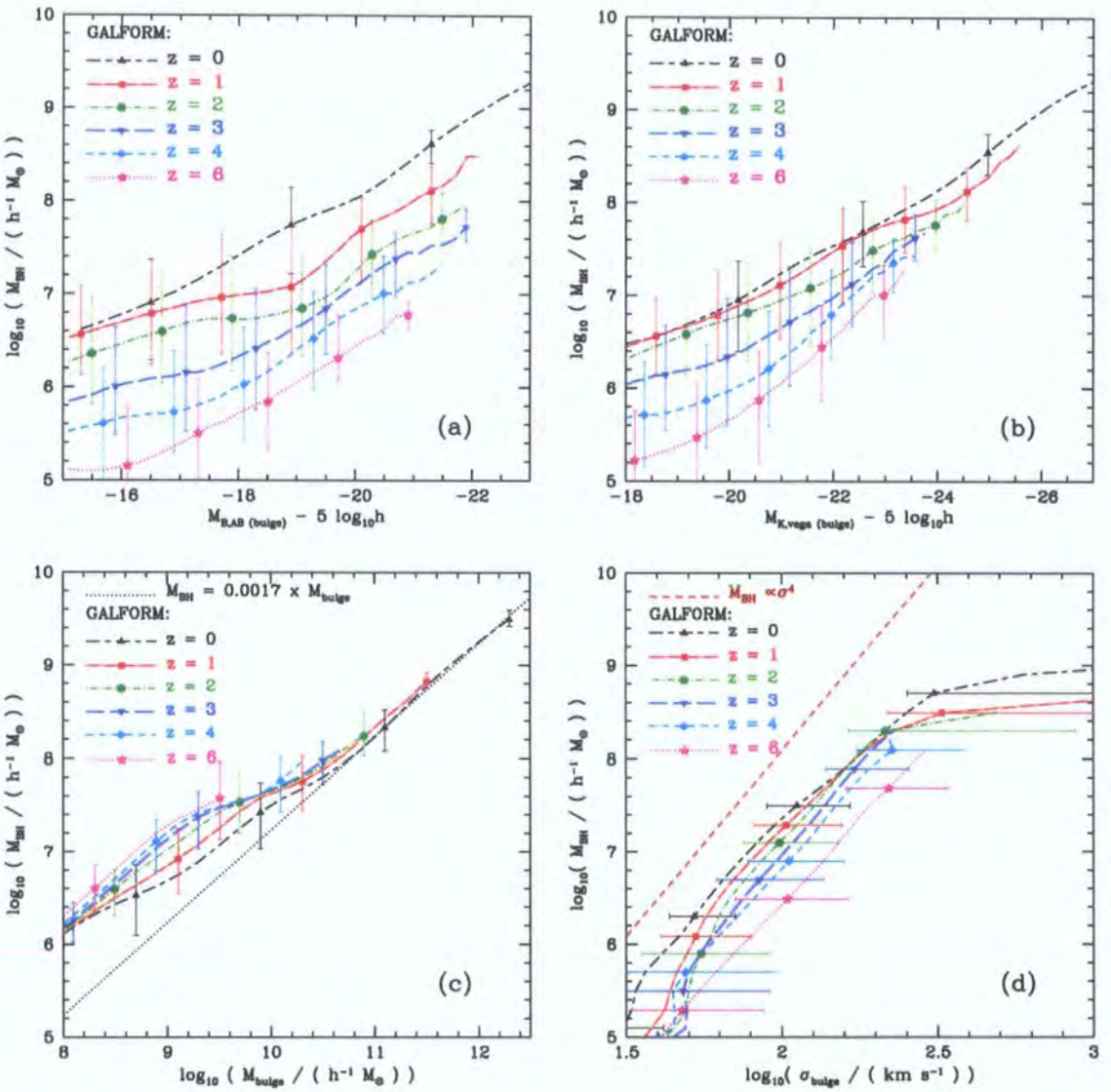

Figure 4.1: The redshift evolution of the relations between central black hole mass, $M_{\rm BH}$, and bulge properties. Each panel shows the relationship between $M_{\rm BH}$ and a different property of the host spheroid : (a) the bulge rest-frame B-band magnitude; (b) the bulge rest-frame K-band magnitude; (c) the stellar mass of the bulge; (d) the velocity dispersion of the bulge. The model predictions are shown by the symbols with errorbars; the lines show the median relations and the errorbars the 10–90 percentile spread of the distributions. Redshifts 0, 1, 2, 3, 4 and 6 are shown in different line types, as indicated by the key.

wards brighter magnitudes at higher redshifts. This is a reflection of the evolving stellar populations. The stellar populations of bulges at low redshift are older and thus less luminous than their high redshift counterparts. This effect more than compensates for any evolution in the opposite direction in the $M_{\rm BH} - M_{\rm bulge}$ relation, which we discuss below. The redshift evolution in the $M_{\rm BH} - M_{\rm B,bulge}$ relation is greater than that in the $M_{\rm BH} - M_{\rm K,bulge}$ relation because stellar populations dim more strongly with time in the B-band than in the K-band. Observationally, however, Peng et al. (2006), selecting high redshift quasars, find little trend in the $M_{\rm BH} - M_{\rm R,bulge,rest}$ relation with redshift, which conflicts somewhat with our prediction of an evolution towards brighter magnitudes as redshift increases.

We show the $M_{\rm BH} - M_{\rm bulge}$ relation in Fig. 4.1(c). There is no significant evolution in either the slope or scatter at large bulge masses. For $M_{\rm bulge} < 10^{10}h^{-1}M_{\odot}$, the black hole mass to bulge mass ratio increases with increasing redshift. Observationally, Peng et al. (2006) find that the ratio of $M_{\rm BH}/M_{\rm bulge}$ was 3–6 times larger at $z \gtrsim 2$ for AGNs than for quiescent galaxies at $z = 0$. McLure et al. (2006), selecting radio galaxies at $z > 0$ from the 3CRR catalogue, argue that $M_{\rm BH}/M_{\rm bulge}$ increases with redshift, and is ~ 4 times greater for radio galaxies at $z = 2$ than for quiescent galaxies at $z = 0$. We find that $M_{\rm BH}/M_{\rm bulge}$ was ~ 2 times greater at $z = 2$ than at $z = 0$ for $M_{\rm bulge} < 10^{10}h^{-1}M_{\odot}$. This evolution is in the same sense as and of comparable size to the observational trend, although the effect in the model is perhaps not as strong. As discussed in §2.3.1, the predicted variation in the $M_{\rm BH}/M_{\rm bulge}$ ratio reflects the variation in the fraction of bulge stars which formed quiescently in discs. Mergers at higher redshift, when discs are more gas-rich and have fewer stars, deposit a lower fraction of (quiescently-formed) disc stars in the bulge (e.g. Croton 2006).

Close inspection of Fig. 4.1(c) shows that a few objects at the highest redshifts have black hole masses that exceed $f_{\rm BH} \times M_{\rm bulge}$. This would appear to be impossible given our definition of $f_{\rm BH}$ in §2.2.3. This apparent anomaly is due to our assumption that the mass of the black hole increases instantaneously at the time of the starburst. In our model, star formation in the bursts extends over ~ 50 dynamical times, while quasars shine only over ~ 0.3 dynamical times, so that the stellar mass builds up much more slowly than the black hole mass. It seems likely, however, that a black hole will still be growing towards its final mass towards the end of the starburst (Archibald et al. 2002; Alexander et al. 2005b; Borys et al. 2005). We defer a study of the co-evolution of the stars and the the black hole mass to future work.

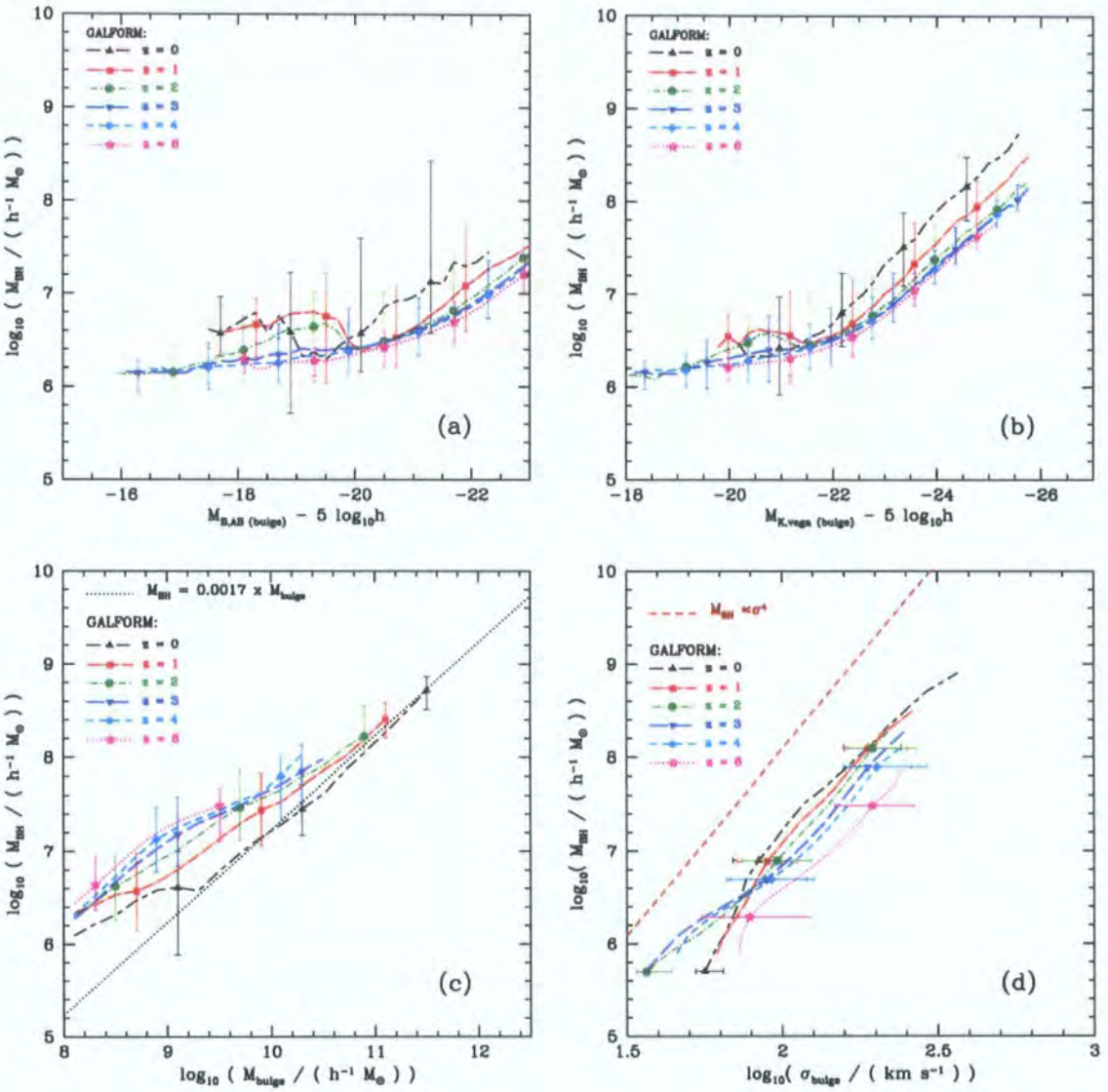

Figure 4.2: The redshift evolution of the relations between central black hole mass, $M_{\rm BH}$, and bulge properties, for galaxies selected to be recent bursts (i.e. galaxies which have had a starburst less than $10^8 yr$ before the redshift of observation). Each panel shows the relationship between $M_{\rm BH}$ and a different property of the host spheroid : (a) the bulge rest-frame B-band magnitude; (b) the bulge rest-frame K-band magnitude; (c) the stellar mass of the bulge; (d) the velocity dispersion of the bulge. The model predictions are shown by the symbols with errorbars; the lines show the median relations and the errorbars the 10–90 percentile spread of the distributions. Redshifts 0, 1, 2, 3, 4 and 6 are shown in different line types, as indicated by the key.

Finally, we show the $M_{\rm BH} - \sigma_{\rm bulge}$ relation in Fig. 4.1(d). There is no evolution in the slope of the relation, but the zeropoint does evolve and the scatter increases significantly towards higher redshift. For a given mass of black hole, the velocity dispersion of the bulge is greater at higher redshift. To some extent, this evolution reflects the expected variation in the properties of dark matter haloes: at a given mass, the halo velocity dispersions scales as $\sigma \propto (z_{\rm form} + 1)^{1/2}$. Alternatively, the evolution could be viewed as a reduction in the black hole mass with increasing redshift, for a fixed bulge velocity dispersion.

Shields et al. (2003) have compared the relative amounts of black hole mass in distant quasars and in galaxies in the local universe. They find a large scatter and an increase of $0 - 0.5$ dex in $M_{\rm BH}/\sigma_{\rm bulge}$ between $z = 0$ and $z \sim 3$. Similarly, Woo et al. (2006) have compared Seyferts at $z = 0.36$ with galaxies at $z = 0$. They too find an increase, of 0.62 ± 0.10 dex, in black hole mass at fixed $\sigma_{\rm bulge}$ at $z = 0.36$ compared to $z = 0$. Thus, the observed trend in $M_{\rm BH}/\sigma_{\rm bulge}$, if any, is in the opposite direction to the trend we find in our simulations. It is possible that our model neglects effects that would cause black holes to be a larger fraction of the galactic bulge mass at higher redshifts. However, it must be remembered that $\sigma_{\rm bulge}$ is one of the more uncertain properties of the galaxies in our model and that dynamical effects which are not included could play a role in determining the properties of merger remnants (Dekel & Cox 2006; Robertson et al. 2006).

As mentioned earlier, the observational samples used to study the evolution of the $M_{\rm BH}$ – bulge relations are small, and any selection effects may be hard to characterize. There does appear to be some tension between the redshift evolution of black hole masses in our model and those in the data (in the sense that our black holes may not be massive enough at high redshift), and we briefly considered whether we could obtain a better match to the data by considering only recent starbursts. At high redshifts, it is only possible to obtain observational black hole masses for galaxies with active nuclei, and it may be that galaxies with active nuclei (which will be recent bursts in our model) have black holes biased in mass with respect to the general galaxy population. In Fig. 4.2, we show the $M_{\rm BH}$ – bulge relations for galaxies selected to have undergone a starburst less than 10^8yr before the redshift of observation. It is quickly apparent that we are not able to obtain a better fit to the observational data by selecting recent bursts. In future work, we intend to improve our modelling of the evolution of AGN luminosities and spectra *during* starbursts, so that we are able to make more realistic predictions of which objects would be selected as quasars or AGN, although it does seem likely that this alone will not improve our match to the observed evolution in the $M_{\rm BH}$ – bulge relations.

4.2 Downsizing in a hierarchical universe

In cosmology, 'downsizing' is an ill-defined term which has been applied to describe the phenomenon whereby luminous activity (e.g. star formation or accretion onto black holes) appears to be occuring predominantly in progressively lower mass objects (galaxies or BHs) as the redshift decreases. Claims of downsizing were first made in connection with the population of star-forming galaxies (Cowie et al. 1996). More recently, the same trend has been inferred from the evolution of the X-ray luminosity function of quasars (Cowie et al. 2003; Steffen et al. 2003; Ueda et al. 2003; Barger et al. 2005; Hasinger et al. 2005): the number of bright X-ray sources peaks at a higher redshift than the number of faint X-ray sources. The optical quasar luminosity function shows similar evolution, with more bright objects seen at increasing redshifts (e.g. Croom et al. 2004).

The apparent downsizing in the quasar X-ray luminosity function has been interpreted by some authors as implying that black holes acquire mass in an '*anti-hierarchical*' manner (Marconi et al. 2004; Merloni 2004; Shankar et al. 2004; Hasinger et al. 2005). In this Section, we demonstrate that the 'downsizing' of the *luminous* growth of black holes is actually a natural feature of our model, despite the fact that the overall assembly of mass into black holes *is hierarchical.* Downsizing in the galaxy population in hierarchical models is promoted by the earlier collapse and more active merging of objects in regions of high overdensity (Kauffmann 1995; Mouri & Taniguchi 2006; Neistein et al. 2006). In recent models of galaxy formation, this natural trend is accentuated by the feedback processes associated with AGN activity in massive haloes (Bower et al. 2006; Croton et al. 2006). However, we wish to emphasize that AGN activity in low redshift cooling flows is very far from being the only ingredient required for downsizing, and that we still find downsizing in our model. We now review some of the indirect evidence already presented in support of this conclusion (§4.2.1), and go on to present explicit predictions which reveal which black holes in our model are accreting mass most rapidly (§4.2.2).

4.2.1 Indirect evidence for downsizing in the model: the evolution of the optical luminosity function

The optical quasar luminosity function, as we have already remarked, reveals a dramatic increase in the space density of bright quasars with increasing redshift (Croom et al. 2004). In §2.3.2, we presented the model predictions for the optical luminosity function, which are in good agreement with this trend in the observations. Two features of our

model are responsible for this success: the increase in the halo merger rate (and hence the galaxy merger rate) with increasing redshift and the increase in the gas content of discs with increasing redshift (see also Kauffmann & Haehnelt 2000). In combination, these phenomena lead to an increase in the frequency and strength of starbursts with redshift. In our model, a starburst results in a *luminous* phase of growth of the supermassive black hole; a fraction $f_{\rm BH}$ of the cold gas which is turned into stars during the burst is accreted onto the black hole.

Galaxy mergers are still an important way of building black hole mass at low redshift. Our model predicts that BH-BH mergers are the most important channel for building black hole mass for the most massive black holes at the present day. This *dark* growth process represents the assembly of mass which is already locked up in black holes into larger units. Galaxy mergers at low redshift tend to be gas poor in our model simply because more time has elapsed to allow galactic discs to turn cold gas into stars quiescently. This effect is accentuated in the case of the most massive black holes which tend to reside in the more massive dark haloes. The process of galaxy formation starts earlier in the progenitors of massive haloes, since these objects collapse into bound structures earlier than is the case in less extreme environments.

4.2.2 Direct evidence for downsizing in the model: which black holes are accreting mass?

Our model allows us to separate the mass assembly of black holes into two contributions: accretion, in which cold gas is turned into black hole mass in a starburst and mergers, in which existing black holes merge to build a more massive black hole. Here we focus on the process of gas accretion. Fig. 4.3 presents two views showing which mass of black holes are accreting material the most vigorously. The left-hand panels of Fig. 4.3 show the distribution of accretion rates, expressed in units of the Eddington mass accretion rate. Since the Eddington accretion rate scales with mass, this is easily scaled to give the distribution of fractional accretion rates. The right-hand panels compare the present accretion rate to the past average accretion rate (calculated as $< \dot{M} > \times t_{\rm age} / < M_{\rm BH} >$), as a function of black hole mass. Each row corresponds to a different redshift (top: $z = 0$, middle: $z = 1$, bottom $z = 2$). In these plots, we have *not* limited the mass accretion rate to be less than or equal to the Eddington limit.

The left-hand side of Fig. 4.3 shows that, at all redshifts, there is a large spread in the Eddington ratios at which black holes are accreting. There is variation amongst mergers

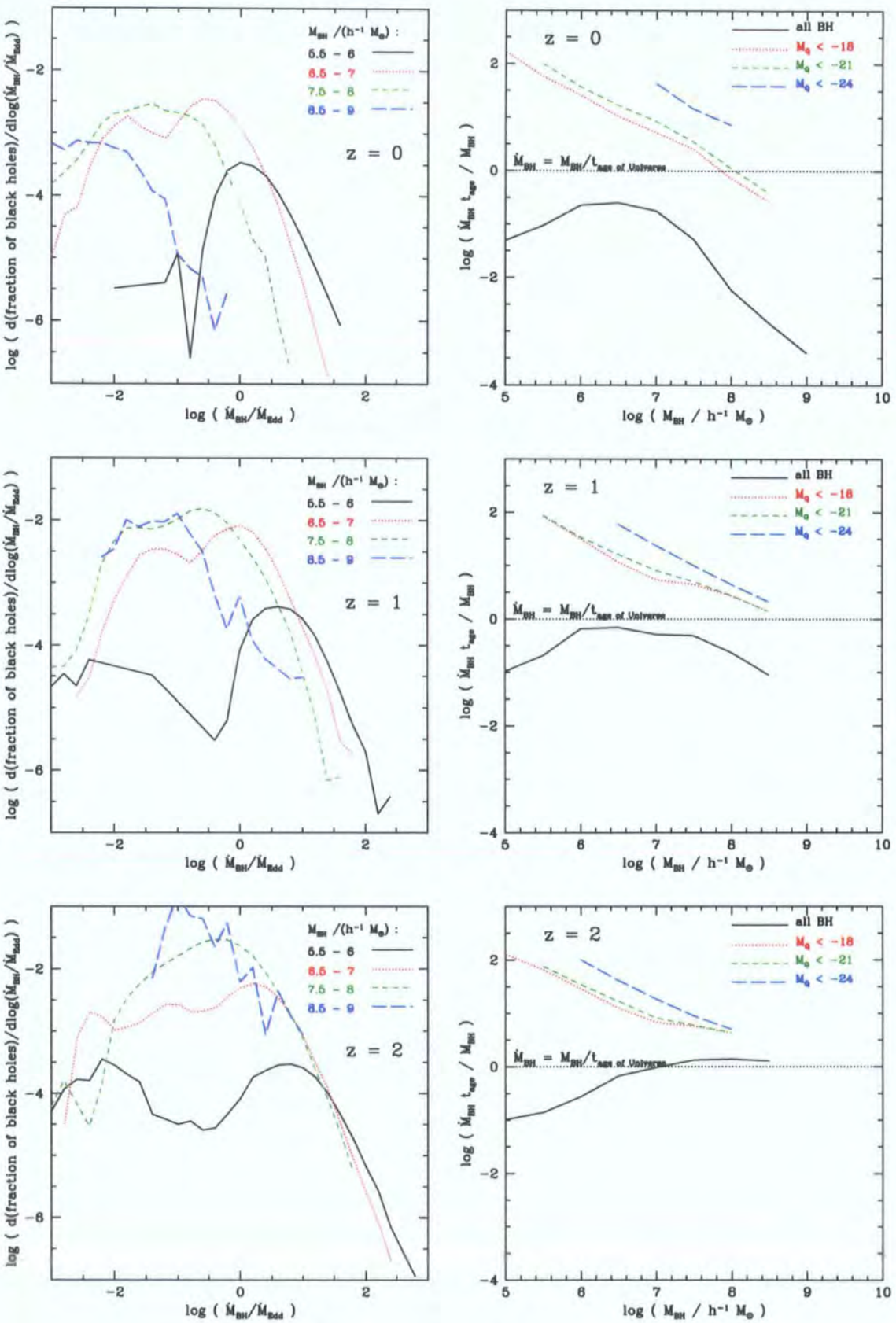

Figure 4.3: Left panels: The distribution of accretion rate normalized by the Eddington mass accretion rate. Right panels: the current mean mass accretion rate normalized by the past average mass growth rate, plotted against black hole mass. A more extensive description of the plots is in the text.

in gas supply, accretion timescale and initial black hole mass. Furthermore, the Eddington ratio evolves during any single accretion event. As expected, the mass accretion shifts towards higher fractions of the Eddington limit at higher redshift since there is more gas available in mergers. At $z = 0$, we see that as black hole mass increases, accretion shifts to lower fractions of the Eddington ratio. This trend is less pronounced at $z = 1$ and practically disappears by $z = 2$. Thus, more massive black holes were accreting mass more rapidly at $z = 2$ than they are today. The predicted distribution of mass accretion rates at $z = 0$ agree reasonably well with the observational results of Heckman et al. (2004). The distribution is normalized to unit area, but most black holes, particularly at lower redshifts, are not accreting at all (i.e. they are in a δ-function at $\dot{M} = 0$).

The right-hand column of Fig. 4.3 shows the ratio of the present accretion rate to the past average accretion rate ($< \dot{M} > \times t_{\rm age\ of\ Universe} / < M_{\rm BH} >$) as a function of black hole mass at $z = 0$, 1 and 2. If this ratio exceeds unity, then the current mass accretion rate exceeds the average rate at which the black hole gained mass in the past (summed over all progenitors). The predictions for this ratio are sensitive to the black hole selection, for example, selection using a cut in quasar luminosity. We show results for all black holes (solid lines) and also for black holes selected as quasars brighter than a given luminosity. Note that the bulk of black holes in the model are not accreting material at any given time. The solid lines in each panel show that there is clear evidence for downsizing in the model. At $z = 0$, more massive black holes are growing less rapidly than less massive black holes. By $z = 1$, this trend is greatly diminished, and at $z = 2$ it is reversed, i.e. the most massive black holes have the highest fractional accretion rate. When only quasars are selected, those with low mass black holes show similar fractional accretion rates as redshift varies from 0 to 2, while quasars with massive black holes show a strong decline in fractional accretion rate towards the present day.

4.3 Possible improvements to our model

In §4.1, we found indications that, for any given galaxy property, the black holes in our model may be somewhat undermassive at high redshifts. This appears to be a generic problem with hierarchical galaxy formation models – Bromley et al. (2004) were the first to demonstrate this explicitly – they fixed their BH model to $z = 6$ BH data and found that in a realistic hierarchical galaxy formation model which produced sufficiently massive black holes at $z = 6$, the black holes at $z = 0$ are too massive. Since we fix our model to

$z = 0$ galaxy and BH properties, we have the converse problem – for a model which has the appropriate feedback to reproduce $z = 0$ galaxy properties and a BH model which matches $z = 0$ BH properties, the BH masses at $z = 6$ are too small. We suggest some improvements to our model which may help to alleviate this problem:

• Black hole growth through starbursts fueled by disc instabilities may be required (e.g. Croton et al. 2006). Disc instabilities occur more at high redshift, where disc scale lengths are lower (allowing less rotational support) and discs are less likely to be stabilised by bulges (Mo et al. 1998). In Chapter 6, we consider a galaxy formation model developed by Bower et al. (2006), which includes this channel for black hole accretion.

• Black holes may lose significant mass through mergers over cosmic time (e.g. mass loss to gravitational waves, expulsion of black holes from galactic nuclei through 3-body interactions or expulsion from the galactic potential by gravitational wave kick). However, as shown in Fig.3.7, a much higher proportion of the mass of larger mass black holes is formed through mergers than smaller mass black holes. Therefore if any of these mechanisms are of sufficient magnitude to allow significant mass loss between high and low redshift, it is likely that the slope of the $M_{\rm BH}$ – bulge relations could become too shallow for $M_{\rm BH} > 10^8 h^{-1} M_\odot$ to fit observations.

• A model where black hole growth is limited more by self-regulated feedback (e.g. Silk & Rees 1998) than by mass supply and is thus more strongly related to the depth of the potential well. For a given mass of dark matter halo, potential wells are deeper at higher redshift, so this may help to bias black hole growth to larger redshift (e.g. Wyithe & Loeb 2003).

• We must still add AGN feedback to the model, particularly from cooling flows in quasi-hydrostatically cooling haloes, and this is likely to improve our predictions of downsizing (Croton et al. 2006; Bower et al. 2006). When gas falls into more massive haloes, it does not have time to cool before it reaches the halo centre, and forms an atmosphere of hot gas in quasihydrostatic equilibrium. Since haloes are more likely to be in this regime at higher halo masses snd at lower redshift, then introducing this feedback might allow us to retune our star formation feedback to be less extreme at higher redshifts whilst still reproducing the $z = 0$ galaxy luminosity function. This may allow there to be higher cold gas masses in mergers at high redshift, allowing us to shift black hole masses upwards at high redshift, and perhaps to shift formation of the most massive black holes towards high redshift. Currently, we need very strong star formation feedback in the Baugh et al. (2005) model to fit $z = 0$ galaxy properties. The Bower et al. model, which

we consider in Chapter 6, does include AGN feedback, although we have not yet explored all the possible retunings of other forms of feedback which the inclusion of AGN feedback allows.

• Radiation drag from radiation fields may be vital in allowing the loss of angular momentum and thus to black hole accretion (Kawakatu & Umemura 2002). Star formation occurs faster at higher redshift (both because the burst mode dominates at higher redshift, and because star formation timescales may be shorter at higher redshift), and this could lead to increased black hole growth at higher redshift. Star formation may occur particularly fast in larger starbursting galaxies at higher redshift, leading to even greater growth of larger black holes than smaller black holes (e.g. Granato et al. 2004). Furthermore, the decreased scale sizes of galaxies at high redshift would increase the radiation field density for any given star formation rate.

• Black hole seeds of mass $\sim 10^6 h^{-1} M_\odot$ may be produced in metal-free haloes with virial temperature greater than 10000 K, if H_2 cooling is supressed (Bromm & Loeb 2003; Volonteri & Rees 2005a). The gas is unable to cool to far below the virial temperature of the halo, which prevents fragmentation into stars, allowing direct collapse into a supermassive black hole (Bromm & Loeb 2003), or allows the formation of a fat accretion disc, allowing quasi-sperical super-Eddington accretion (Volonteri & Rees 2005a). The presence of these seeds could accelerate black hole growth in the most highly biased regions of the Universe at high redshift. We hope to include a simple model for black hole seeding at high redshift in future work.

Chapter 5

The nature of $z \sim 6$ quasars in hierarchical cosmologies

5.1 Introduction

The discovery of quasars at $z > 6$ (Fan et al. 2001a) has caused a great deal of excitement in the astronomy community. In particular, it is possible to infer the state of the intergalactic medium (IGM) from the spectra of these objects. Hydrogen in the Universe is neutral after recombination at $z \sim 1100$ but ionized by $z = 0$, due to energetic photons emitted by galaxies and quasars. Neutral hydrogen strongly absorbs photons at the wavelength of Lyman-alpha (1216Å), while ionized hydrogen does not, and the absorption of redshifted Lyman-alpha indicates that there is neutral hydrogen along the line of sight to the quasar. Small clouds will be seen as narrow absorption lines (the 'Lyman-alpha forest') while large regions of neutral hydrogen still present before the Universe is reionized will be seen as a Gunn-Peterson trough (Gunn & Peterson 1965). In recent years, as observations of quasars have been extended to $z > 5.8$, Gunn-Peterson troughs have been observed for the first time, indicating that the epoch of reionization is coming to an end around $z \sim 6$ (Becker et al. 2001; Fan et al. 2002; Songaila 2004; Fan et al. 2006a,b).

Apart from being useful probes of the reionization history of the Universe, high redshift quasars are very interesting objects in themselves. At $z \sim 6$, the Universe is only ~ 1 Gyr old, which is not very long compared to the Salpeter time (the timescale of mass accretion assuming that the black hole mass grows exponentially, at the Eddington rate – see Introduction for full definition). Therefore, the number of e-folds of Eddington-limited growth possible before $z \sim 6$, assuming the most favorable conditions (i.e. constant accretion at the Eddington rate over the whole of that period) is modest, around 25 for $\epsilon = 0.1$. Growth may happen more rapidly than this, however, if there is a lower efficiency

of photon emission, ϵ. However, at $z \sim 6$, several authors claim that there are already black holes of mass a few times $10^9 M_\odot$, as we review in §5.3. If these black holes grew from the typical masses of pop III stellar remnants ($\sim 100 M_\odot$, Carr et al. 1984), then around 16 e-folds of mass increase are required, after the first generation of stars have formed. This would require almost continuous growth if this all occurs via accretion with a high efficiency of photon emission. Furthermore, observations of the molecular gas in the galaxies or protogalaxies surrounding high redshift quasars may give valuable clues regarding the earliest stages of galaxy formation (Walter et al. 2003). If high redshift quasars are hosted in the rarest density fluctuations, which collapse earliest, then it may also be possible to find out about galaxy formation before reionization by observing their environments (Willott et al. 2005).

In this Chapter, we study whether objects such as those found by Fan et al. (2001, 2004, 2006) can be reproduced in a ΛCDM model of hierarchical galaxy formation, and explore whether the abundance of such objects can provide a useful constraint on the basic cosmological parameters.

We first discuss the recent Fan et al. observations of luminous quasars (§5.2), the inferred black hole masses and their reliability (§5.3) and simple inferences regarding the masses of their dark matter halo hosts (§5.4). We then discuss existing theoretical work which aims to explain $z \sim 6$ quasars in hierarchical models of galaxy formation (§5.5). The principal advances on previous work which we present are outlined in (§5.6). We then discuss our calculation (§5.7) and the results of our work (§5.8). We summarise and conclude in §5.9.

5.2 Observations of high redshift quasars

The Sloan Digital Sky Survey (SDSS; York et al. 2000) covers 10,000 contiguous square degrees, with imaging in 5 bands (u,g,r,i and z) down to a depth of 23 magnitudes in g. The magnitudes and colours are then used to select objects which are likely to be quasars to target for spectroscopic observations (Fan et al. 2001b). To find quasars at $z > 5.7$ in the SDSS, Fan et al. (2001*a*,*b*, 2004, 2006) select i-band dropout objects (specifically, they require a colour $i - z > 2.2$, and also apply a magnitude cut in z.). Their best estimate of the space density of these objects (based on 12 quasars) is $2.2 \times 10^{-9} h^3 \mathrm{Mpc}^{-3}$ (Fan et al. 2004). They have recently expanded the colour-selected sample to 19 objects (Fan et al. 2006), but still require spectroscopic confirmation of the new objects (which is now

scheduled), and so have not yet recalculated the space density. We follow Fan et al., and define 'luminous quasars' in this Chapter to be quasars with an absolute magnitude at a rest frame wavelength of 1450Å brighter than $M_{1450} = -26.81$.

5.3 The masses of luminous quasars at $z \sim 6$

We now assess the published estimates of the masses of the black holes fuelling the luminous quasars discovered by Fan et al.

Fan et al. (2001a) argue, from the observed magnitudes ($M_{1450} < -26.81$), that these luminous quasars have very massive black holes, with masses around $10^9 M_\odot$. They derive these masses assuming that the quasars are accreting at the Eddington rate and that the spectral energy distribution (SED) matches the mean observed SED for radio-quiet quasars found by Elvis et al. (1994). There are several reasons why this may not necessarily be an accurate assumption. Firstly, and perhaps most importantly, Elvis et al. found a very wide variation in the spectra of individual objects, with a correspondingly large uncertainty in the appropriate bolometric correction. Furthermore, Elvis et al. selected only quasars with $M_V < 17$ (for the V-band, $\lambda_{\rm eff} \approx 5510$Å). The SDSS quasar samples are magnitude-limited in the i-band (Fan et al. 2001b; Richards et al. 2003), which is at a longer wavelength ($\lambda_{\rm eff} \approx 7460$Å), and so this survey has greater sensitivity to redder quasars (Richards et al. 2003). The tail of red quasars is due to dust-reddening (Hopkins et al. 2004), and these quasars tend to have different spectra from the less reddened quasars which will be selected in a survey which is magnitude limited in the blue.

The quasar found with the highest redshift so far in the Fan et al. samples ($z = 6.41$) has attracted a great deal of attention. Whilst there are some models which permit quasars to radiate at super-Eddington luminosities (Abramowicz 2005), these remain controversial. Other largely independent mass estimators support the mass derived using the Eddington limit argument, at least for this one quasar. Willott et al. (2003) estimate a mass of $3 \times 10^9 M_\odot$ for the $z = 6.41$ quasar using the line-width of MgII. The width of this line is equated to the velocity of the gas clouds responsible for its emission. The virial argument ($M_{\rm BH} = r \times v^2/G$) is then used to estimate the black hole mass, using the calibration of McLure & Dunlop (2002). McLure & Dunlop estimate the distance from the black hole of the gas emitting the broad MgII line (r) using the radius-luminosity relationship of Kaspi et al. (2000), which is determined from local quasars using reverberation

mapping (Netzer & Peterson 1997). The mass derived in fact varies as $L^{0.47}$, so does have some dependence on the intrinsic luminosity of the quasar, but less so than the variation as $M_{\rm BH} \propto L$ which applies when using the Eddington argument. Note that, in principle, the intrinsic luminosity could be affected by gravitational lensing, as discussed below). Note, however, that there is no assumption regarding whether the luminosity is equal to the Eddington luminosity in the estimate by Willott et al. (2003). Barth et al. (2003) use a higher resolution spectrum of the quasar to estimate the mass of the quasar using the virial argument. Using the McLure & Dunlop (2002) calibration for the MgII line they derive a mass of $2 \times 10^9 M_\odot$, and using the Vestergaard (2002) calibration for the CIV line, they derive a mass of $6 \times 10^9 M_\odot$. Both estimates have an uncertainty of a factor of 2.5–3, due to the intrinsic scatter in the McLure & Dunlop (2002) and Vestergaard (2002) scaling relations.

It may be that the brightest observed quasars are gravitationally lensed. If a quasar is lensed, its intrinsic luminosity would be less than the luminosity inferred observationally without taking the lensing magnification into account. Upon correcting for lensing amplification, the Eddington argument would imply black hole masses lower than a few $\times 10^9 M_\odot$. The steeper the slope of the quasar luminosity function, the higher the probability that a quasar in any given luminosity bin will have been scattered from a bin of lower intrinsic luminosity. The probability that many quasars selected in a magnitude limited survey will be magnified by gravitational lensing will be correspondingly higher (Comerford et al. 2002). If gravitational lensing is important, then we may expect to see multiple images of quasars, which are a signature of strongly lensed objects. Most highly magnified objects have multiple images, but in some cases (Keeton et al. 2005), an object may have a single, highly magnified image (this usually involves lensing by a cluster-sized halo, of mass $\gtrsim 10^{13.5} M_\odot$), or may have counter-images too faint to be detectable (this is more likely to occur with lower mass lenses ($\lesssim 10^{12} M_\odot$) and is very dependent upon the inner profile of the lensing halo). Richards et al. (2004) imaged 4 of the Fan et al. (2001a) $z > 5.7$ quasars, and found no multiple images, implying that the luminosities of the Fan et al. (2001a) are not significantly boosted by gravitational lensing. Even considering cases where strong lensing does not lead to multiple images, Keeton et al. (2005) find that the probability that a quasar which is not observed to have a multiple image to be strongly lensed is only 9%-29%, which gives an overall probability of $\sim 0.1^4$ that all four objects observed by Richards et al. are strongly lensed. Therefore, it is very unlikely that the extreme luminosities of the Fan et al. (2001a) objects can be explained, in general,

by strong gravitational lensing.

5.4 The dark matter halo hosts of luminous quasars

Fan et al. (2004) compared the space density of luminous quasars to the space density of dark matter haloes more massive than some threshold mass at $z = 6$. Their implicit assumption is that *all* dark matter haloes more massive than the threshold mass host one luminous quasar. The duty-cycle is the fraction of time for which the black hole is active and observed as a quasar. Fan et al. (2004) assume that the duty-cycle of quasars is 1, i.e. that a central supermassive black hole in a dark matter halo will always be seen as a quasar. With these assumptions, their conclusion is that luminous quasars are hosted in dark matter haloes of mass greater than $M_{\rm halo} \sim 10^{13} M_{\odot}$. However, given their assumptions, $\sim 10^{13} M_{\odot}$ should be regarded simply as an upper limit to the minimum halo mass required to host a quasar more luminous than $M_{1450} = -26.81$. The duty-cycle of quasars is likely to be less than one, particularly if the accretion of gas onto black holes is driven by mergers, which are a somewhat intermittent process. Furthermore, the $M_{\rm BH} - M_{\rm halo}$ relation may have significant scatter. This is particularly the case in a model where black hole accretion is highly dependent on galaxy mergers; for any given mass of halo this process may depend strongly upon its merging history.

Willott et al. (2005) attempted to measure the mass of the dark matter halo hosting the Fan et al. (2001a) quasar at $z = 6.41$. They intended to do this by observing galaxies surrounding the quasar, then measuring the velocity dispersion of the galaxies which appear to be satellite galaxies. However, they were unable to observe any galaxies, despite expecting to see galaxies if the quasar were indeed in a dark matter halo of mass $M_{\rm halo} \sim 10^{13} M_{\odot}$. Their inference was that there may be a significant scatter in the $L_{\rm Q} - M_{\rm halo}$ relationship. Since the dark matter halo mass function is extremely steep at halo masses $\gtrsim 10^{13} M_{\odot}$ at $z \sim 6$, the typical masses of haloes hosting quasars could be significantly smaller than might be expected given a $L_{\rm Q} - M_{\rm halo}$ relation without any scatter.

5.4.1 Bias and clustering – the 'HOD' model

We described in the introduction (Chapter 1) how the evolution of dark matter in the Universe can be approximated by a model of dark matter haloes which form and merge. To a first approximation, dark matter halo merger histories are independent of the large scale environment (Lemson & Kauffmann 1999). If we further assume that the evolution

of galaxies within a dark matter halo is not affected by environmental effects arising from outside that halo (e.g. inhomogeneous reionization, stellar or AGN winds from neighbouring dark matter haloes), then we have a 'halo model of galaxy formation'. In a such a model, the spatial distribution of some object (galaxies brighter than some magnitude, quasars more luminous than some flux limit etc.) is completely described by the 'halo occupation distribution' or HOD (Benson et al. 2000; Peacock & Smith 2000; Seljak 2000; Berlind & Weinberg 2002; Cooray & Sheth 2002), in combination with the distribution of dark matter haloes and the density profile of dark matter haloes, both of which can be obtained from a simulation (e.g. Sheth et al. (2001); Springel et al. (2005b) in the case of halo distribution, and Navarro et al. (1997); Power et al. (2003) in the case of density profile).

The concept of bias was introduced by Kaiser (1984), in the context of explaining why Abell clusters, which are hosted by the most massive dark matter haloes, are significantly more clustered than would be predicted soley from the clustering of dark matter in N-body simulations. Specifically, the square of the bias of a given mass of dark matter halo is the factor by which the amplitude of their correlation function is increased relative to that of the overall dark matter clustering amplitude, and it is higher for more massive dark matter haloes. To a reasonable approximation, the clustering of dark matter haloes depends solely on their mass (Cole & Kaiser 1989; Mo & White 1996; Sheth et al. 2001). To second order, it also depends on other properties, such as concentration, spin, formation time and environment (Gao et al. 2005a; Wechsler et al. 2006; Bett et al. 2007; Harker et al. 2006). Once the halo occupation distribution of a class of objects (in our case, quasars) is known, a fairly accurate overall bias factor for those objects is calculated from the weighted mean of the bias factors of their matter halo hosts. In this way, the clustering amplitude of those objects can be predicted analytically, without recourse to N-body simulations. This is the procedure we adopt in this Chapter for quasars and black holes, in anticipation of future measurements of quasar clustering at $z \sim 6$. In future work, we would also like to predict the quasar-galaxy cross-correlation function (e.g. Kauffmann & Haehnelt 2002), since, due to the low space density of $z \sim 6$ quasars, this may be more viable observationally, but for now, we restrict ourselves to a few illustrations of the environments of the most massive $z \sim 6$ black holes.

5.5 Black holes and quasars at $z \sim 6$ in hierarchical models

There has been much theoretical work recently which attempts to explain the highest redshift quasars within the current paradigm of hierarchical structure formation. Using extended Press-Schechter theory merger trees (which we discussed in §1.4.2), Volonteri & Rees (2006) have studied the growth of black holes in $10^{13} M_{\odot}$ dark matter haloes. Their choice to consider only this mass was motivated by the Fan et al. (2001a) estimate of the dark matter halo mass which hosts luminous quasars, as described in the previous section (§5.4). Volonteri & Rees have a very crude model for galaxy formation (they basically consider dark matter haloes to host only one galaxy per halo, with a simple estimate of the galaxy mass), but a very sophisticated model for black hole evolution. Some of the most important processes which they consider and which we do not are the following:

- Black hole seeds. The specific model they use is described in Volonteri & Rees (2005b). They posit a phase of rapid accretion onto pop III stellar remnants in metal free haloes with virial temperature $T_{\rm vir} > 10^4$K to form black hole seeds of masses $\sim 10^5 - 10^6 M_{\odot}$.

- The spin-up of black holes during gas accretion and the spin interaction and evolution during black hole mergers.

- The limiting of the black hole mass accretion rate due to the Eddington limit, and the effect upon this due to black hole spin (highly spinning black holes accrete at a higher efficiency of photon emission, and thus the Eddington accretion rate $\dot{M}_{\rm BH,Edd}$ will be lower, as $\dot{M}_{\rm BH,Edd} \propto 1/\epsilon$).

- The efficiency of BH-BH merging after dark matter halo merging – the orbital evolution of black holes within dark matter haloes and the resulting BH-BH separation are tracked. After a halo-halo merger, it takes time for black holes to sink to the centre of the new halo, due to dynamical friction against the dark matter halo, stars or gas. In comparison, we consider black holes to merge when their host galaxies merge, which is on a longer timescale than the halo merger. However, the BH-BH merger timescale could be larger still, as the orbits of black holes must shrink down to an even smaller radius.

- The gravitational recoil of black hole merger remnants after the merger of two black holes of different masses. The black hole escapes from the dark matter halo if its velocity is greater than the escape velocity. Otherwise, its orbit within the dark matter halo is tracked, including the effects of dynamical friction.

Some of these processes and differences in modelling will enhance the growth of high

redshift black holes relative to our model, whilst others will slow it down. Assessing the net impact of these differences is beyond the scope of this Thesis.

Springel et al. (2005b) performed an N-body simulation called the 'Millennium Run'. This is a state of the art simulation, and has been used for a large number of projects which had not previously been possible. It has a volume of $1.25 \times 10^8 h^{-3}\mathrm{Mpc}^3$ and uses $\sim 10^{10}$ particles to represent the dark matter, each of which represents a mass of $8.6 \times 10^9 h^{-1} M_\odot$. Assuming that dark matter haloes are sufficiently resolved by 20 particles, it resolves dark matter haloes down to a mass of $1.72 \times 10^{10} h^{-1} M_\odot$. Springel et al. (2005b) attempted to identify the dark matter haloes in the simulation which are likely to the black holes fuelling luminous quasars. They used a simple galaxy formation model to attach galaxies to the dark matter, and in particular they assigned these galaxies a stellar mass, a star formation rate and a black hole mass. The 10 candidate galaxies which they suggest are most likely to host luminous quasars have central black holes masses in the range $10^8 - 10^9 M_\odot$. In fact, Springel et al. (2005b) do not specifically select the most massive black holes in their simulation, but select candidates with either the most massive dark matter halo or the most massive host galaxy stellar mass. Either method used results in a largely overlapping set of objects (8/10 objects are identical), and it is likely that the same is true if black hole mass is considered instead. However, when they select the 10 objects with the highest star formation rate, there is less overlap (4/10 objects overlap with the 8 objects in common between the other two samples). It is likely that quasar activity is more correlated with star formation rate than with stellar mass; at least, this is the case in a model such as that of Malbon et al. (2006) where quasar accretion is triggered by starbursts. Therefore, it is not necessarily true that Springel et al. are correct in their approach of identifying the brightest quasars with the largest dark matter haloes.

To attain luminosities brighter than $M_{1450} = -26.81$, black holes with masses in the range found by Springel et al. must accrete with efficiencies significantly greater than the canonical value of $\epsilon = 0.1$, and Springel et al. suggest that this may well be the case, in which case the most luminous quasars in the simulation would be analagous to those found by Fan et al. (2001a, 2004). This seems to be a rather optimistic conclusion. Probably the most serious problem with the claim of Springel et al. is the small volume which they have simulated. Given a volume of $1.25 \times 10^8 h^{-3}\mathrm{Mpc}^3$, around 0.3 quasars of the luminosity observed by Fan et al. would be expected given the observationally inferred space density. The 10 candidates proposed by Springel et al. all have black hole masses around an order of magnitude lower than that claimed by Fan et al. , and these

objects probably correspond to quasars slightly fainter and more numerous than those found in the Fan et al. samples. One interesting finding of Springel et al. is that the haloes they identify as hosting the most luminous quasars at $z = 6$ end up becoming part of the most massive haloes at $z = 0$. This is basically a consequence of their identification of the brightest quasars with the most massive dark matter haloes, and the tendency of the most massive haloes at high redshifts to end up in the most massive haloes at $z = 0$. We will explore this claim in §5.8.5.

Li et al. (2006) study the formation of a $z \sim 6$ quasar via hierarchical galaxy mergers in an N-body simulation. They overcome many of the shortcomings in the Springel et al. (2005b) work. They perform an N-body simulation with a very large volume ($10^9 h^{-3}\mathrm{Mpc}^3$), 8 times larger than the Millennium simulation. They select the highest mass dark matter halo at $z = 0$ to resimulate. It was not possible for them to select directly the most massive dark matter halo at $z = 6$ due to the poor mass resolution of their initial simulation, but it is very likely that one of the most massive $z = 6$ haloes, if not the most massive $z = 6$ halo, will be selected by this method. The chosen volume is then resimulated at much higher resolution (dark matter particle mass, $m_{\mathrm{dm}} \sim 2.8 \times 10^8 h^{-1} M_\odot$), sufficient to characterize the dark matter merger history of the most massive $z = 6$ dark matter halo in detail. They find that its merger tree consists of eight distinct progenitor haloes, which they assume, using simple models, to host eight galaxies. These galaxies engage in seven major mergers to form the final quasar host. Before galaxies enter the merger tree at $z \gtrsim 14$, one $200 M_\odot$ black hole seed is assumed to form in the centre of each galaxy at $z = 30$ due to pop III star formation. This is assumed to grow at the Eddington limit (assuming $\epsilon = 0.1$) until the galaxy enters the merger tree (they are rather unclear about what exactly they mean by the phrase 'enters the merger tree', and do not give details of their merger tree, or the redshifts at which galaxies 'entered' it). Each galaxy merger is followed using a high resolution SPH simulation, similar to those used in Springel et al. (2005a), and the black hole growth is tracked, taking into account black hole feedback.

Li et al. find that they can easily produce a black hole of mass $\sim 2 \times 10^9 M_\odot$ by $z = 6.5$ and a luminous quasar at $z \sim 6.5$, without requiring constant accretion at the Eddington rate or super-Eddington accretion. In practice, the frequent galaxy mergers allow almost constant accretion at the Eddington rate. It seems very likely that the high frequency of mergers (i.e. the time between mergers is not long compared to the timescale for any given merger) ensures that there is an almost continuous supply of gas available

for accretion, allowing essentially constant accretion at the Eddington rate throughout most e-folds of black hole growth (which occur whilst the black hole is not very massive, so that a low accretion rate is required for growth at the Eddington rate). It is only in the final few e-folds of black hole growth that the exact rate of gas supply becomes relevant. During these last few e-folds, the black hole spends periods accreting at a sub-Eddington rate, and the significant boosting of gas supply provided by mergers becomes relevant. As the black hole approaches its final mass (which is strongly correlated to the depth of the galactic potential well due to the black hole feedback model), it is largely feedback from the black hole itself which heats up gas and shuts off further accretion.

Li et al. have a sophisticated model for quasar luminosity. As they showed in Hopkins et al. (2005a), the accretion rate onto a black hole varies substantially during mergers, tracing the variation in the gas supply. This leads to very complicated quasar lightcurves, which do not have a simple form. The local column density of obscuring gas in the galaxy or galaxy merger along all lines of sight comes directly from the simulation, and using the methodology of Hopkins et al. (2005b), they are able to make sophisticated predictions of the quasar luminosity which will actually be observed. They calculate the colum density of obscuring gas and its metallicity. Using the Springel & Hernquist (2003) model for a two-phase interstellar medium, they calculate the column density of gas which will be in the hot phase, and assume that the probability that any cool phase gas is in the line of sight will be small, since it is condensed into compact clumps. They model the intrinsic quasar spectrum following the work of Marconi et al. (2004). They make a simple assumption for the gas-to-dust ratio, and calculate the dust extinction assuming the SMC reddening curve of Pei (1992). They do not include a full radiative transfer calculation, or absorption and re-radiation of light by dust.

5.6 The main advances we make upon existing work

In this Chapter, we present predictions for the abundance of $z = 6$ quasars which represent a number of improvements over previous work:

• We use grids of haloes with a fine spacing in mass (masses are separated by a factor of 1.06) covering all regions of the halo mass-function which are of interest. This is a significant improvement upon models such as Volonteri & Rees (2006) and Li et al. (2006), where only one halo or a small number of the most massive haloes ($M_{\rm halo} \sim 10^{13} M_{\odot}$) are considered.

• We use a much higher resolution in our dark matter merger trees than is available in the Springel et al. (2005b) work.

• Unlike Springel et al. (2005b), we calculate quasar luminosities for all objects, under different assumptions for the Eddington limiting of the luminosity, relating this to recent starbursts which the galaxy has undergone. Springel et al. rely almost exclusively on their calculation of black hole masses in their discussion of luminous quasars.

• In our semi-analytic simulations, the maximum halo mass probed is $3.29\times10^{13}h^{-1}M_{\odot}$ for ΛCDM1 and $1.51\times10^{13}h^{-1}M_{\odot}$ for ΛCDM2 (see next section for the definition of these cosmological models). In both cases, this is significantly larger than the most massive dark matter halo in the Millennium simulation (Springel et al. 2005b) of $5.1\times10^{12}h^{-1}M_{\odot}$. The space density of the most massive haloes we consider is $1.9\times10^{-14}h^{3}\mathrm{Mpc}^{-3}$, and we would require a volume 40000 times bigger than the Millennium simulation to find such a halo.

• When we do use an N-body simulation, for example to probe the environment of the most massive black holes and to track these through to the present day (§5.8.5), we use a much larger simulation (volume $2.4\times10^{9}h^{-3}\mathrm{Mpc}^{3}$) than the Millennium simulation (volume $1.25\times10^{8}h^{-3}\mathrm{Mpc}^{3}$). Taking the space density of luminous quasars measured by Fan et al. (2004), we would expect ~ 0.3 such objects in the Millennium simulation, and ~ 5 such objects in our N-body simulation of the ΛCDM1 cosmology.

5.7 Model

5.7.1 Cosmological model

Although there has been a dramatic improvement over the past decade in observational constraints on the basic parameters of the cosmological model, uncertainties still remain, some of which may have significant implications for our modelling of galaxy and quasar formation, particularly at high redshift. It may be that the abundance of high redshift quasars provides a useful test of the cosmological model. Here, we consider two cosmologies:

•ΛCDM1 – the cosmology based on the WMAP 1-year results and initial 2dFGRS $P(k)$ measurement (Spergel et al. 2003). This is the cosmology which was used in the 'Millennium' simulation (Springel et al. 2005b).

•ΛCDM2 – the cosmology based on the WMAP 1-year results and the final 2dFGRS $P(k)$ results (Sánchez et al. 2006). This model is also consistent with the best fit to

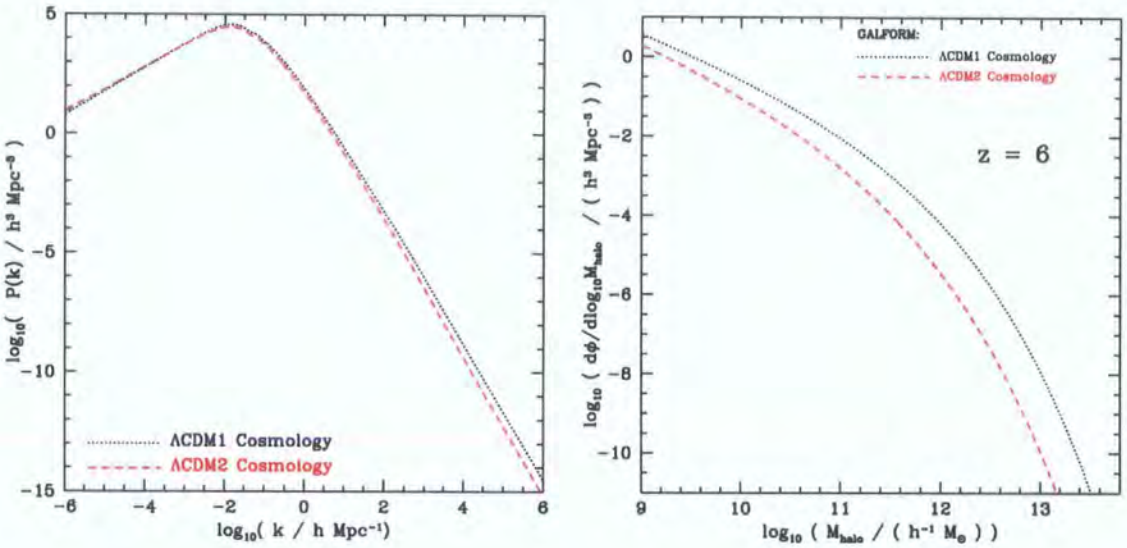

Figure 5.1: Left: The linear theory dark matter power spectrum at $z = 0$ in the ΛCDM1 and ΛCDM2 cosmologies. Right: The halo mass function at $z = 6$ in the ΛCDM1 and ΛCDM2 cosmologies.

WMAP3 + 2dFGRS $P(k)$ found by Spergel et al. (2006).

The cosmological parameters for ΛCDM1 and ΛCDM2 are summarised in table 5.1. For comparison, we also show in this table the corresponding values of the parameters in the 'Concordance Cosmology' which was used by Baugh et al. (2005) and Malbon et al. (2006) and in Chapters 2-4 of this Thesis. The change in the cosmological parameters from ΛCDM1 to ΛCDM2 is mainly driven by the improvement achieved in the 2dFGRS power spectrum on using the final dataset (Cole et al. 2005). The main evidence for tilt comes from the addition of the 2dFGRS power spectrum to the CMB data, although the CMB data weakly favours tilt (Sánchez et al. 2006).

The abundance of bright quasars could potentially cause problems for the ΛCDM2 cosmology. We show the power spectrums of density fluctuations for the two cosmologies in Fig. 5.1 (left), and the halo mass functions at $z = 6$ in Fig. 5.1 (right). The reduced power in the ΛCDM model on small scales leads to a reduction in the number of haloes of any given mass. This reduction is greater at higher redshifts, since structure formation is increasingly dominated by the collapse of smaller scales with increasing redshift. For haloes more massive than $10^{11.5} h^{-1} M_{\odot}$, there are approximately 10 times fewer haloes at $z = 6$ in the ΛCDM2 cosmology than in the ΛCDM1 cosmology, mainly because ΛCDM2 has a spectral tilt of $n_s < 1$ and a lower amplitude of fluctuations on scales of $8h^{-1}$Mpc, σ_8.

Table 5.1: In the top half of the table, we compare the cosmological parameters used in the ΛCDM1 and ΛCDM2 models which vary from those of the 'Concordance' cosmology used in Baugh et al. (2005) and Malbon et al. (2006), and in Chapters 2-4. $\Omega_{\rm m}$, Ω_Λ and $\Omega_{\rm b}$ are the cosmological matter density, the energy density of dark energy and the cosmological baryon density, all in units of the critical density. n_s is the spectral index, σ_8 is the linear *rms* fluctuation in spheres of radius $8h^{-1}$Mpc and h is the Hubble constant in units of $100{\rm kms}^{-1}{\rm Mpc}^{-1}$. In the bottom half of the table, we show the corresponding variations in the galaxy formation parameters, as discussed in §5.7.4. Where $v_{\rm cond}$ is defined, then we switch off cooling in haloes with circular velocities greater than $v_{\rm cond}$. $v_{\rm sw,disc}$ and $v_{\rm sw,burst}$ are circular velocities below which superwinds are most effective (the superwind parameterization is described quantitatively in Chapter 1). $f_{\rm gas,burst}$ and $F_{\rm BH}$ have the same definitions as in Chapter 2.

Cosmology	Concordance	ΛCDM1	ΛCDM2
$\Omega_{\rm m}$	0.3	0.25	0.237
Ω_Λ	0.7	0.75	0.763
$\Omega_{\rm b}$	0.04	0.045	0.041
n_s	1	1	0.954
σ_8	0.9	0.9	0.773
h	0.7	0.73	0.73
Galaxy/BH			
$v_{\rm cond}/{\rm kms}^{-1}$	–	200	–
$v_{\rm sw,disc}/{\rm kms}^{-1}$	200	300	300
$v_{\rm sw,burst}/{\rm kms}^{-1}$	200	300	300
$f_{\rm gas,burst}$	0.75	0.5	0.5
$F_{\rm BH}/10^{-2}$	2.2	2.3	2.5

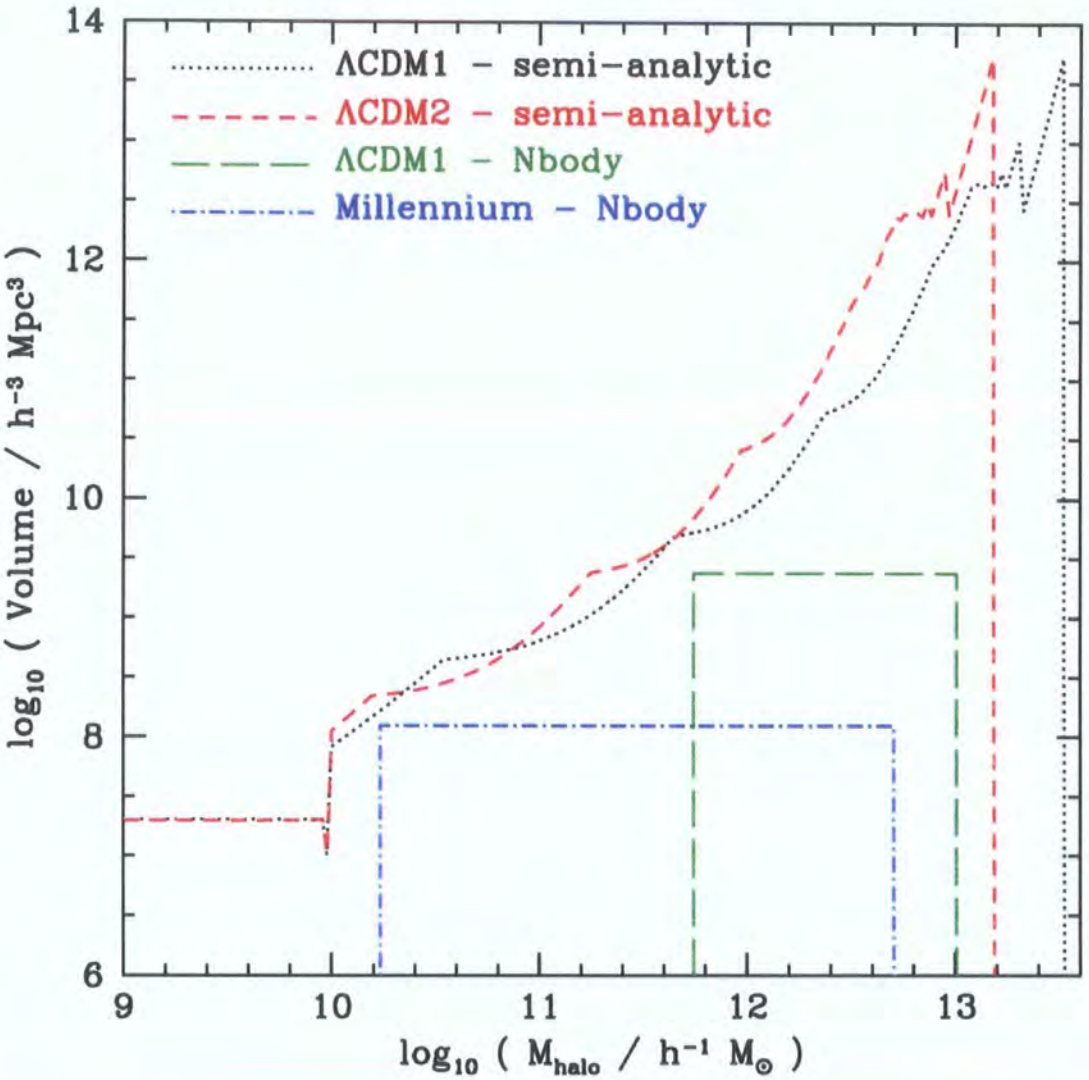

Figure 5.2: The volume probed by our semi-analytic merger trees as a function of halo mass, for the ΛCDM1 and ΛCDM2 cosmologies. We compare this to the volume probed by our N-body simulations of ΛCDM1, and the volume probed by the 'Millennium' run of Springel et al. (2005b). The N-body lines begin at the dark matter halo mass resolution limit of the N-body simulation and end at the maximum halo mass present in the simulation.

5.7.2 Semi-analytic dark matter merger trees

We use Monte-Carlo merger trees, which we generate using the extended Press-Schechter formalism (Bond et al. 1991; Bower 1991), in particular, the expressions for dark matter halo merger rates presented by Lacey & Cole (1993). The exact procedure is described in Cole et al. (2000). For the majority of the results presented in this Chapter, we use

grids of merger trees weighted by the halo mass function. We also attach Monte-Carlo merger trees of the appropriate mass to all of the haloes found in an N-body simulation (as in §5.7.3, §5.8.5). The use of Monte-Carlo merger trees allows us to probe halo masses much lower (here, at least an order of magnitude) than those in the highest resolution simulation available to us. Furthermore, using grids of Monte-Carlo trees, we are able to target large numbers of the most massive haloes, including haloes larger than those found in the N-body simulations, effectively allowing us to probe to a much higher volume. The effective volume probed for any given halo mass varies with the mass, as shown in Fig. 5.2. We have oversampled the largest halo masses where we expect to find interesting, rare objects such as bright quasars and massive black holes. We have not sampled lower mass haloes so heavily, but it will become apparent when we study the halo occupation distribution of quasars and black holes that we have adequately sampled all halo masses relevant to the objects in which we are interested. As shown in Fig. 5.2, the minimum halo mass for which we have generated trees is $10^9 h^{-1} M_{\odot}$, 600 times less massive than the minimum halo mass in our N-body simulations, and 17 times less massive than the minimum halo mass in the 'Millennium' simulation (Springel et al. 2005b).

When calculating statistics for the universal halo, galaxy and black hole populations, we have used the Jenkins et al. (2001) mass function to assign effective space densities to all objects we produce when generating our grid of dark matter merger trees. The Jenkins et al. mass function is an empirical fit to N-body simulations. It has also been shown specifically to be accurate at high redshifts ($z > 6$) including highly biased, massive dark matter haloes (Reed et al. 2003; Springel et al. 2005b), which is the regime we probe in this Chapter. Whilst it is possible to make accurate analytic estimates of the overall abundance of dark matter haloes at any given redshift, there are a number of differences between extended Press-Schechter theory merger histories and N-body merger trees when higher order statistics are considered. These differences have been highlighted and emphasized in recent papers which make use of high resolution N-body simulations (e.g. Springel et al. 2005b; Li et al. 2006). For example, for any given mass of dark matter halo, haloes which formed earlier tend to be more clustered (Gao et al. 2005a). Surprisingly perhaps, given the poor agreement of the Press-Schechter mass function in comparison to the Jenkins et al. mass function at high redshifts, the extended Press-Schechter merger histories in fact appear to agree very well with progenitor mass functions derived from N-body simulations (Gao et al. 2005b).

5.7.3 N-body simulations

We have performed an N-body simulation of the ΛCDM1 cosmology, in order to study the environments of $z \sim 6$ quasars. The N-body simulation has a volume of $(1340h^{-1}\mathrm{Mpc})^3 = 2.4\times10^9\mathrm{h}^{-3}\mathrm{Mpc}^3$, approximately 20 times larger than the Millennium simulation (Springel et al. 2005b). The particle mass is $5.59 \times 10^{10}h^{-1}M_\odot$. We require a minimum of 10 particles to be linked by the friends of friend groupfinder to qualify as a resolved halo, so our dark matter halo resolution limit is $5.59\times10^{11}h^{-1}M_\odot$. The N-body simulation and the group-finding algorithm are described in further detail in Angulo et al. (2006). We have constructed dark matter merger histories semi-analytically for each of the haloes found in the N-body simulation, using the extended Press-Schechter algorithm. Disk space, RAM limitations and limited resolution did not allow us to take halo merger histories directly from the simulation. Whilst the dark matter halo resolution limit in the simulation was fairly poor, we were able to achieve a very high mass resolution ($5 \times 10^8h^{-1}M_\odot$) in our dark matter halo merger histories, 30 times better than that available from the Millennium simulation merger trees.

5.7.4 Galaxy and black hole formation model

We use the formation galaxy model of Baugh et al. (2005). The parameters in Baugh et al. (2005) were set using the 'concordance cosmology' (see Table 5.1). When we replace the concordance cosmology with the ΛCDM1 or ΛCDM2 cosmology, we find that the fit of our model to the $z = 0$ galaxy data used to set the model parameters is altered. We make a small number of minor adjustments to the model parameters to solve this problem, which we summarize in Table 5.1.

We compare the predictions of the ΛCDM1 and ΛCDM2 versions of our galaxy formation model to our standard set of redshift zero galaxy data, as described in Chapter 1. The comparisons are shown in Figs. 5.3, 5.4 , 5.5 and 5.6. The fits are very good, and we are satisfied that our galaxy formation model, following these minor tunings, is appropriate irrespective of the cosmology. We have also reset the parameter F_{BH} for the two new models. We normalized the black hole masses by comparison with the observed $z = 0$ M_{BH} – bulge relations, as described in Chapter 2. With minor adjustments of F_{BH}, we find very similar results to those found in Chapter 2 for the original (the 'concordance cosmology', as used in Baugh et al. and Malbon et al.) and so we do not show them here.

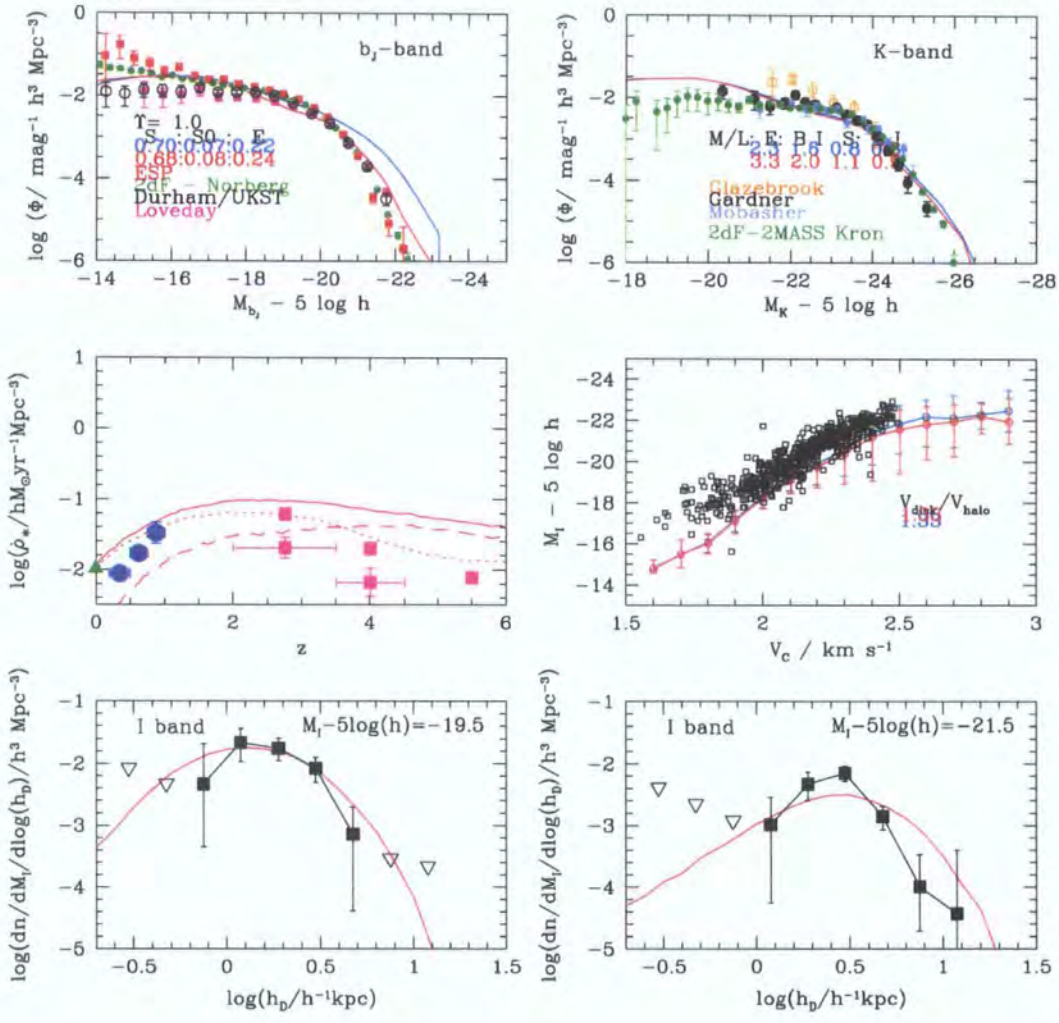

Figure 5.3: The predictions of the ΛCDM1 GALFORM model using our standard dust extinction model (red lines) and switching off dust extinction (blue lines). Where the blue line is not visible, the two results are identical. We also plot a variety of observational data – the observational data in each plot are described in Chapter 1. Top-left: The $z = 0$ b_J-band luminosity function. Top-right: The $z = 0$ K-band luminosity function. Middle-left: The star formation rate per unit volume as a function of redshift (solid – total; dotted – quiescent; dashed – burst). Middle-right: The I-band Tully-Fisher relation at $z = 0$. Bottom-left: The distribution of disc scale-lengths (for galaxies with $-20 < M_I < -19$). Bottom-right: The distribution of disc scale-lengths (for galaxies with $-22 < M_I < -21$).

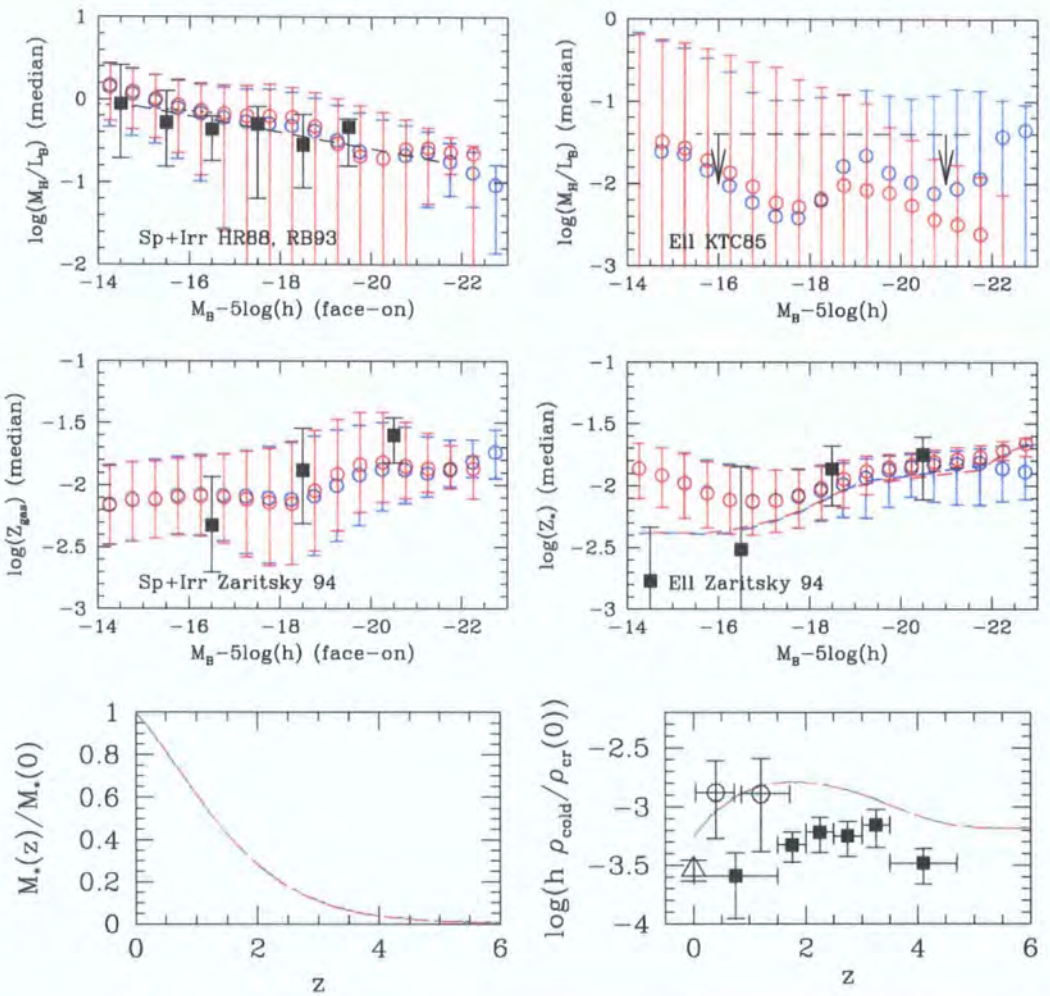

Figure 5.4: Further predictions of the ΛCDM1 GALFORM model using our standard dust extinction model (red) and switching off dust extinction (blue). We also plot a variety of observational data – the observational data in each plot are described in Chapter 1. Top-left: Gas mass to B-band luminosity ratio as a function of M_B for spirals and irregulars. Top-right: Gas mass to B-band luminosity ratio as a function of M_B for ellipticals. Middle-left: Gas metallicity of spirals and irregulars as a function of M_B at $z = 0$. Middle-right: Stellar metallicity of ellipticals as a function of M_B at $z = 0$. Bottom-left: The fraction of the $z = 0$ stellar mass formed as a function of redshift. Bottom-right: The redshift evolution of the universal density of cold gas (as a fraction of the critical density).

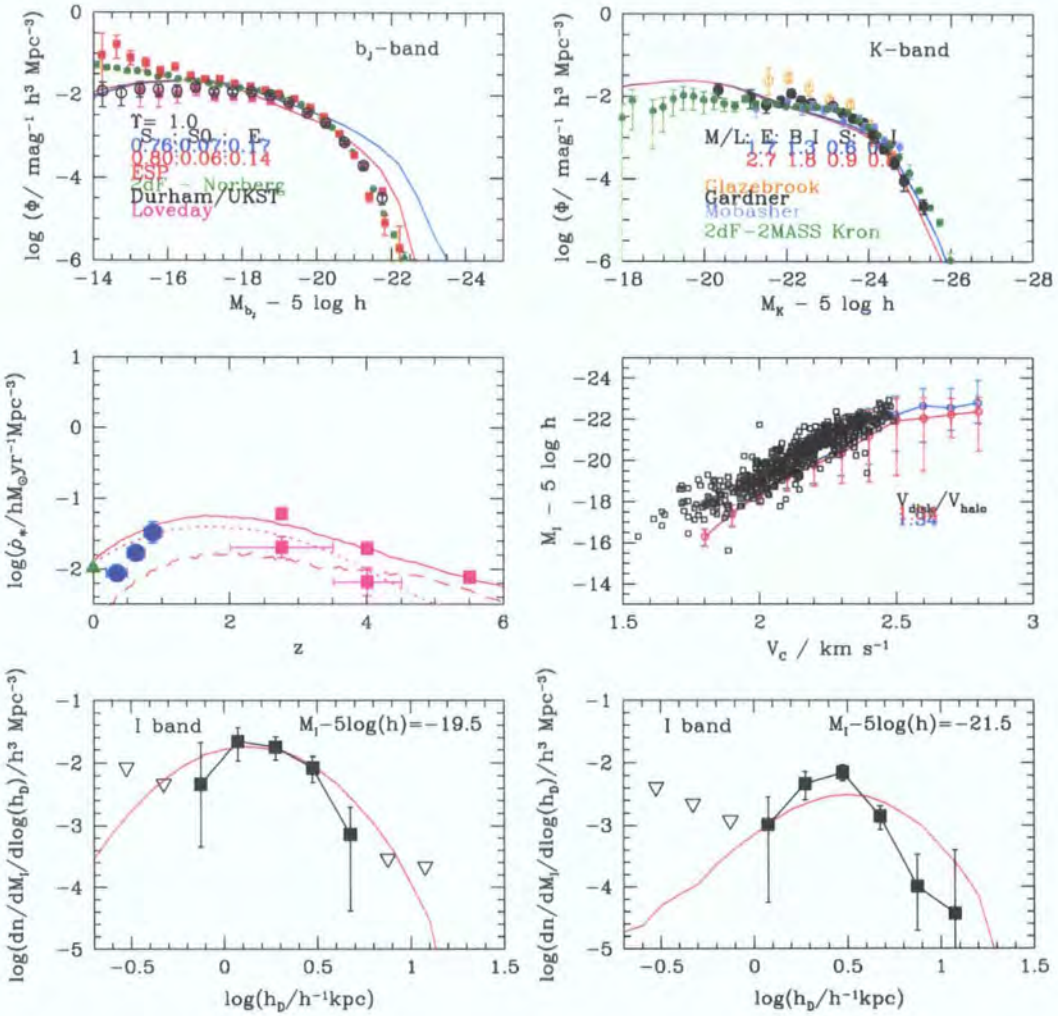

Figure 5.5: The predictions of the ΛCDM2 GALFORM model using our standard dust extinction model (red lines) and switching off dust extinction (blue lines). Where the blue line is not visible, the two results are identical. We also plot a variety of observational data – the observational data in each plot are described in Chapter 1. Top-left: The $z = 0$ b_J-band luminosity function. Top-right: The $z = 0$ K-band luminosity function. Middle-left: The star formation rate per unit volume as a function of redshift (solid – total; dotted – quiescent; dashed – burst). Middle-right: The I-band Tully-Fisher relation at $z = 0$. Bottom-left: The distribution of disc scale-lengths (for galaxies with $-20 < M_I < -19$). Bottom-right: The distribution of disc scale-lengths (for galaxies with $-22 < M_I < -21$).

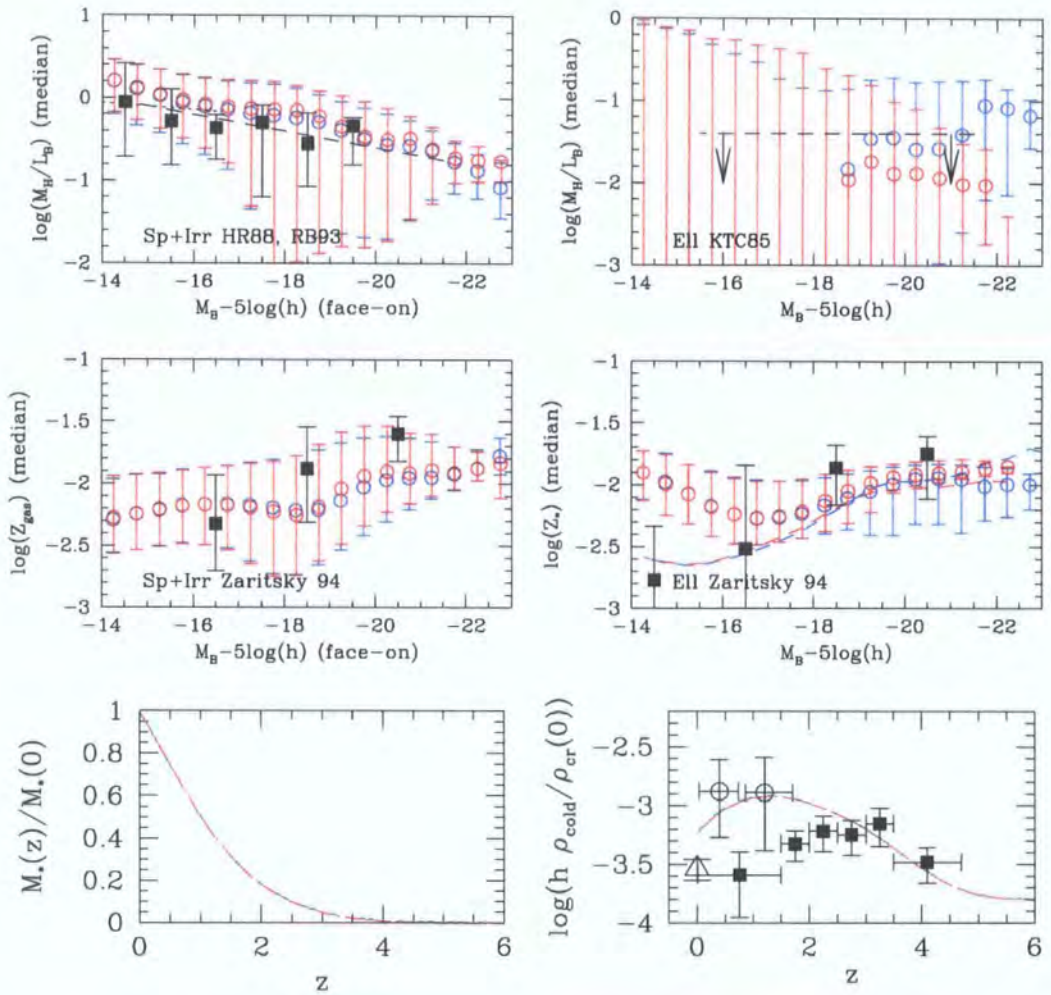

Figure 5.6: Further predictions of the ΛCDM2 GALFORM model using our standard dust extinction model (red) and switching off dust extinction (blue). We also plot a variety of observational data – the observational data in each plot are described in Chapter 1. Top-left: Gas mass to B-band luminosity ratio as a function of M_B for spiralsand irregulars. Top-right: Gas mass to B-band luminosity ratio as a function of M_B for ellipticals. Middle-left: Gas metallicity of spirals and irregulars as a function of M_B at $z = 0$. Middle-right: Stellar metallicity of ellipticals as a function of M_B at $z = 0$. Bottom-left: The fraction of the $z = 0$ stellar mass formed as a function of redshift. Bottom-right: The redshift evolution of the universal density of cold gas (as a fraction of the critical density).

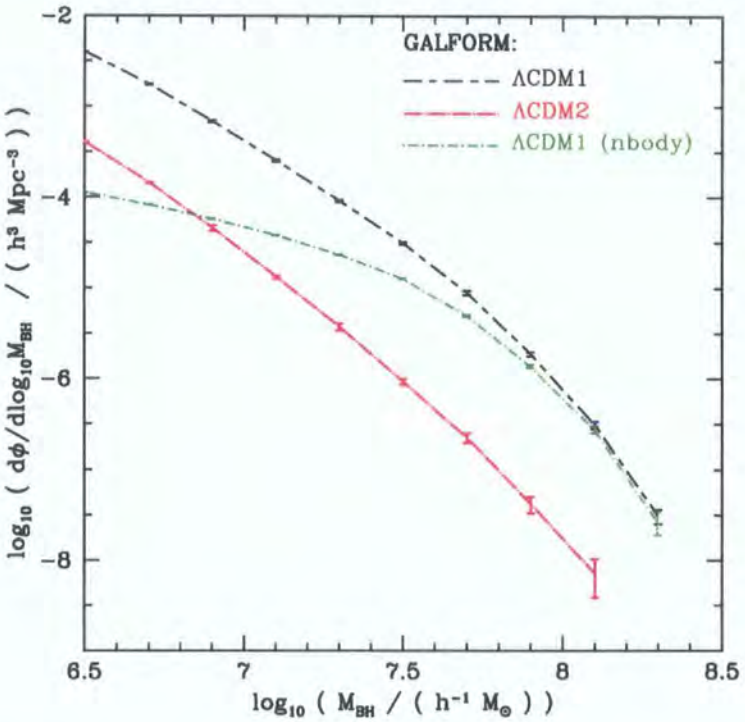

Figure 5.7: The black hole mass function at $z = 6$ in the ΛCDM1 and ΛCDM2 cosmologies. The black and red lines represent the black hole mass function predicted in the ΛCDM1 and ΛCDM2 cosmologies respectively, calculated using finely spaced grids of haloes covering a wide mass range. The green line shows, in comparison, the mass function computed using the haloes in the ΛCDM1 N-body simulation.

5.8 Results

In §5.8.1, we study the black hole mass function in our model, and in §5.8.2 we study the quasar luminosity function. We discuss the limitations of our predictions, and where possible, relate them to the observational data discussed in §5.2-5.3. In §5.8.3 we study the masses of dark matter haloes which are likely to host interesting objects such as massive black holes and luminous quasars. The results so far are (nearly) all derived using a purely semi-analytical model. We then go on to attach our semi-analytic trees to the N-body simulation of the ΛCDM1 cosmology to show the merger histories (§5.8.4) and environments (§5.8.5) of the 5 most massive black holes in the simulation box.

5.8.1 The black hole mass function

In Fig. 5.7, we show the black hole mass function at $z = 6$ from our semi-analytic simulations for both the ΛCDM1 and ΛCDM2 cosmologies. In addition, we also show the

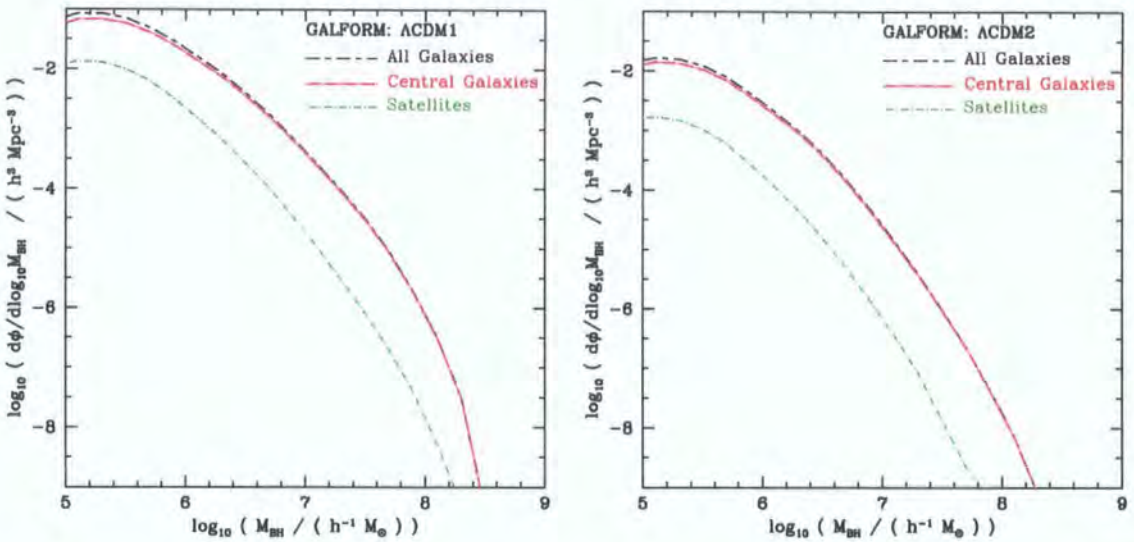

Figure 5.8: The black hole mass function at $z = 6$, divided into black holes residing in 'central' galaxies and 'satellite' galaxies. The black hole mass function for all galaxies, for central galaxies only and for satellite galaxies only are shown by different colours and line types, as indicated in the keys. This is shown for the ΛCDM1 cosmology (left) and the ΛCDM2 cosmology (right).

black hole mass function for the ΛCDM1 cosmology derived using our N-body simulation. The black hole mass function has a very similar form in both cosmologies, the main difference being that the normalization is a factor of ~ 10 lower in the ΛCDM2 cosmology than in the ΛCDM1 cosmology. We show later that black hole mass is a very similar function of halo mass in both cosmologies. Therefore, the difference in the black hole mass functions is probably a direct reflection of the difference in the halo mass functions at $z = 6$ in the two cosmologies, as was shown in Fig. 5.1.

Also in Fig. 5.7, we compare the black hole mass functions we obtain for the ΛCDM1 cosmology when using a semi-analytic description of the halo mass function, and when attaching semi-analytic dark matter trees to our N-body simulation. The two main conclusions we reach are that:

- When using the N-body simulation, we only resolve the black hole mass function for black hole masses above $10^8 h^{-1} M_{\odot}$. This is unsurprising given that, as we go on to show in §5.8.3, the majority of lower mass black holes are hosted in dark matter haloes below the resolution limit of our N-body simulation.

- Counter to our hopes, we have not managed to probe the black hole mass function to significantly lower space densities when using a semi-analytic description of the halo

mass function than the halo mass function available from the N-body simulation. This is a significant point which we now consider in more detail.

When calculating our grid of semi-analytic haloes, we hoped to reach space densities of black holes much lower than is possible using just the N-body simulation, because we are able to target the most massive haloes where we expect to find massive black holes and luminous quasars, and to generate multiple realisations of these halo masses. However, comparison of the N-body and semi-analytic black hole mass functions for the ΛCDM1 cosmology shows that we have not in fact probed significantly lower space densities. We therefore call into question the naïve expectation that the most luminous quasars should reside exclusively in the most massive dark matter haloes. As we will soon show (§5.8.3) it turns out that the majority of black holes of mass $> 10^8 h^{-1} M_\odot$ in fact tend to be hosted in haloes of mass $\sim 10^{12} h^{-1} M_\odot$. As can be seen in Fig. 5.2, for haloes of this mass, the volume probed is only slightly greater than the volume probed by our N-body simulation. We believe that we have more than enough dark matter haloes of mass $> 10^{13} h^{-1} M_\odot$ to probe for rare objects – our next step in probing for rare objects will be to increase the numbers of slightly less massive dark matter haloes, particularly in the mass range $10^{11.5} - 10^{12.5} h^{-1} M_\odot$.

In Fig. 5.8, we plot the black hole mass function in the two cosmologies, showing the contribution to the mass function from central galaxies and from satellite galaxies. The majority of black holes and quasars are hosted in central galaxies, however we find that approximately 10% of black holes are hosted in satellite galaxies. This is perhaps not so surprising since halo-halo mergers occur rapidly at high redshift, and the time-scale between halo mergers at high redshift is fairly short compared to the Hubble time. i.e. $t_{\rm halo-merger} \lesssim t_{\rm Hubble}$. It is common in models of black hole formation using hierarchical mergers of dark matter haloes, such as the work by Volonteri & Rees and Li et al. which we discussed earlier, to assume that there is one galaxy per halo, and to ignore satellite galaxies. Our result here, that 90% of black holes are found in central galaxies, shows that this is actually a reasonable assumption at $z = 6$.

5.8.2 The quasar luminosity function

In Fig. 5.9, we show the quasar luminosity function, in both its differential and cumulative forms. We show results for the ΛCDM1 and ΛCDM2 cosmologies, both with and without Eddington limiting of the quasar luminosity (we explained Eddington limiting of the luminosity in §1.6, and described our procedure for applying this limit in §2.3.2).

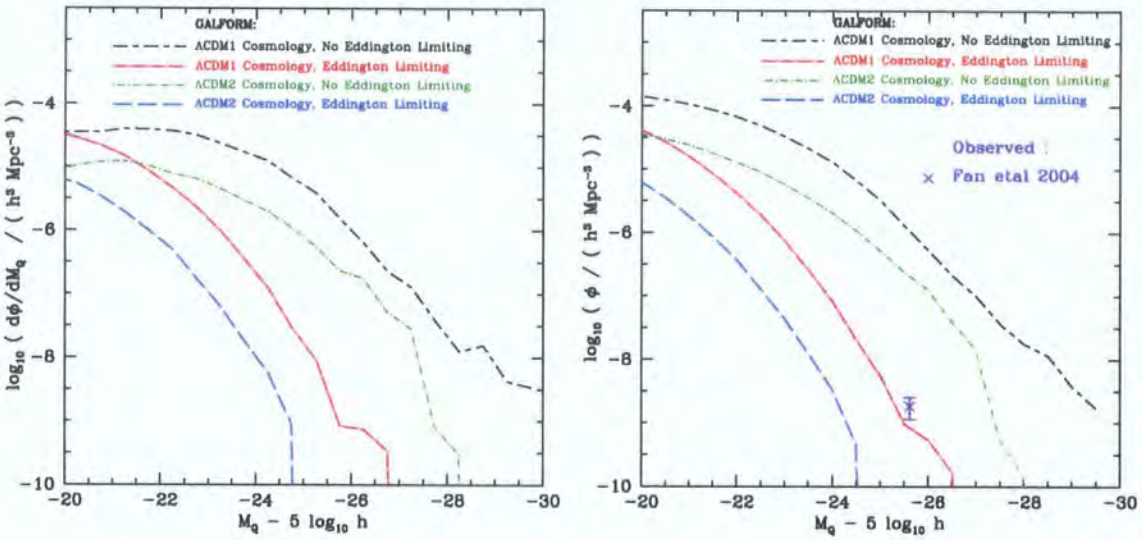

Figure 5.9: Left: The (differential) quasar luminosity function at $z = 6$. Right: The cumulative version of the quasar luminosity function, compared to data from Fan et al. (2004). Both plots show results for the ΛCDM1 and ΛCDM2 cosmologies, assuming either Eddington limiting or no Eddington limiting of the quasar luminosity. The cosmology and the regime for Eddington limiting are indicated in the keys.

If Eddington limiting is ignored, then the predicted luminosites of accreting black holes are vastly greater than observed for quasars at the same space density. The effect of Eddington limiting is far greater than the change in the quasar luminosity function resulting from the differences between the ΛCDM1 and ΛCDM2 cosmologies. Given the observational evidence that $z \sim 6$ quasars are accreting near the Eddington rate, it seems sensible to assume that luminous quasars at $z \sim 6$ are not in fact emitting super-Eddington luminosities. However, we do caution that there are very large uncertainties in the physics of accretion onto $z \sim 6$ black holes. Furthermore, we have not yet fully explored parameter space yet in our galaxy and quasar modelling for either cosmology. Therefore, it is somewhat premature at this stage completely to rule out the ΛCDM2 cosmology on the basis of the observed space densities of quasars. It may well be that, with further refinements to the model, it will be possible to explain the abundance of $z \sim 6$ quasars in the ΛCDM2 cosmology. However, structure formation is biased towards lower redshifts in the ΛCDM2 cosmology, and there is a much greater decrease in ithis model in dark matter halo abundances towards higher redshifts. Therefore, the ΛCDM2 cosmology is likely to present a greater challenge to our modelling of galaxy and black hole formation, if the redshift evolution in the quasar luminosity function is not to be

overestimated.

Admittedly, despite the caveats relating to the black hole mass estimates outlined in §5.3, it is still somewhat concerning that we have not yet managed to produce any black holes of the masses ($\sim 3 \times 10^9 M_\odot$) which seem to be implied by the analyses of Fan et al. (2001a), Willott et al. (2003) and Barth et al. (2003). However, it is reassuring that, at least for the ΛCDM1 cosmology, our prediction of the number density of quasars brighter than $M_{1450} = -26.81$ is remarkably close to the Fan et al. (2004) measurement. It is difficult to infer a space density from a single object at $z = 6.41$ (i.e. the Fan et al. 2001a quasar which is the most luminous ($M_{1450} \sim -27.8$), at the highest redshift, and quite possibly has the largest black hole), so it is difficult to say how many such objects we should find in our simulation. It seems likely that we have probed sufficient volume to uncover numerous objects brighter than $M_{1450} = -26.81$, but that we may not yet have probed quite a large enough volume to be confident of our predictions of the abundance of objects with $M_{1450} \sim -27.8$. Even if our current model does appear to produce sufficiently luminous quasars at $z = 6$, at least for the ΛCDM1 cosmology, this is less likely to be the case once we take into account any feedback limiting (e.g. Eddington limiting of black hole growth) of the black hole accretion rate and the ejection of black holes from the shallow galaxy potentials at $z \gtrsim 6$ due to gravitational processes. However, we also neglect black hole seeds which may ameliorate these differences.

5.8.3 The dark matter haloes hosting black holes and quasars

In Fig. 5.10, we show the halo occupation distribution (HOD) for black holes exceeding various thresholds in mass (left), and the mass function of haloes hosting black holes[1] (right). In Fig. 5.11, we show the results of the same calculations for quasars exceeding various thresholds in luminosity, both with and without Eddington limiting of luminosity. As the threshold black hole mass increases, then the minimum halo mass required to host a black hole increases, and the peak of the mass function of haloes hosting such objects increases (i.e. the typical $M_{\rm halo}$ for a given $M_{\rm BH}$ increases). The same effect is seen just as strongly when thresholds in quasar luminosity are considered, at least when quasar luminosities are assumed to be restricted by the Eddington limit. This is unsurprising,

[1]We refer to the product of the HOD for some selection, X, and the halo mass function as 'the mass function of haloes hosting X'. Where dark matter haloes only host one such object, this is unambiguously a good description. Where dark matter haloes host more than one such object, this is still a good description where it is recognized that these haloes are multiply counted.

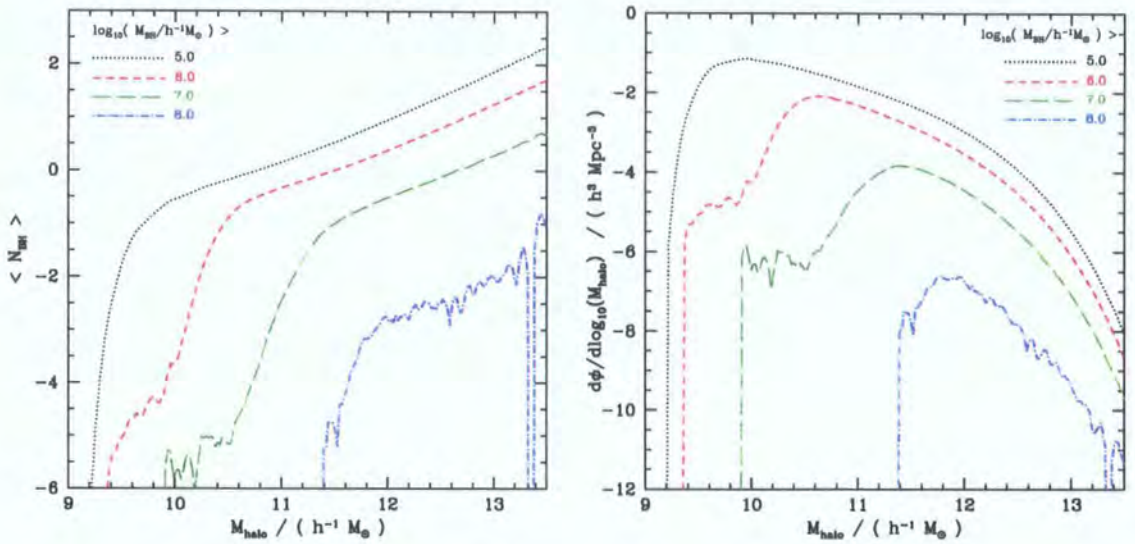

Figure 5.10: The halo occupation distribution (left) and the halo occupation distribution multiplied by the abundance of black holes (right). This is shown for 4 thresholds which must be exceeded by the black hole mass, as indicated in the keys.

and is presumably due to the majority of quasars in our model having luminosities equal to the Eddington luminosity. The trends with luminosity in the minimum halo mass required to host a quasar are far less strong when Eddington limiting of luminosity is switched off.

We plot the host dark matter halo mass (median, with 10–90 percentile spread) as a function of black hole mass (Fig. 5.12, left) and as a function of quasar luminosity (Fig. 5.13, left). Since the dark matter halo mass function at $z \sim 6$ is very steep in the mass range of interest, the typical mass of dark matter halo hosting a given object (as given by the peak of the mass function of dark matter haloes hosting such objects – left hand graphs) tends to be only very slightly more massive than the minimum halo mass required to host such an object (the halo mass range where the HOD is rising very quickly from extremely low values to typical values, where it often plateaus). In Figs. 5.10-5.13, we see that the typical mass of dark matter halo host increases strongly as black hole mass or (Eddington limited) quasar luminosity increases. However, for any given black hole mass or quasar luminosity bin, there is a significant scatter (about an order of magnitude) in the dark matter halo host masses. This large scatter, combined with the steeply descending dark matter halo mass function, means that we predict the typical dark matter halo host masses of the Fan et al. (2004) luminous quasars to be around $10^{12} h^{-1} M_{\odot}$, around an order of magnitude less massive than the Fan et al. estimate of

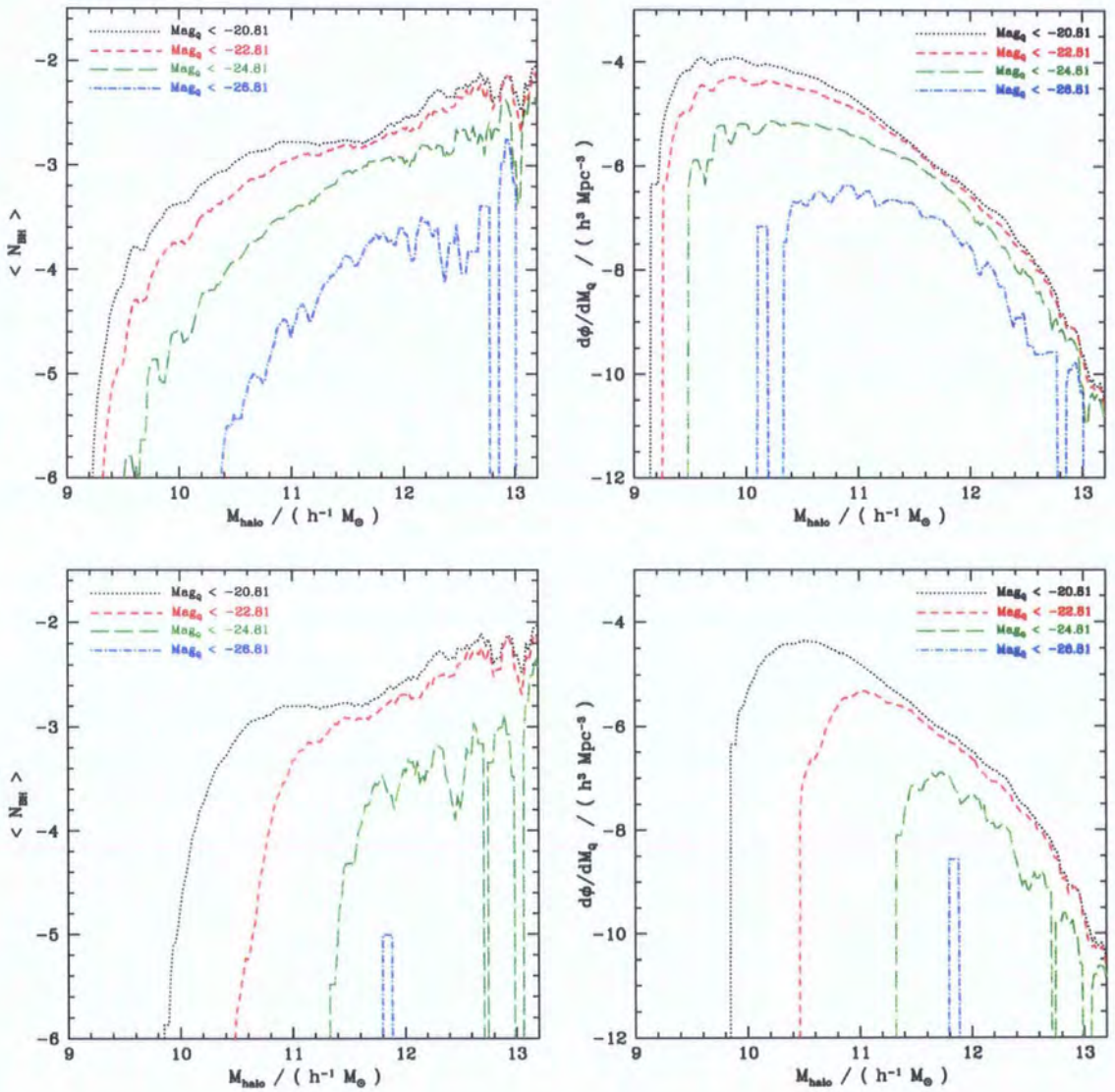

Figure 5.11: The halo occupation distribution of quasars (left) and the quasar halo occupation distribution multiplied by the abundance of black holes (right). We select quasars which exceed 4 thresholds in luminosity, as indicated in the keys. We calculate quasar luminosities both with Eddington limiting (top) and without Eddington limiting (bottom).

$10^{13} M_{\odot}$. This conclusion agrees with the work of Willott et al. (2005), who were unable to observe any bright galaxies surrounding the $z \sim 6.41$ Fan et al. (2001a) quasar, and inferred that a likely explanation is that it is hosted by a dark matter halo significantly less massive than $10^{13} M_{\odot}$, and that this is a reasonable proposition once scatter in the $M_{\rm BH} - M_{\rm halo}$ relations is considered.

For concision, we have only shown the halo occupation distributions and the mass functions of dark matter haloes hosting various objects (Figs. 5.10 and 5.11) for the ΛCDM1 cosmology, although we have also made these calculations for the ΛCDM2 cosmology. We found that the HODS are almost identical, while the mass functions of dark matter haloes hosting objects have a very similar form in both, the main difference being the normalization which is a factor of ~ 10 lower in the ΛCDM2 case, a direct consequence of the difference in mass function normalization. This can be seen in Figure 5.12 (left) – black holes of any given mass are hosted in similar mass haloes in both cosmologies. Quasars of any given luminosity are hosted in similar mass haloes in both cosmologies, assuming that the modelling of the quasar luminosity is identical. However, if the modelling of the luminosity is chosen to match the observed space density of quasars, more Eddington limiting is required in the ΛCDM1 model, while we must switch off Eddington limiting altogether to match the space densities in the ΛCDM2 model. Thus, if the quasar model is chosen to match the data without consideration of the likelihood that luminosities are Eddington limited, quasars in the ΛCDM2 cosmology of any given luminosity would correspond to lower mass black holes with super-Eddington luminosities, and hosted by lower mass haloes.

Admittedly, it does seem somewhat extreme that only 1% of haloes of mass $10^{13} h^{-1} M_{\odot}$ host a black hole more massive than $10^{8} h^{-1} M_{\odot}$. This might be partly due to a general underestimation of black hole masses at high redshifts in our model (see §4.1). It is also quite possible that we have over-estimated the scatter in the $M_{\rm BH} - M_{\rm halo}$ and $M_{\rm BH} - M_{\rm halo}$ relations. This is somewhat supported by some of the low-redshift predictions of the Baugh et al. (2005) galaxy formation model – it has a problem matching the relationship between clustering strength and galaxy luminosity at low redshift, which is likely to be partly due to a scatter in galaxy-halo relations which is a little too large, although perhaps mainly due to a $L_{\rm galaxy} - M_{\rm halo}$ relationship with too shallow a slope (Carlton Baugh, priv. comm.)

The majority of quasars, in particular faint quasars, are found in dark matter haloes with masses lower than the mass resolution limit of our simulation, and furthermore our

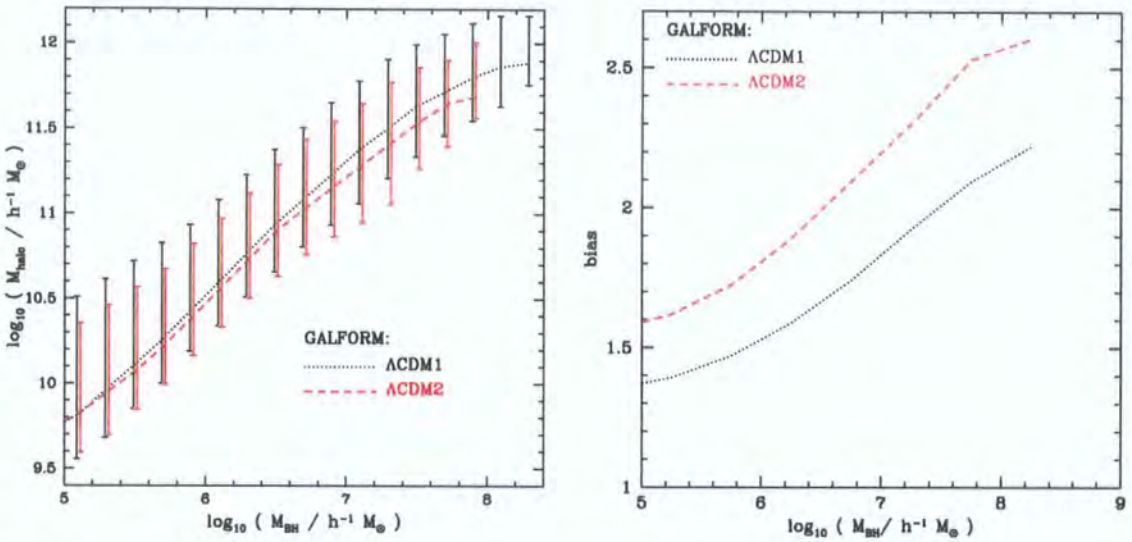

Figure 5.12: Left: The median dark matter halo mass as a function of black hole mass at $z = 6$. 10 and 90 percentiles of the distribution are indicated by error bars. Right: The bias as a function of black hole mass at $z = 6$ for the ΛCDM1 and ΛCDM2 cosmologies. For both plots, the cosmology is indicated in the keys.

N-body simulation has insufficient volume to measure the clustering of brighter quasars directly. Therefore, we make estimates of black hole and quasar clustering using a simple analytic bias model. For each black hole, we calculate the bias of its dark matter halo host. The effective bias of black holes in any given mass range is calculated as the weighted sum of all the bias factors for all of these black holes. The correlation function expected can then be found by multiplying the correlation function of the dark matter by the appropriate bias factor (squared). The bias of black holes as a function of their mass is shown in Fig. 5.12 (right). The bias of quasars as a function of their luminosity is calculated in the same way, and is shown in Fig. 5.13 (right). As expected, higher mass black holes, and brighter (Eddington limited) quasars are more biased, and hence more clustered. Whilst black holes of the same mass are hosted in very similar mass dark matter haloes for both cosmologies, they are more biased in ΛCDM2 than in ΛCDM1 since dark matter haloes of any given mass are rarer in ΛCDM2 than in ΛCDM1. However, since the bias factors differ between the two cosmologies by the same factor as σ_8, we would expect the same clustering amplitude in each cosmology for samples defined by black hole mass.

Predictions of bias made using the Sheth et al. (2001) formalism are fairly accurate when compared to numerical simulations, although they are likely to be modified by a

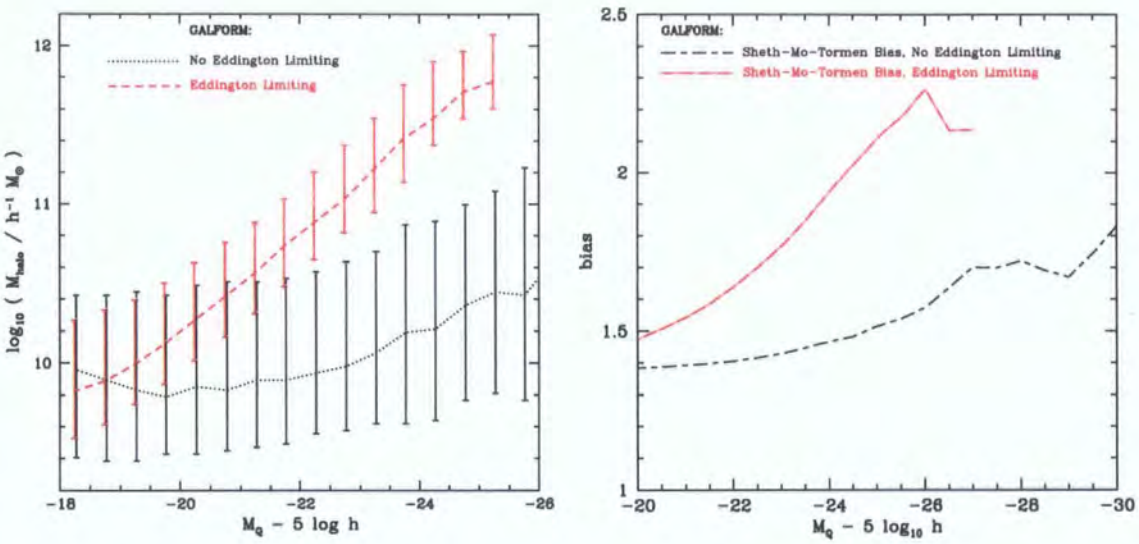

Figure 5.13: Left: The median halo mass as a function of quasar luminosity at $z = 6$. The 10–90 percentile spread is shown as error bars. Right: the bias of quasars as a function of their luminosity at $z = 6$. Both of these plots show results for the ΛCDM1 cosmology, and with quasar luminosities calculated both with and without Eddington limiting, as indicated by the keys.

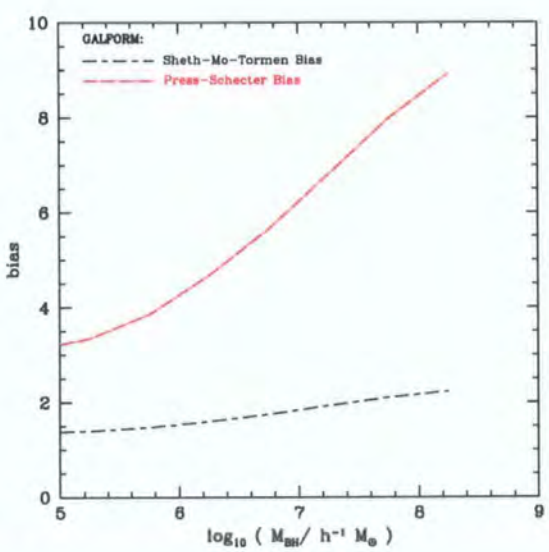

Figure 5.14: The bias of black holes as a function of their mass. This is shown using the Press-Schechter bias (Mo & White 1996) and the Sheth, Mo & Torman (2001) bias, as indicated in the key.

factor of the order of $10 - 20\%$ once 2nd order effects are taken into account, such as halo formation time, large-scale environment or halo concentration (e.g. Gao et al. 2005a; Harker et al. 2006; Wechsler et al. 2006) and this may well have an observable effect of a similar magnitude on observables such as galaxies (Reed et al. 2007; Croton et al. 2007). Finally, we have also briefly looked at the effect of using different approximations to calculate the bias. Mo & White (1996) outlined an analytic prescription for computing bias using Press-Schechter theory. In Fig. 5.14, we show the bias factors of black holes, as a function of their mass, for the Mo & White and the Sheth et al. models. The two results are very different, particularly for more massive haloes. The Mo & White (1996) bias model is highly inaccurate in the regime we are probing in this Chapter (high redshift, rare peaks). This is a useful demonstration of the care which must be taken to check the validity with respect to numerical simulations of any result derived using Press-Schecter theory.

Why is the bias far greater for any given mass (at the high mass end of the mass function) for the Mo & White (Press-Schecter) model than for the Sheth et al. model? The bias is the excess of peaks (let us assume that these are all peaks above some mass threshold which is required to host a galaxy brighter than some luminosity, or some other observable object) above the average number density of such peaks in the Universe. An overdense region of the Universe (i.e. a high mass dark matter halo) can be considered as a small patch of the Universe evolved to a slightly later time (lower redshift). Thus the bias for this overdense region is given by the evolution of the number density of peaks greater than a given mass, and thus of the halo mass function (integrated to be the mass function of haloes greater than a threshold mass), between the current time and a slightly later time (this time difference does not vary significantly between the Press-Schecter andSheth et al. models). The Press-Schechter mass function is far steeper than the Sheth et al. mass function at the high mass end. The evolution of the mass function with time can be considered largely as a shift in mass, as the characteristic mass M_* increases. The shift in M_* is fairly similar for the Press-Schecter model and the Sheth et al. model. Therefore, the increase in the integrated number of haloes above some given threshold of mass corresponding to this shift in M_* is greater for the Sheth et al. mass function (whilst there is some difference is the normalisation of the bias factor, i.e. the mean galaxy/peak density, this evolution is small compared to the evolution in the excess galaxy/peak density). Therefore the bias of overdense regions of the Universe is greater in the Press-Schecter model than the Sheth et al. model.

All of the results in this section agree very well with the hypothesis that the differences between the ΛCDM1 model and the ΛCDM2 model are due to the difference in normalization between the two mass functions. This suggests that predictions for various cosmologies are well characterised by a decomposition into the halo mass function and, given a fixed physical model, the abundance of objects in haloes as a function of halo mass. This confirms that the 'halo model of galaxy formation' (Neyman & Scott 1952; White & Rees 1978; White & Frenk 1991; Benson et al. 2000; Cooray & Sheth 2002), where galaxy formation can be decomposed into galaxy formation as a function of halo mass and the halo mass function, is a very useful formalism.

However, although we have not yet demonstrated this ourselves, the difference between ΛCDM1 and ΛCDM2 is in fact somewhat more subtle than a simple difference in the normalization of the dark matter halo mass function, although this is indeed a useful first approximation. Not only does the normalization of the dark matter halo mass function at a given redshift change[2], but the merger history of any given halo also changes. Specifically, structure formation is pushed towards lower redshifts in the lower σ_8 ΛCDM2 cosmology than in the ΛCDM1 cosmology. Li et al. (2006) perform simulations for two cosmologies very similar to the ΛCDM1 and ΛCDM2 cosmologies (they refer to these as WMAP1 and WMAP3). They used initial conditions with exactly the same phases for the two cosmologies. This allowed them to identify individual objects in both simulations, and to compare their formation. They found that individual dark matter haloes at $z \sim 6$ were slightly less massive in the WMAP3 (ΛCDM2) cosmology than in the WMAP1 (ΛCDM1) cosmology, by a factor of ~ 1.6. They also compared the merger histories of corresponding haloes in the two cosmologies. The first halo to form in the simulation forms at $z \sim 16.8$ in WMAP1 (ΛCDM1) and at $z \sim 14.4$ in WMAP3 (ΛCDM2). The formation history of the most massive dark matter halo at $z \sim 6$ moves towards lower redshifts in the WMAP3 (ΛCDM2) than in the WMAP1 (ΛCDM1) cosmology, although the number of major mergers is the same.

We may investigate this difference in halo formation time in the future. For now, we

[2]The difference between the two dark matter halo mass functions is better characterized as a difference in the *mass* normalization (or of the characteristic mass, M_*), but, since the halo mass function is not highly curved in the region of interest, a difference in normalization of halo space density describes the difference in the halo mass function fairly well. For the most part, we describe the difference in the halo mass function as a difference in the space density normalization since our galaxy formation modelling is largely the same in both cosmologies for any given dark matter halo mass.

speculate that for our standard model, without Eddington limiting of black hole growth[3], black hole formation would vary little between the two cosmologies apart from a shift in merger and formation redshifts (and, of course, the difference of a factor of ~ 10 in space density). However, if we consider effects such as Eddington limiting of mass accretion, then the time differences represented by these redshift differences in halo and black hole formation times/redshifts could become important; the more limited time available in the ΛCDM2 between dark matter halo formation redshift and observed quasar redshift could be important in reducing the number of e-folds of black hole growth allowed, compounding the difficulty of producing massive black holes in the ΛCDM2 cosmology.

5.8.4 The growth histories of the most massive black holes at $z = 6$

In Figs. 5.16-5.19, we show the growth histories of the five most massive black holes in our N-body simulation volume.

It is interesting to compare these growth histories to those of black holes at $z = 0$, as were shown in Chapter 3. Examining these plots, we find that, in contrast to black holes at $z = 0$, black holes of mass $\sim 2 \times 10^8 h^{-1} M_\odot$ at $z = 6$ grow almost exclusively by accretion onto a single progenitor. This is a conclusion which we have also verified statistically using the same analysis of black hole merger histories as was done in Chapter 3. We in fact find that black holes of all masses are assembled almost exclusively by accretion onto their main progenitor. Furthermore, examination of the 5 examples shown here suggests that black hole growth occurs in a very small number of accretion episodes, with $\sim 50 - 99\%$ of the growth occuring in the final accretion episode, which occurs between $z \sim 7$ and $z = 6$.

We find that the 5 most massive black holes in our simulation appear to grow in 3-7 starbursts (mass ratio < 20:1). This is comparable to the finding of Li et al. (2006) that their analogue of the most luminous Fan et al. quasar grows in 7 major mergers (mass ratio < 5:1). They will probably find more major mergers in the merger tree they use for their most luminous quasar, since they have explicitely chosen the most massive (or one of the most massive) $z = 6$ dark matter halo in their simulation. Our most massive black holes are found in lower mass dark matter haloes, which is perhaps consistent with our use of both major and minor galaxy mergers to trigger growth onto the black hole.

[3]Recall that Eddington limiting of black hole growth is not the same as Eddington limiting of black hole luminosity – for most of this Chapter we consider Eddington limiting of *luminosity*, but here we are referring to Eddington limiting of black hole *growth*.

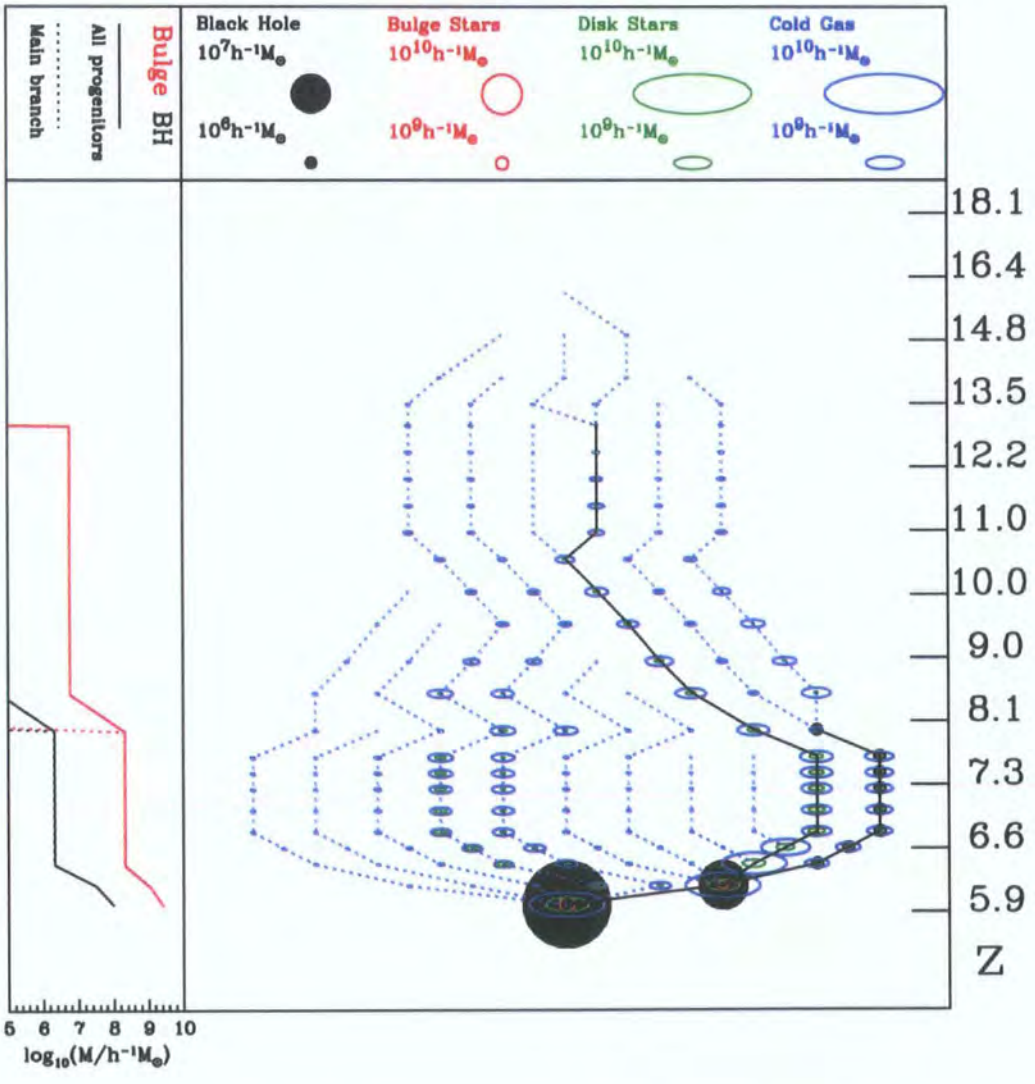

Figure 5.15: Formation history of the most massive black hole in the N-body simulation volume. The black hole has a mass of $2.29 \times 10^8 h^{-1} M_\odot$, and is hosted in a dark matter halo of mass $7.81 \times 10^{11} h^{-1} M_\odot$

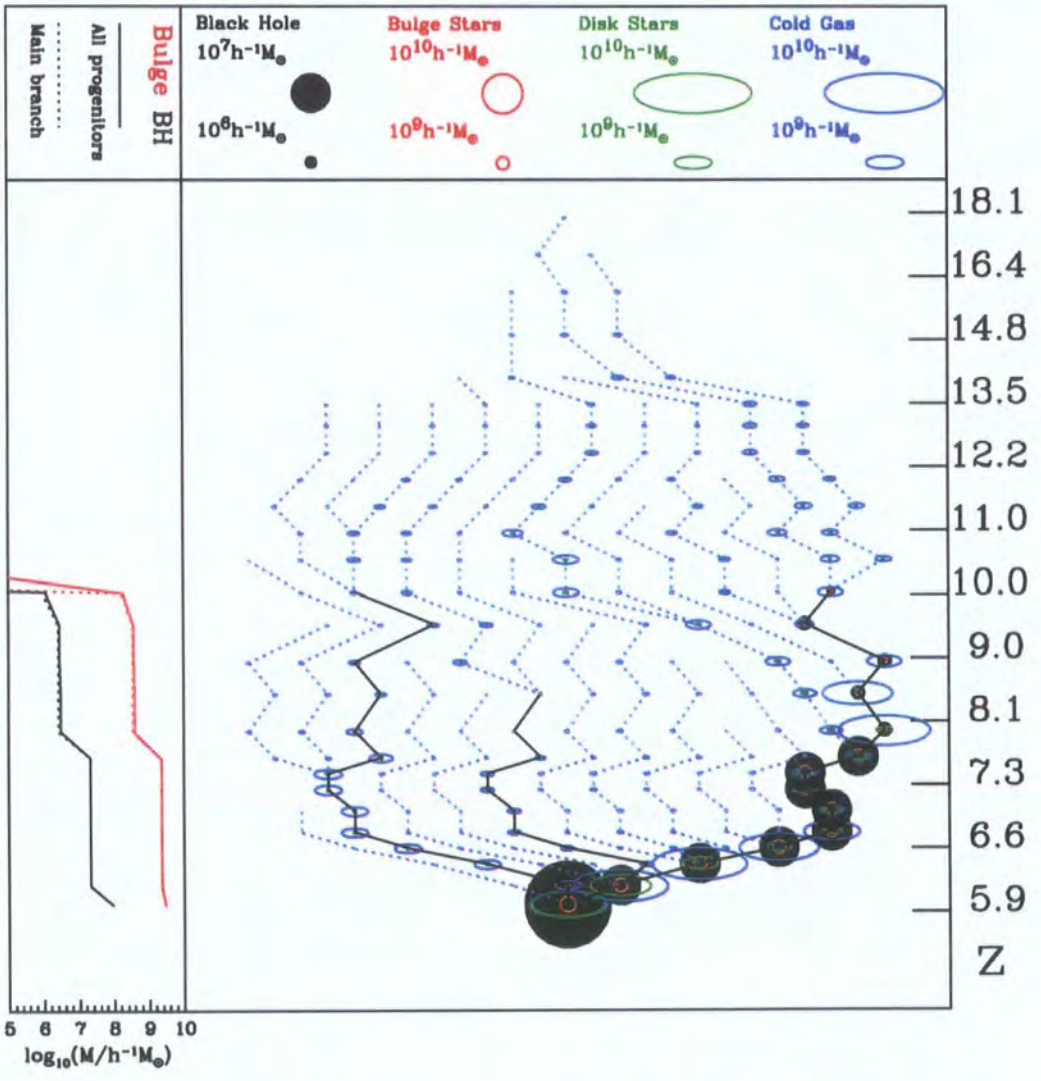

Figure 5.16: Formation history of the 2nd most massive black hole in the N-body simulation volume. The black hole has a mass of $2.24 \times 10^8 h^{-1} M_\odot$, and is hosted in a dark matter halo of mass $3.18 \times 10^{12} h^{-1} M_\odot$

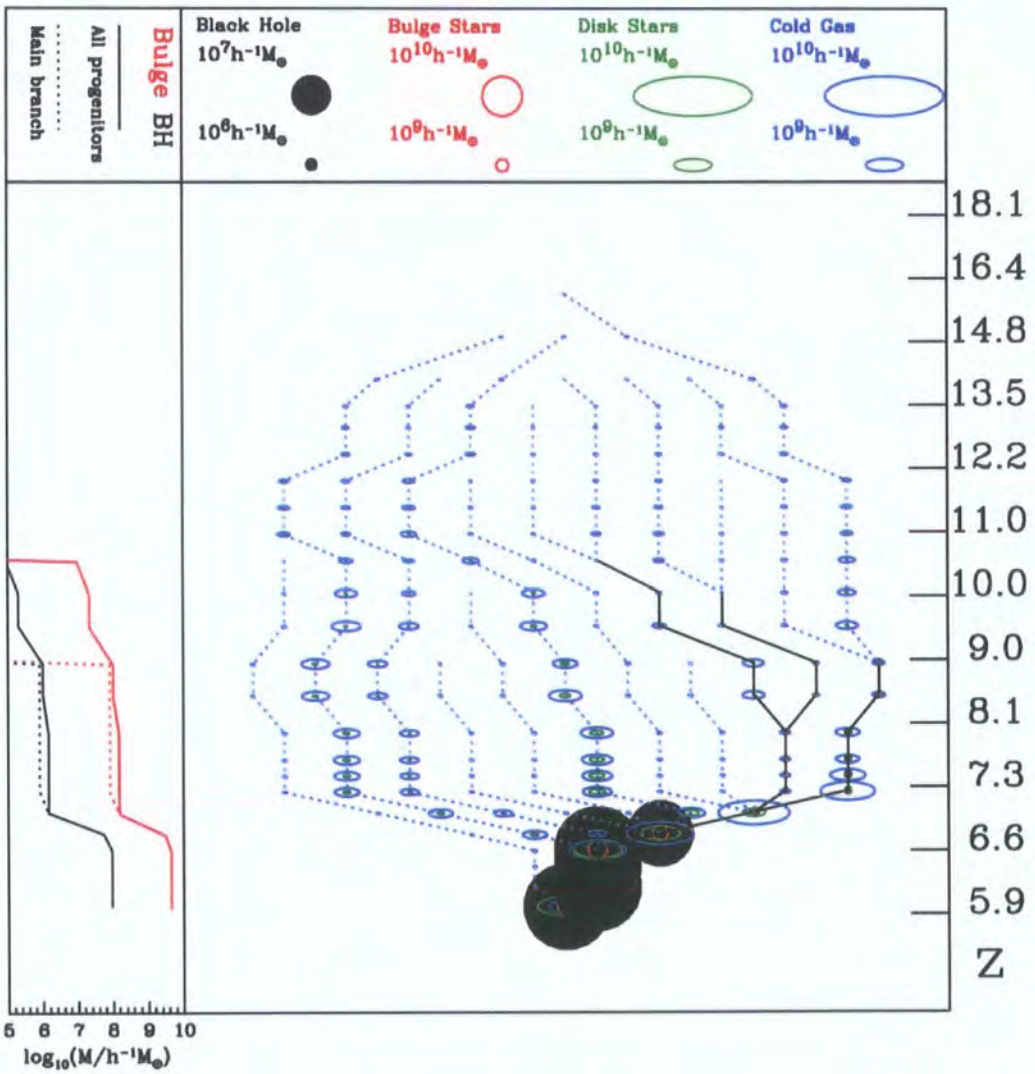

Figure 5.17: Formation history of the 3rd most massive black hole in the N-body simulation volume. The black hole has a mass of $2.21 \times 10^8 h^{-1} M_\odot$, and is hosted in a dark matter halo of mass $9.49 \times 10^{11} h^{-1} M_\odot$

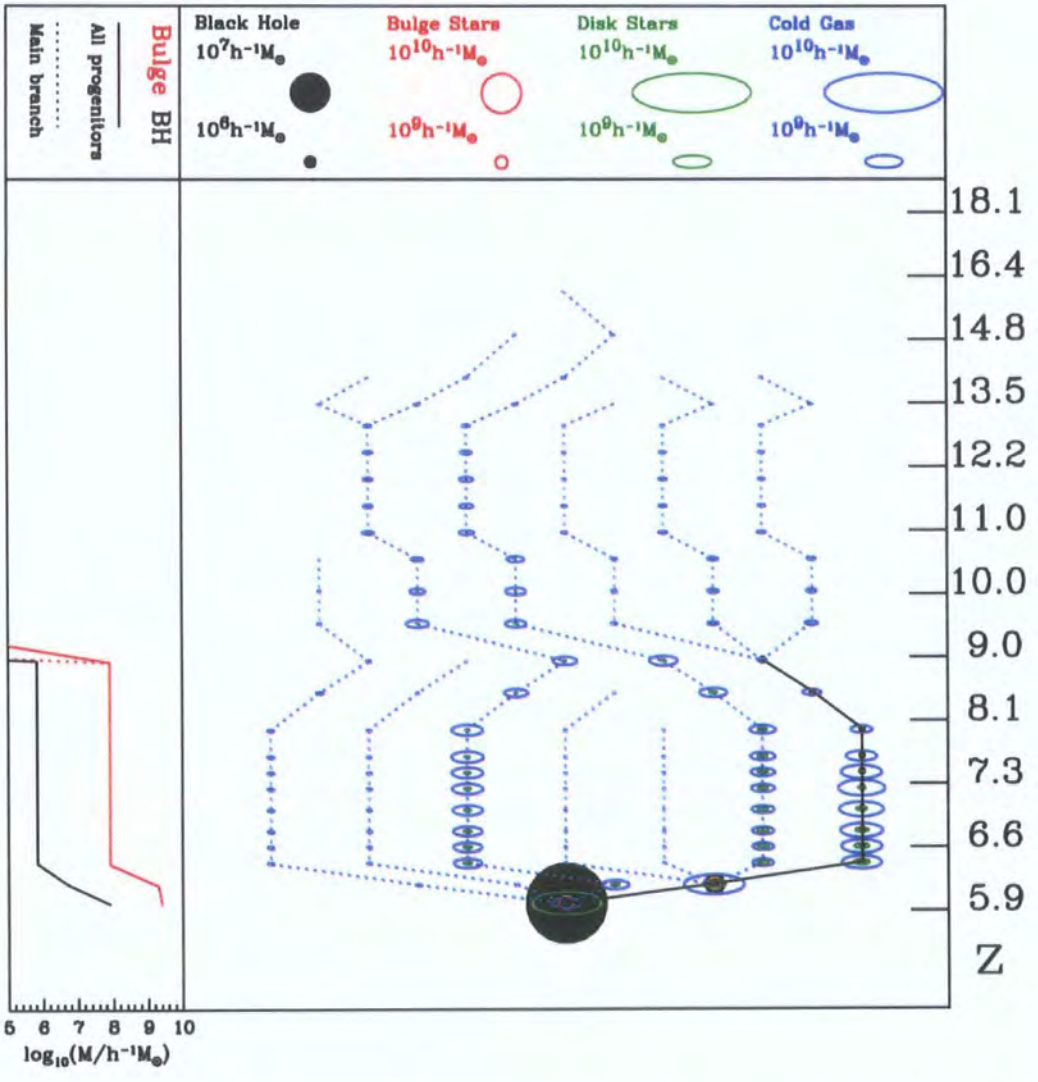

Figure 5.18: Formation history of the 4th most massive black hole in the N-body simulation volume. The black hole has a mass of $1.94 \times 10^8 h^{-1} M_\odot$, and is hosted in a dark matter halo of mass $6.14 \times 10^{11} h^{-1} M_\odot$

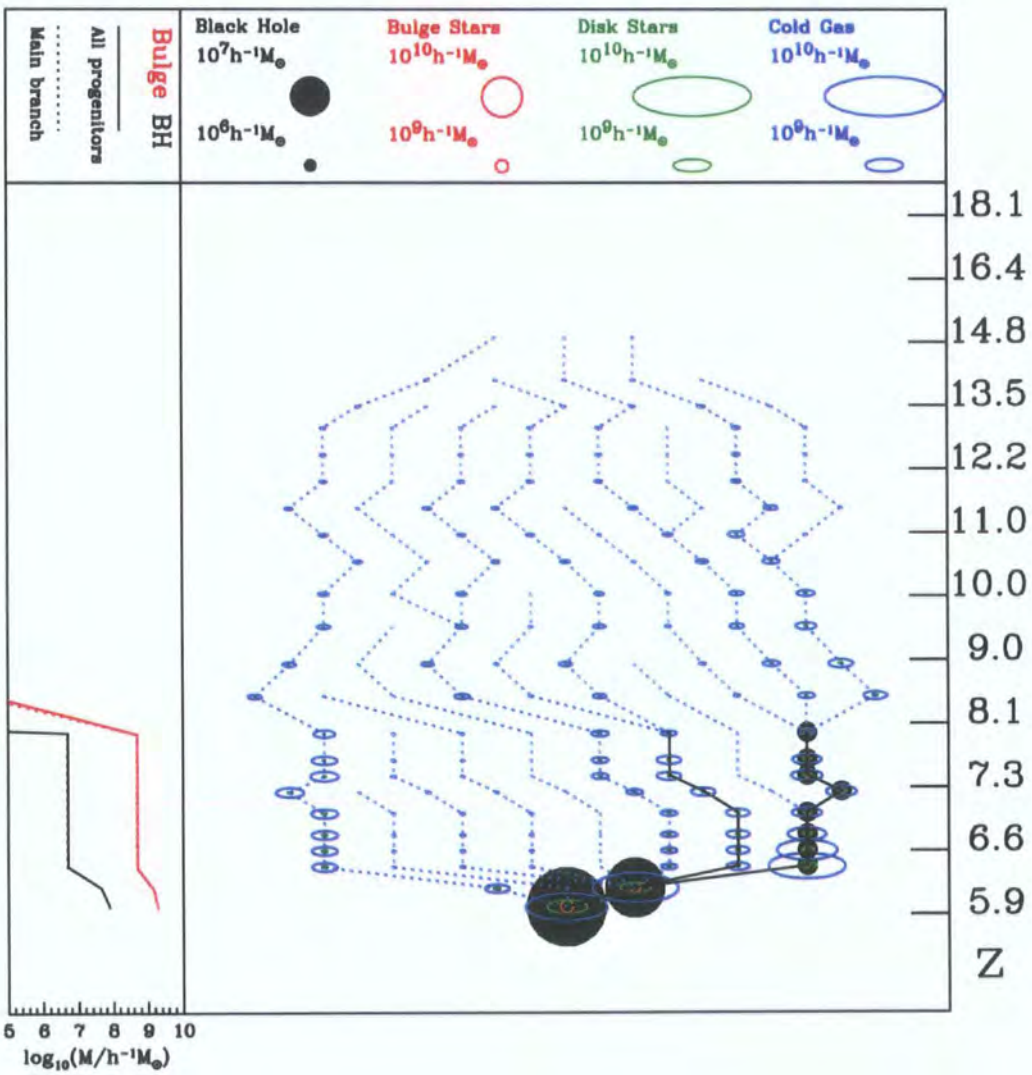

Figure 5.19: Formation history of the 5th most massive black hole in the N-body simulation volume. The black hole has a mass of $1.86 \times 10^8 h^{-1} M_\odot$, and is hosted in a dark matter halo of mass $7.25 \times 10^{11} h^{-1} M_\odot$

5.8.5 The environments of the most massive black holes at $z = 6$ and their descendants

In Fig. 5.20 (left 5 panels), we show the environments of the five most massive black holes in our N-body simulation volume ($2.4 \times 10^9 h^{-3}\mathrm{Mpc}^3$). Visually inspecting the black holes in the environments surrounding the 5 most massive black holes in our simulation, it is obvious that black holes live in a range of environments. In the future, we would like to make predictions of the distribution of observable galaxies surrounding the most luminous quasars at $z \sim 6$, but this is difficult to do with such a limited resolution dark matter simulation. In the right 5 panels of the figure, we show the same regions in the simulation at $z = 0$, marking the positions of the black holes which we identified at $z = 6$. We see that at $z = 0$, in many but not all cases, there are fewer black holes in the simulation. In one or two cases, black holes have left the simulation box, but in most cases, we believe that this is because the dark matter haloes hosting these black holes have merged, and quite probably, the black holes have also merged, or are close to merging[4]. There appears to be more merging of black holes in those simulation boxes with a higher number of black holes, and which are in denser, more highly biased regions of the simulation. This is as might be expected; we believe that more highly biased regions of the Universe undergo more merging during their evolution. However, with only five regions of the simulation studied, this can as yet be regarded only as an anecdotal result, which we would like to study more quantitatively. In Table 5.2, we list the masses of the 5 most massive black holes we find in the simulation box at $z = 6$, the masses of their dark matter halo hosts at $z = 6$, and the masses of the descendent dark matter haloes of these hosts at $z = 0$. We find that, as expected (Springel et al. 2005c; Li et al. 2006), the most massive black holes at $z = 6$ end up in the most massive dark matter haloes at $z = 0$ – these would be expected to be massive clusters.

[4]In this particular calculation, we were only able to track the dark matter haloes from $z = 6$ to $z = 0$. Therefore, black holes at $z = 0$ have been placed on the centre of mass of their host dark matter halo. In some cases, we may have identified black holes in distinct galaxies in the same dark matter halo to be at the same position. In future work, we hope to track the galaxy and black hole formation with our semi-analytic model using N-body merger trees, which will remove this ambiguity.

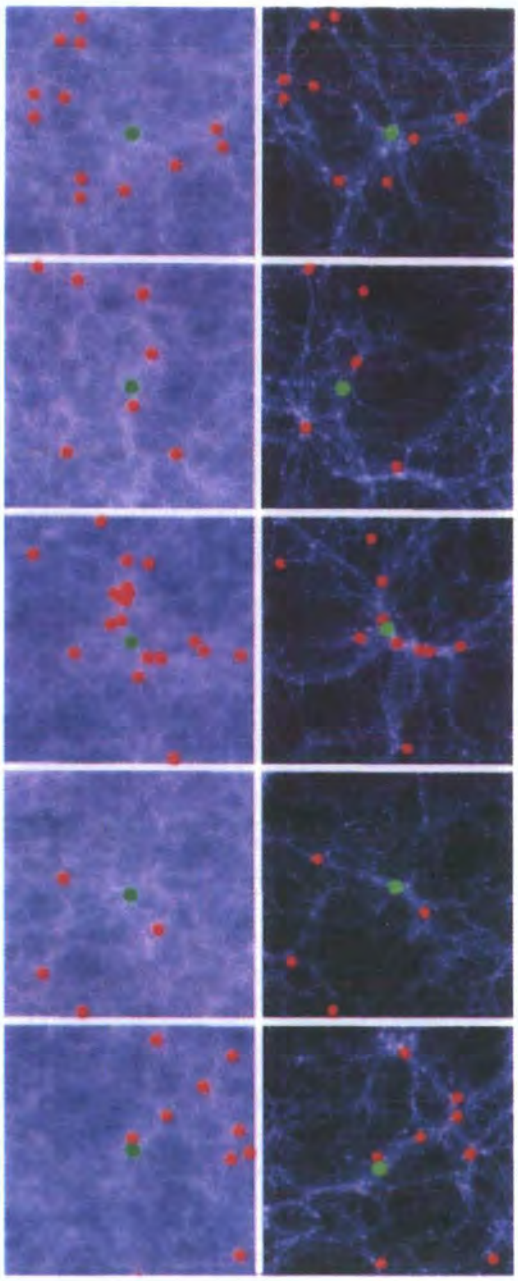

Figure 5.20: Left: The environments (box width = $85h^{-1}$Mpc comoving) of the 5 most massive black holes in the simulation at $z = 6$, marked by green dots. Other black holes in the region are marked by red dots. The dark matter is shown using a black (lowest density) → blue → white (highest density) colour scheme. Right: The corresponding regions of the simulation at $z = 0$. The positions of all black holes marked in the $z = 6$ images are also marked in the $z = 0$ image, where they still lie within the $z = 0$ volume. Colour scheme as before. Both left and right panels are ordered from 1st most massive black hole (top) to 5th most massive black hole (bottom). These images were made by Raul Angulo, using a catalogue of black hole positions at $z = 6$ provided by RKM. Raul Angulo was also responsible for tracing the positions of the black holes and haloes from $z = 6$ to $z = 0$.

Table 5.2: In this table, we list the masses of the 5 most massive black holes in the N-body simulation at $z = 6$, of their host dark matter haloes at $z = 6$, and of the descendent host dark matter haloes at $z = 0$. The $z = 0$ halo masses were provided by Raul Angulo, who traced the $z = 6$ dark matter haloes in the simulation forwards to $z = 0$.

1st 5 most massive BH at $z = 6$	$M_{\rm BH}/h^{-1}M_{\odot}$ at $z = 6$	$M_{\rm halo}/h^{-1}M_{\odot}$ at $z = 6$	$M_{\rm halo}/h^{-1}M_{\odot}$ at $z = 0$
1st	2.28×10^8	7.81×10^{11}	1.57×10^{15}
2nd	2.24×10^8	3.18×10^{12}	1.12×10^{15}
3rd	2.21×10^8	9.49×10^{11}	1.69×10^{15}
4th	1.94×10^8	6.14×10^{11}	1.10×10^{15}
5th	1.86×10^8	7.25×10^{11}	1.29×10^{15}

5.9 Discussion and conclusion

We are able to match the abundance of luminous quasars measured by Fan et al. (2004) with our standard galaxy and black hole formation model in the ΛCDM1 cosmology. We struggle to produce sufficiently luminous quasars when we assume a ΛCDM2 cosmology without invoking implausible assumptions of greatly super-Eddington luminosities, although we hesitate to rule it out on this basis since we have yet to explore models of galaxy and black hole formation more thoroughly. The most massive black holes at $z \sim 6$ appear to grow in a qualitatively different way – almost completely by accretion and in a small number of starbursts – compared to those at $z = 0$, which grow largely by mergers of pre-existing black holes. Whilst the scatter in the $M_{\rm BH} - M_{\rm halo}$ and $M_{\rm Q} - M_{\rm halo}$ relations is quite possibly smaller than presented here, it is unlikely to be zero, as is often assumed (e.g. in the work of Li et al. 2006; Volonteri & Rees 2006). Thus the brightest quasars and the most massive black holes will be found in dark matter haloes somewhat less massive than is often assumed. This result may one day be verified by comparing the clustering of quasars to the results we present here.

Chapter 6

Black hole growth in a galaxy formation model with AGN feedback

So far in this Thesis, we have concentrated exclusively on predictions of black hole growth using the Baugh et al. (2005) galaxy formation model, or close variants of it. However, this model does not include an explicit treatment of AGN feedback, although the parameters we require for superwind feedback may well require a contribution from AGN feedback to be energetically feasible. We now consider another galaxy formation model developed at Durham using a closely related GALFORM code, the Bower et al. (2006) model, which does includes AGN feedback from low efficiency accretion of mass onto supermassive black holes.

In §6.1, we describe the motivation for including feedback from black holes accreting mass at low fractions of the Eddington rate in massive haloes with haloes of hot gas in quasi-hydrostatic equilibrium. In §6.2, we describe the Bower et al. model of galaxy formation, concentrating in particular on its differences from that of Baugh et al., in particular gas cooling (§6.2.1), AGN feedback (§6.2.2), N-body dark matter halo merger trees (§6.2.3), star formation (§6.2.4) and additional channels of black hole growth (§6.2.5). We then compare the predictions of the Bower et al. model of the $M_{\rm BH}$ – bulge relations (§6.3) and quasar luminosity function (§6.4) to observational data. We show predictions of the evolution of the black hole mass function (§6.5) and of the contributions to it from haloes various $z = 0$ masses (§6.6). In §6.7, we predict the contributions of mergers and accretion to the growth of black holes of various masses, and the evolution of the BH-BH merger rate. In §6.8, we show the predictions for the evolution of the $M_{\rm BH}$ – bulge relations, and compare these briefly to those of the Baugh et al. model and to the observational data. In §6.9, we study the formation times (considered both in terms of mass *'transformation'* and *'assembly'*) of black holes of various masses, and compare these predictions to those of the Baugh et al. model. We compare the predictions of downsizing

in black hole growth in the Bower et al. and Baugh et al. models in §6.10. Finally, in §6.11 we summarize and conclude.

6.1 The motivation for AGN feedback

Hierarchical models of galaxy formation have long suffered from overcooling in massive haloes, and AGN feedback has been proposed as the important missing ingredient required to circumvent this. The overcooling and previous attempts to avoid it have led to a number of concerns over the predictions of earlier models, specifically :

• Hierarchical galaxy formation models have often had problems reproducing the bright end of the $z = 0$ galaxy luminosity function, and in particular the sharpness of the break brightwards of L^*, whilst at the same time using realistic energy budgets for feedback (Benson et al. 2003a). This has been even more of a problem in the past few years, since the baryon content of the Universe as measured from the abundances of light elements and nucleosynthesis arguments has been revised upwards from $\Omega_b = 0.02$ to $\Omega_b = 0.04$ (e.g. Cuoco et al. 2004). Benson et al. (2003a), using the new determination of Ω_b, were able to reproduce the $z = 0$ K-band luminosity function, but only by using an energy budget in supernovae ~ 5 times greater than that which we believe to be available.

• It has long been noted that hierarchical galaxy formation models tend to produce massive galaxies at the centres of clusters which are too blue (White & Frenk 1991). This has often been dealt with in a rather ad hoc fashion, by imposing a cut-off to cooling of gas in haloes more massive than some threshold mass (e.g. Kauffmann et al. 1993). Ideally, we would like our models to have a more physically motivated way of shutting off star-formation in cooling flows at the centres of massive haloes.

• Galaxies seem to lie on one of two clearly demarcated sequences in colour-magnitude space: the red sequence of passive galaxies and the blue sequence of active, star-forming galaxies (e.g. Kauffmann et al. 2003; Baldry et al. 2004). Very few galaxies lie in the colour-magnitude space between these sequences. There appears to be a characteristic galaxy mass of $2 - 3 \times 10^{10} M_{\odot}$ (Kauffmann et al. 2003; Baldry et al. 2004). Less massive galaxies are more likely to lie on the blue sequence, whilst more massive galaxies are more likely to lie on the red sequence. Except for the most massive galaxies, however, there are still galaxies of any given mass on both sequences. The clear demarcation between the red and blue sequences suggests that galaxies move from the blue sequence to the red sequence very rapidly, and thus that the shut-off of star formation is very rapid.

• Hierarchical galaxy formation models have sometimes struggled to reproduce simultaneously the $z = 0$ galaxy properties *and* galaxy stellar mass functions at high redshift. Either the models produce too many bright galaxies at $z = 0$, or too few bright, massive galaxies at $z > 0$. For example, Kauffmann et al. (1999) managed to reproduce the high abundances of massive galaxies at high redshift, but at the same time underproduced L^* galaxies at $z = 0$. Conversely, Benson et al. (2003a) and Baugh et al. (2005) managed to reproduce the $z = 0$ galaxy luminosity function, but at the same time did not produce sufficiently high stellar masses at high redshifts. We must remember, of course, that observationally inferred stellar mass functions are derived using specific assumptions for the IMF and star formation history, which may well be wrong, leading to errors in the conversion from photometry to stellar mass. Therefore, hierarchical models are perhaps most consistently judged by their fits to K-band luminosity functions, which are a direct output of the model.

In the past year, a number of groups apart from the Durham group have published semi-analytic models of galaxy formation including the effect of AGN feedback. In this Chapter, we consider the Bower et al. (2006) model of galaxy formation, which is able to resolve all of the above problems, largely due to the inclusion of AGN feedback. Other work which has recently included the effects of AGN feedback includes that of Cattaneo et al. (2006a), Croton et al. (2006), Kang et al. (2006), Menci et al. (2006) and Monaco et al. (2007). These authors arrived at conclusions similar to those of Bower et al. regarding the importance of AGN feedback in solving these longstanding issues in galaxy formation.

6.2 The Bower et al. galaxy formation model

In Figs. 6.1 and 6.2, we show a comparison of the Bower et al. (2006) model with the standard set of redshift zero observed galaxy properties presented in Chapter 1. As can be seen, the model is very successful in reproducing the observed $z = 0$ galaxy properties. In particular, we would like to emphasize the excellent agreement we find with the $z = 0$ B-band and K-band luminosity functions. We obtain a break at L^* very much like that observed, with a near exponential decline in the abundance of galaxies brighter than L^*. The match we find to the $z = 0$ luminosity functions is perhaps slightly better than that achieved with the Baugh et al. model (see Figs. 1.5 and 1.6), although both are very good.

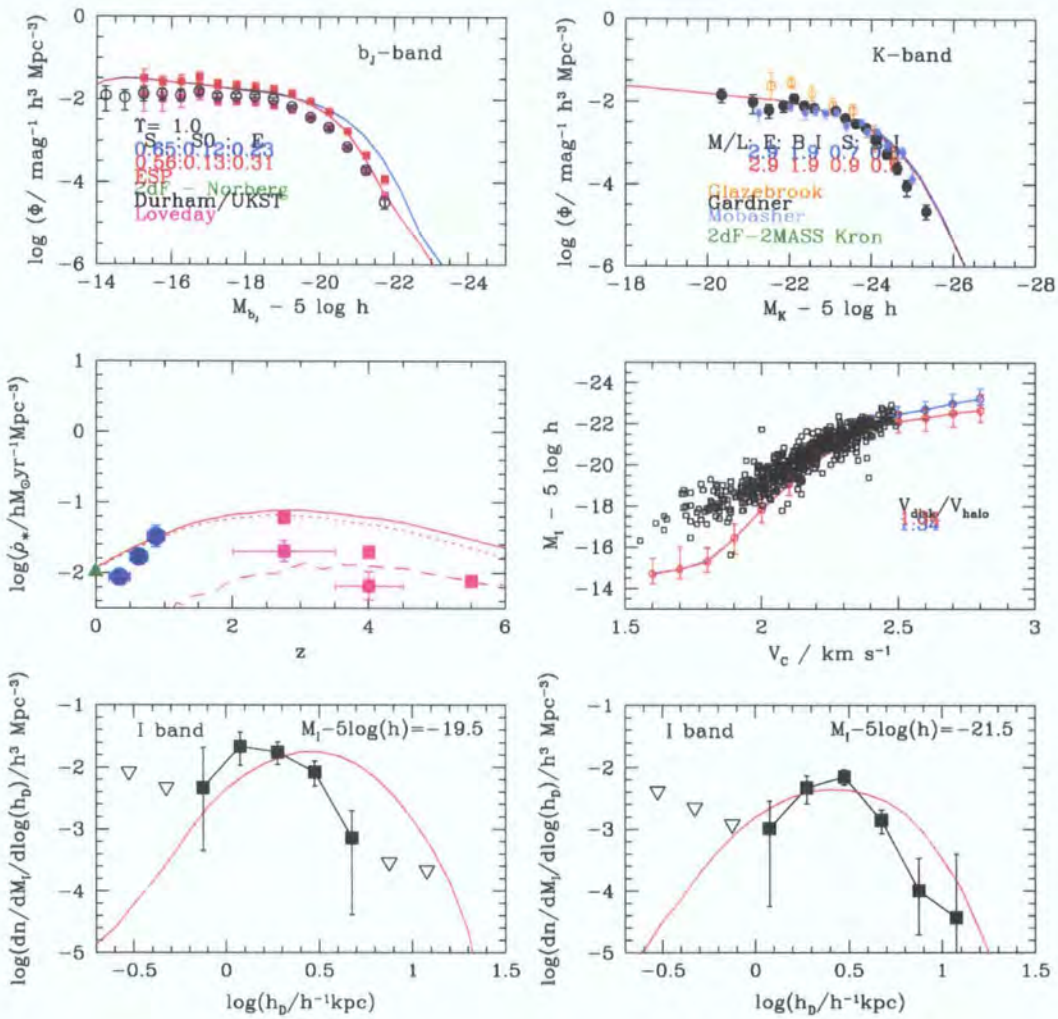

Figure 6.1: The predictions of the Bower et al. (2006) GALFORM model using our standard dust extinction model (red lines) and switching off dust extinction (blue lines). Where the blue line is not visible, the two results are identical. We also plot a variety of observational data – the observational data in each plot are described in Chapter 1. Top-left: The $z = 0$ b_J-band luminosity function. Top-right: The $z = 0$ K-band luminosity function. Middle-left: The star formation rate per unit volume as a function of redshift (solid – total; dotted – quiescent; dashed – burst). Middle-right: The I-band Tully-Fisher relation at $z = 0$. Bottom-left: The distribution of disc scale-lengths (for galaxies with $-20 < M_I < -19$). Bottom-right: The distribution of disc scale-lengths (for galaxies with $-22 < M_I < -21$).

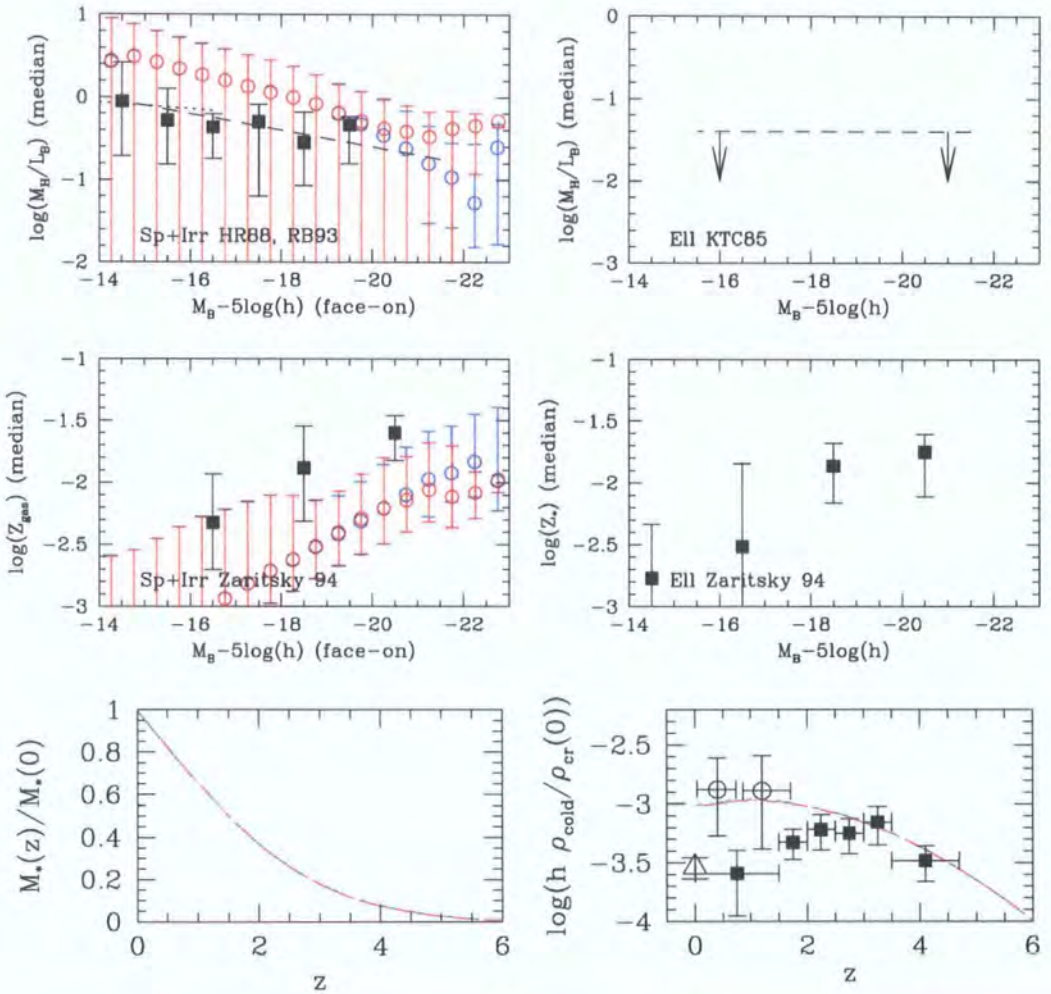

Figure 6.2: Further predictions of the Bower et al. (2006) GALFORM model using our standard dust extinction model (red) and switching off dust extinction (blue). We also plot a variety of observational data – the observational data in each plot are described in Chapter 1. Top-left: Gas mass to B-band luminosity ratio as a function of M_B for spirals and irregulars. Top-right: Gas mass to B-band luminosity ratio as a function of M_B for ellipticals. Middle-left: Gas metallicity of spirals and irregulars as a function of M_B at $z = 0$. Middle-right: Stellar metallicity of ellipticals as a function of M_B at $z = 0$. Bottom-left: The fraction of the $z = 0$ stellar mass formed as a function of redshift. Bottom-right: The redshift evolution of the universal density of cold gas (as a fraction of the critical density).

Both the Baugh et al. model (used in earlier Chapters of this Thesis) and the Bower et al. model (used in this Chapter) are offspring of the Benson et al. (2003a) study of feedback in galaxy formation, and develop different aspects of the feedback models explored by Benson et al. in an effort to match the $z = 0$ K-band luminosity function, in particular its bright end. Baugh et al. concentrated on superwind feedback, which ejects gas from dark matter haloes, whilst Bower et al. concentrated on energy input into the hot gas halo, preventing its cooling. We now review the main aspects of the Bower et al. model which differ from that of Baugh et al..

6.2.1 Gas cooling

A striking feature of the galaxy luminosity function is that it has a sharp break (e.g. Norberg et al. 2002). Since the dark matter halo mass function has far less curvature than the galaxy luminosity function, this suggests that a scale is imposed by some baryonic, physical process in galaxy formation, as was discussed in detail by Benson et al. (2003a). In the White & Frenk (1991) and Cole et al. (2000) cooling model, we simply assume that when gas falls into dark matter haloes, it is shock-heated to the virial temperature of the halo. However, in less massive dark matter haloes, from galactic-sized haloes downwards, gas-cooling may be so rapid that any shock is ineffective at heating the gas (Birnboim & Dekel 2003; Kereš et al. 2005). This is particularly the case at higher redshifts, where the transition between the cooling regimes moves to lower mass dark matter haloes.

Semi-analytic modellers have long argued that it is not necessary to consider whether or not gas falling into dark matter haloes is shock-heated to the virial temperature of the halo, since in low mass haloes, it will cool almost immediately whether or not it has been shock-heated to the virial temperature (e.g. Cole et al. 2000). However, feedback from a central source (specifically an AGN) is likely only to be effective at preventing cooling when gas forms a hot atmosphere in quasi-hydrostatic equilibrium (Binney 2004); if gas has already cooled before falling into the centre of the halo, then a central energy source will not be able to prevent cooling. So, there is a boundary between two qualitatively different regimes of gas cooling: cooling before infall (in lower mass haloes) and the shock-heating of infalling gas to form a quasi-hydrostatic hot halo before it has had sufficient time to cool (in higher mass haloes). It may well be this boundary which imparts the physical scale necessary to produce the observed break in the galaxy luminosity function.

We now consider the dark matter halo mass at which this boundary may occur, and how it may vary with redshift. Since the density of dark matter haloes of any given mass

increases as their formation redshift increases, then their cooling times will be shorter at higher redshift. Thus the halo mass below which cooling can take place before efficient shock-heating takes place is greater at high redshift than at low redshift (Birnboim & Dekel 2003; Kereš et al. 2005). This means that the transition between the two regimes of cooling moves to lower masses of dark matter halo at lower redshift. This may well be fundamental in producing the observed downsizing of star formation in galaxies, where downsizing is the inference that star formation moves to lower mass objects at lower redshift.

In the Bower et al. model, we distinguish between cooling that that occurs on a free-fall timescale, and cooling which occurs from a quasi-hydrostatic hot halo. We assume that if the cooling time is shorter than the free-fall timescale, then gas is able to fall to the centre of a dark matter halo on a free-fall timescale. If the cooling time is longer than the free-fall timescale, then we assume that gas is able to form a quasi-hydrostatic hot halo. In practice, we in fact replace the free-fall timescale in this calculation by 0.58 times the free-fall timescale, in order to achieve a good fit to observed galaxy properties. We actually work in terms of a 'free-fall radius' and a 'cooling radius', which we calculate for a halo at every timestep. We assume that when a halo is formed, then the initial gas profile is identical to that of the dark matter, which we assume to have an NFW profile (Navarro et al. 1997). The 'free-fall' radius is the maximum radius in a halo for which the free-fall timescale is less than the time since halo formation. The 'cooling radius' is the maximum radius in a halo for which the cooling timescale is less than the time since halo formation. All gas inside both the freefall radius and the cooling radius is assumed to cool immediately, and cannot be prevented from cooling by AGN feedback. All gas outside both the freefall radius and the cooling radius is assumed to form a quasi-hydrostatic hot halo. As the cooling radius propagates outwards, then all gas inside the freefall radius which at the latest timestep enters the cooling radius is allowed to cool from the quasi-hydrostatic hot halo. Only that gas which cools from the quasi-hydrostatic hot halo can be prevented from cooling by AGN feedback.

6.2.2 AGN feedback

In the model used in Chapters 2 - 5 of this Thesis and in Malbon et al. (2006), we were able to treat black holes simply as 'sink' particles which are affected by galaxy formation, but which do not themselves influence galaxy formation. We now allow AGN feedback to prevent the cooling of hot gas which is in quasi-hydrostatic equilibrium in its dark matter

halo.

We first calculate the energy per unit time, $L_{\rm cool}$, which would be radiated by hot gas in the halo if it were allowed to cool. We then parameterise the AGN power available to offset this cooling luminosity as a fraction, $\epsilon_{\rm SMBH}$, of the Eddington luminosity of the central black hole in the galaxy, $L_{\rm Edd}$. Gas in a halo is prevented from cooling when:

$$L_{\rm cool} < \epsilon_{\rm SMBH} L_{\rm Edd} \ , \tag{6.1}$$

and the rate of energy output of the AGN, $L_{\rm BH}$, is given by:

$$L_{\rm BH} = L_{\rm cool}/\epsilon_{\rm SMBH} \ . \tag{6.2}$$

To achieve a good fit to galaxy properties, we take $\epsilon_{\rm SMBH} = 0.5$.

6.2.3 N-body dark matter merger trees, and resolution effects

The Bower et al. model uses dark matter merger trees drawn from the 'Millennium' simulation (Springel et al. 2005b, Helly et al. in prep.). In Fig. 2.1 in §2.2.4, we showed the effect of varying the resolution of the dark matter halo merger trees on the predicted black hole mass function at $z = 0$ and $z = 6$. We compared the black hole mass function obtained with our standard resolution of $M_{\rm res} = 5 \times 10^8 h^{-1} M_\odot$ with that obtained using the resolution of the 'Millennium' simulation merger trees of $M_{\rm res} = 1.72 \times 10^{10} h^{-1} M_\odot$. Particularly at $z = 6$, there is a dramatic difference in the black hole mass function obtained using these two resolutions; the use of the 'Millennium' resolution degrades the mass resolution we obtain for our black holes from $10^5 h^{-1} M_\odot$ to $10^6 h^{-1} M_\odot$ for $z = 0$ and to $6 \times 10^6 h^{-1} M_\odot$ for $z = 6$. We have not yet specifically done a resolution test using the Bower et al. model, but we would expect that there is a similar degradation in resolution.

6.2.4 Star formation

There are some significant differences in the modelling of star formation in the Bower et al. and Baugh et al. models of galaxy formation. In the Baugh et al. model, we assumed a constant timescale for quiescent star formation, whilst in the Bower et al. model, we assume that the timescale of quiescent star formation scales with the dynamical time of the disc. In the Bower et al. model, we use the same IMF for burst star formation as we do for quiescent star formation. In contrast, in the Baugh et al. model, we used a top-heavy IMF in starbursts.

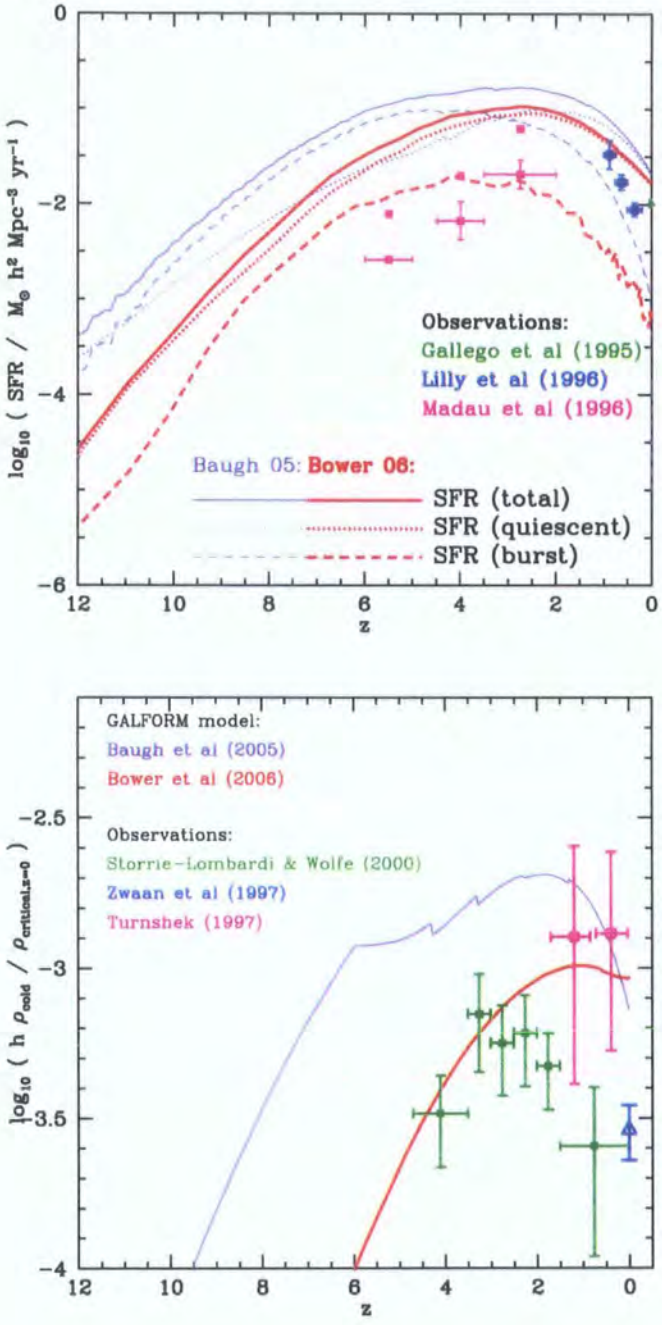

Figure 6.3: Global star formation rate and cold gas fraction in the Bower et al. model (thick red lines) compared to that in the Baugh et al. (thin purple lines) model. Top: For each model, we show the global star formation rate (solid lines), and the contributions to the star formation rate from quiescent star formation (dotted lines) and burst star formation (dashed lines). We also show some observational determinations of the global star formation rate, as indicated in the key. Bottom: For each model, we show the global density of cold gas as a fraction of the critical density. We also show some observational determinations, as indicated in the key. The abrupt drop in cold gas at $z = 6$ is due to the very simple photoionization model we are using.

Dynamical times are shorter at high redshift. In the Bower et al. the timescale for quiescent star formation scales with dynamical time. Therefore, star formation, which is dominated by quiescent star formation, is accelerated at high redshifts compared to low redshifts. In Fig. 6.3, we show the global star formation rate (top) and the global cold gas content of the Universe (bottom) in both models. At high redshifts, the cold gas content of the Universe is far lower in the Bower et al. model than in the Baugh et al. model, but decreases far less rapidly between $z = 1$ and $z = 0$. The depletion of cold gas as it forms stars or as it is reheated due to star formation feedback is accelerated at high redshift in the Bower et al. model and at low redshift in the Baugh et al. model. To some extent, the lower global cold gas content in the Bower et al. model may also be due to the poorer mass resolution of the N-body dark matter merger trees used; gas is only able to cool once it is in a resolved halo. We used $\sim 30\times$ higher mass resolution in our calculations of the Baugh et al. model, which means that dark matter merger trees extend to much higher redshifts, and gas is able to cool earlier. We also note that the faster rate of quiescent star formation at higher redshifts in the Bower et al. model means that there is less cold gas available in galaxy mergers than there is in the Baugh et al. model, and therefore there is a much lower rate of star formation in merger driven starbursts.

As seen in In Fig. 6.3 (top), the star formation rate is significantly higher in the Baugh et al. model than in the Bower et al. model, particularly towards higher redshifts. There are a number of reasons for this:

- The increased resolution in the Baugh et al. model, and the increased cold gas content of the Universe at higher redshifts in the Baugh et al. model, as discussed above.

- In the Baugh et al. model, merging satellite galaxies are allowed to trigger starbursts in galaxies 20 times their mass, whilst in the Bower et al. model they are only able to trigger bursts in galaxies 10 times their mass. Furthermore, there are more of these small galaxies available in the Baugh et al. model to trigger starbursts due to the much better mass resolution. Thus there is a significantly enhanced rate of burst star formation in the Baugh et al. model.

- The larger contribution to high redshift star formation of dust-enshrouded starbursts means that the Baugh et al. model is able to match observations of high redshift star-forming galaxies, in particular the $z = 3$ and $z = 4$ luminosity functions of Lyman break galaxies and the number counts of sub-mm galaxies. The model was tuned to give the high rate of burst star formation which appeared to be required to match these observations. These predictions have not yet been made for the Bower et al. model.

• The Baugh et al. model assumes that the burst mode of star formation, which dominates at high redshifts, has a top-heavy IMF. Using a top-heavy IMF, most of the mass involved in star formation is in fact recycled into the intergalactic medium, and only a small fraction is locked up in long-lived stars.

We now briefly compare the metallicity of gas and of stars in the two models. In Fig. 6.2, we see that using the Bower et al. model, the metallicity of gas in galaxies (middle-left) and of stars in galaxies (middle-right) is lower than it is when we use the Baugh et al. model (Fig. 1.6 of Chapter 1, middle-left and middle-right). This is very likely to be a direct consequence of the lower global star formation rate in the Bower et al. model. Furthermore, the top-heavy IMF used in starbursts in the Baugh et al. model returns a higher yield of metals from high mass supernovae to the interstellar medium than does the standard IMF used in bursts in the Bower et al. model.

6.2.5 Additional channels of black hole growth

In the Malbon et al. (2006) model used in the first few Chapters of this Thesis, we only considered black hole accretion in merger driven starbursts. In this Chapter, following Bower et al. , we allow three different channels for accretion onto black holes:

• Starbursts triggered by galaxy mergers. As in Malbon et al. , in starbursts triggered by galaxy mergers, we assume that a fraction, $F_{\rm BH}$, of the gas mass which is turned into stars[1] is accreted onto the black hole.

• Starbursts triggered by disc instabilities. When a galactic disc does not have sufficient rotational support to offset its self-gravity, it is unstable to small perturbations, for example from dark matter substructures or satellite galaxies (Efstathiou et al. 1982; Mo et al. 1998). We follow Cole et al. (2000), and take as our stability criterion:

$$\epsilon_{\rm disc} = V_{\rm max}/(GM_{\rm disc}/r_{\rm disc})^{1/2} \quad . \tag{6.3}$$

where $V_{\rm max}$ is the circular velocity at the half-mass radius (not the maximum velocity in the rotation curve). If, at any timestep, $\epsilon_{\rm disc} < \epsilon_{\rm stable\,disc}$, then we consider the disc to be unstable. We consider value of $\epsilon_{\rm stable\,disc}$ around 1, choosing the exact value to fit the observed properties of galaxies and of the $M_{\rm BH}$ – bulge relations. We take $\epsilon_{\rm stable\,disc} = 0.8$, which is fairly close to the value of 1.1 found by Efstathiou et al. (1982). When

[1]The mass of stars produced in the burst is the mass *after* taking into account feedback processes which may expel gas from the galaxy and the recycling of mass from stars. This is the case irrespective of whether a starburst is triggered by a merger or a disc instability.

calculating the scale-size of the unstable disc and the resulting bulge, we assume that the disc shrinks (this occurs both through infall of cooling gas to form the disc and through adiabatic contraction of any existing disc) only until it first becomes unstable (not until it is rotationally supported, as was assumed by Cole et al.), at which time a spheroid is formed. The scale size of the spheroid is calculated using energy conservation from the properties of the disc at the largest radius at which it would become unstable. An important point to note is that $F_{\rm BH}$ is the same for all starbursts, irrespective of whether they are triggered by a major merger, a minor merger or a disc instability.

• Accretion of cooling gas in quasi-hydrostatic haloes. AGN feedback can only occur when mass is accreted onto the central black hole. We must allow for the increase in the mass of the black hole corresponding to the accretion required to generate the feedback energy needed to offset gas cooling. Assuming that accretion occurs with an efficiency ϵ_1, then:

$$\Delta E = \epsilon_1 \Delta M_{\rm BH} c^2 \Longrightarrow L_{\rm BH} = \epsilon_1 \dot{M}_{\rm BH} c^2 \quad . \tag{6.4}$$

since:

$$L_{\rm cool} = \epsilon_{\rm SMBH} L_{\rm BH} \quad , \tag{6.5}$$

then:

$$\dot{M}_{\rm BH} = L_{\rm cool}/(\epsilon_{\rm SMBH} c^2) \quad , \tag{6.6}$$

where E is energy, $\Delta M_{\rm BH}$ is the increase in black hole mass, c is the speed of light, $\dot{M}_{\rm BH}$ is the black hole accretion rate, ϵ_1 is the accretion efficiency, $L_{\rm BH}$ is the bolometric energy output of the AGN, $L_{\rm cool}$ is the luminosity which would be produced if gas were allowed to cool and $\epsilon_{\rm SMBH}$ is the parameter defined in §6.2.2. We calculate the growth of the black hole as follows:

$$\dot{M}_{\rm BH} = L_{\rm cool}/\epsilon_1 c^2 \quad . \tag{6.7}$$

We take a value of the accretion efficiency for AGN fuelled by quasi-hydrostatic cooling flows, $\epsilon_1 = 0.016$, by choosing a value which is high enough to provide the significant feedback energy required for a good match to $z = 0$ galaxy properties, without requiring so much accretion onto black holes in central galaxies that they would have black hole masses significantly higher than the observed $M_{\rm BH}$ – bulge relations[2].

[2]In the published version of Bower et al. (2006), there was a numerical error. The results published in the paper were obtained using a value $\epsilon_1 = 0.016$, as stated in this Thesis Chapter, rather than the value

We fix our parameters $F_{\rm BH}$ and $\epsilon_{\rm stable\,disc}$ by requiring that the model reproduce both the observed $z = 0$ galaxy properties, and, in particular, the observed relations at $z = 0$ between black hole mass and bulge properties. We take $F_{BH} = 0.005$ and $\epsilon_{\rm stable\,disc} = 0.8$.

In Fig. 6.4, we show the global rate of increase in black hole mass due to accretion as a function of redshift, split up into the three channels of black hole growth outlined above. At $z > 1.4$, the global rate of black hole growth is dominated by disc instabilities – discs are far more likely to be unstable at high redshifts. Dark matter haloes of any given mass are smaller and denser at high redshift, so discs are more compact, and therefore more susceptible to gravitational instability. Furthermore, the cumulative effect of feedback means that discs have larger scale sizes by lower redshift. At $z < 1.4$, the global rate of black hole growth dominated by the accretion of cooling gas during the feedback phase. There are two reasons for this. Firstly, black holes are more massive at lower redshift, so have higher Eddington accretion rates, and therefore are able to accrete at higher rates in order to offset gas cooling. Secondly, more haloes have hot gas atmospheres which are in quasi-hydrostatic equilibrium at low redshift, both because high mass haloes are more abundant and also because haloes of any given mass are more likely to be in quasi-hydrostatic equilibrium due to the longer cooling times at lower redshift.

Accretion due to disc instabilities and accretion due to mergers are both associated with the burst mode of star formation. Both disc instabilities and mergers trigger starbursts. In the case of merger-driven starbursts, there will be a single starburst triggered by each merger, leading to a single episode of accretion. In the case of disc instability-driven starbursts, there will typically be a number of episodes of burst star formation, one at every timestep of the simulation, until the bulge formed from the starbursting gas is sufficiently massive to stabilize any newly formed disc: For any halo, at each timestep, gas is assumed to cool and form a disc of mass $M_{\rm disc}$. $\epsilon_{\rm disc} = V_{\rm max}/(GM_{\rm disc}/r_{\rm disc})^{1/2}$ is calculated at each timestep, and if $\epsilon_{\rm disc} < 0.8$, then we assume that the gas in the disc undergoes a starburst and is added to the bulge component. Thus the bulge component increases in mass at each timestep, and $V_{\rm max}$ (the circular velocity at the half-mass radius) increases at each timestep. This leads to a decrease in $\epsilon_{\rm disc}$ at each timestep (at least for constant $M_{\rm disc}$). Eventually, $\epsilon_{\rm disc}$ will go below the threshold of 0.8, and the newly formed disc will be stable. The newly formed disc is stable, gas continues to cool, and $M_{\rm disc}$ will increase – this may then lead to a further disc instability and starburst, which increases the mass of the bulge, once again stabilising the disc. Eventually, the central bulge will be

$\epsilon_1 = 0.2$, as is stated in the paper.

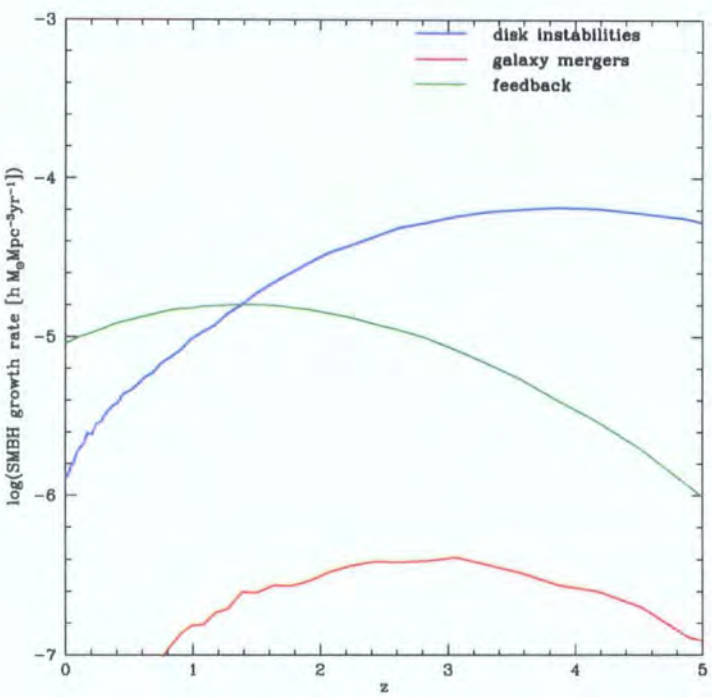

Figure 6.4: Contributions to the global SMBH accretion rate from the three growth channels – disc instabilities (blue line), merger driven starbursts (red line) and accretion of cooling gas in quasi-hydrostatic haloes (green line). This plot was taken directly from Bower et al. (2006), and was made by Richard Bower using a version of the `GALFORM` code modified by RKM to include black hole growth during merger-driven starbursts. Note that there is an error in the y-axis label – the units of SMBH growth rate should be $h^2 M_\odot \mathrm{Mpc}^{-3}\mathrm{yr}^{-1}$.

sufficiently massive that $V_{\rm max}$ is large enough to ensure that the disc is always stable when more mass is added to the disc. Fundamentally, our disc instability mechanism depends upon internal quantities relating to a single galaxy, its host halo and the gas cooling from that host halo. In contrast, the merger-triggered starburst mechanism is triggered by a merger of a galaxy with an external galaxy, and so is more stochastic.

At no redshift does growth by merger-driven starbursts dominate in the Bower et al. model. To a large extent, this is because mergers at high redshift are gas poor, due to the scaling of the quiescent star formation rate with the dynamical time of the disc. It may also largely be due to the relatively poor mass resolution of the N-body merger trees we use; there are not enough low mass galaxies available to trigger bursts via mergers, so nearly all of the star formation in bursts is in fact triggered by disc instabilities. Since black hole accretion in bursts depends on the star formation in exactly the same way irrespective of how those bursts were triggered, this means that the black hole mass accreted during bursts comes mainly from bursts triggered by disc instabilities. It is only by including black hole growth in starbursts triggered by disc instabilities that we are able to match the observed masses of black holes with a realistic value of $F_{\rm BH}$. In a similar model of galaxy formation with AGN feedback, Croton et al. (2006) used merger trees made from the same N-body simulation we are using. They decided not to allow black hole accretion in starbursts triggered by disc instabilities, and were forced to assume that in merger-triggered starbursts, an unrealistically high fraction ($\sim 30\%$) of the mass involved in starbursts is accreted onto a central supermassive black hole. As discussed in Chapter 2 of this Thesis, we believe that there may be a strong relationship between black hole growth, AGN activity and merger driven starbursts. Whilst there may indeed also be a strong contribution to the build-up of black hole mass from bursts triggered by disc instabilities, it seems likely that the Bower et al. model underestimates the contribution to the build-up of black hole mass in merger-driven starbursts. In future work, we hope to revisit this issue with the Bower et al. model and its variants using higher resolution merger trees.

6.3 The $M_{\rm BH}$ – bulge relations

As in Chapter 2, we set the parameters of the black hole growth model, in particular $F_{\rm BH}$, to match the zero-point of the observed $M_{\rm BH}$ – bulge relations. In Fig. 6.5, we show the relationship of black hole mass to a number of galaxy properties – the bulge mass, the

K-band bulge magnitude, the B-band bulge magnitude and the bulge velocity dispersion.

In Fig. 6.5 (a), we show the relationship between black hole mass and bulge mass. Our theoretical relation agrees very well with the data for bulges less massive than $2 \times 10^{11} h^{-1} M_{\odot}$, which have corresponding black hole masses less than $2 \times 10^{8} h^{-1} M_{\odot}$. Above this bulge and black hole mass however, the predicted relationship becomes significantly steeper than the best fitting slope of a linear fit to the $M_{\rm BH} - M_{\rm bulge}$ relationship. To some extent, there is an indication in the data to support an upturn in the $M_{\rm BH} - M_{\rm bulge}$ relationship at high black hole masses. The slope of the relationship found using the Bower et al. model is steeper than that found using the Baugh et al. model at the high mass end, and more closely matches the observed slope. As we suggested may be the case in Chapter 2, accretion of cooling gas onto black holes helps to steepen the slope of the $M_{\rm BH} - M_{\rm bulge}$ relation, and to attain a better fit to the data. We also note that there is a greater scatter in the $M_{\rm BH} - M_{\rm bulge}$ relation in the Bower et al. model than in the Baugh et al. model. In the Baugh et al. model, black hole accretion was exclusively tied to galaxy mergers, which also build up the stellar mass of the bulge, leading to a tight relationship particularly at high bulge stellar masses. However, in the Bower et al. model, black holes can also grow by accretion of cooling gas; star formation or the assembly of stars into a galactic bulge are not associated with this process. This increases the scatter of the $M_{\rm BH} - M_{\rm bulge}$ relation in the Bower et al. model, especially for more massive black holes, which have more growth due to accretion required to offset the cooling of gas.

We show in Fig. 6.5 (b) the relationship between black hole mass and bulge K-band luminosity, and in Fig. 6.5 (c) that between black hole mass and bulge B-band luminosity. In both cases, the slope of the relation is steeper than that found using the Baugh et al. model, and more closely matches the data. The scatter in both the $M_{\rm BH} - M_{\rm K,bulge}$ relation and the $M_{\rm BH} - M_{\rm B,bulge}$ relation is greater in the Bower et al. model than in the Baugh et al. model, especially for brighter bulges. This is presumably largely as a result of the increased scatter in the $M_{\rm BH} - M_{\rm bulge}$ relation, although there may also be a greater scatter in the star formation histories.

In Fig. 6.5 (d), we show the relationship between black hole mass and bulge velocity dispersion. Again, the slope of the relation we find is steeper than that found using the Baugh et al. model, and more closely matches the data. As noted by Wyithe (2006), it is difficult to know which, out of $M_{\rm BH}$ and $\sigma_{\rm bulge}$, to treat as the independent variable, and which to treat as the dependent variable. We have demonstrated in this plot that this makes a huge difference. There is a tail of galaxies in our model with unphysically large

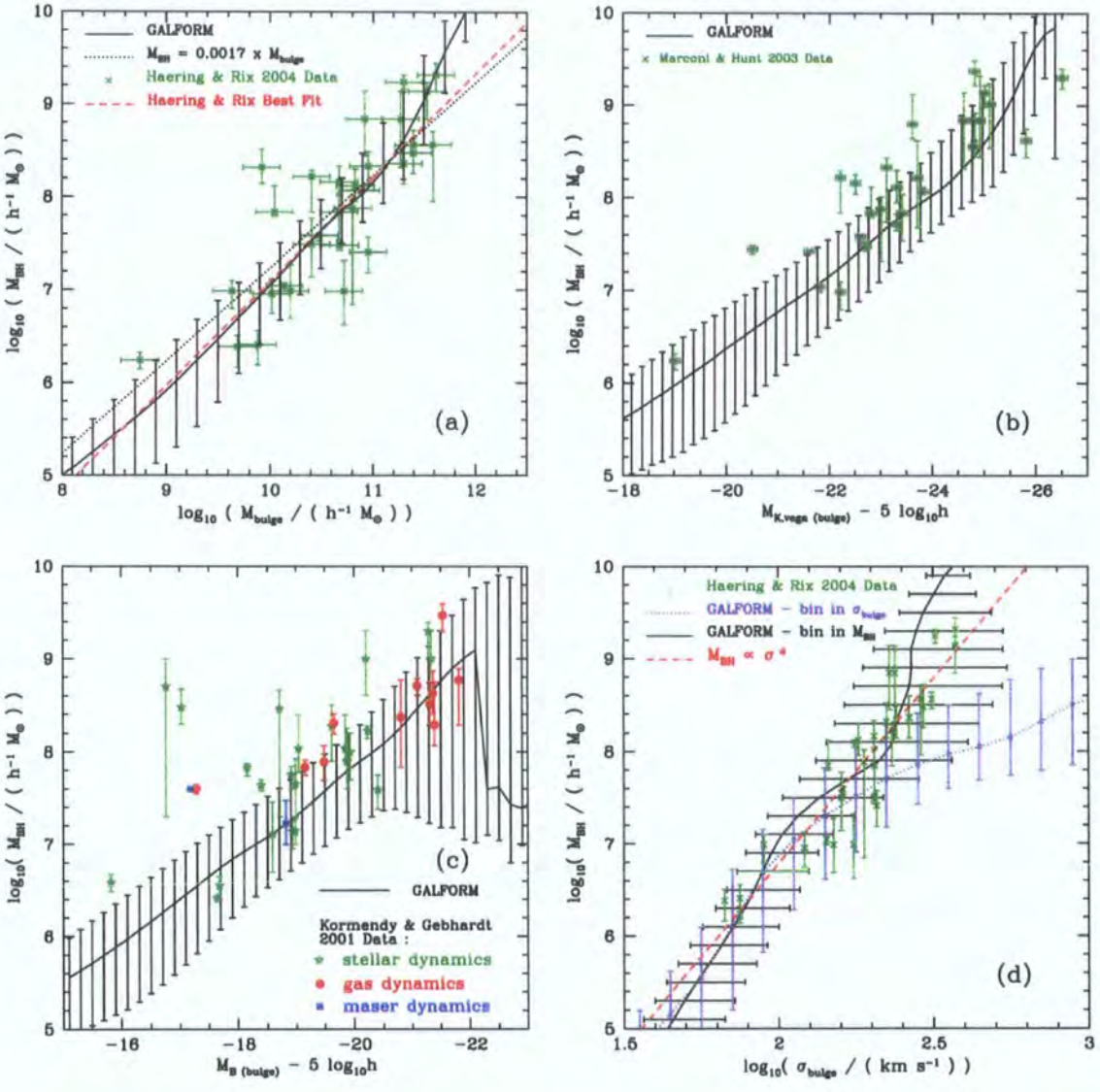

Figure 6.5: The relation between black hole mass, $M_{\rm BH}$, and a selection of properties of the spheroid of the host galaxy. Each panel shows the correlation with a different bulge property: (a) the stellar mass of the bulge; (b) the bulge rest-frame K-band magnitude; (c) the rest-frame B-band bulge magnitude; (d) the velocity dispersion of the bulge. The model predictions are shown by the line with errorbars: the line shows the median and the errorbars the 10–90 percentile spread of the distribution. The observational measurements are shown by symbols, with sources indicated in each panel.

velocity dispersions and tiny radii, possibly due to a problem with our model of adiabatic contraction.

We examined briefly whether the $M_{\rm BH} - M_{\rm bulge}$ relations might differ for central galaxies as opposed to the general galaxy population. However, we found that there was little, if any difference. Perhaps this is not so surprising. At the bright end of the galaxy bulge luminosity function, most galaxies are the central galaxy of their dark matter halo, whilst at the faint end, there has been little opportunity for accretion from a quasi-hydrostatic cooling flow, since fainter galaxies typically are hosted in lower mass dark matter haloes. In future work, we intend to study the environmental dependence of the $M_{\rm BH} - M_{\rm bulge}$ relations. In particular, we would like to see if the central galaxies of large groups and of clusters have black hole masses significantly above the mean relation. This may well be the case, since cooling must be prevented in large groups and clusters, possibly requiring significant accretion of mass onto the central black hole to provide the required feedback. Since the majority of the most massive galaxies in the Universe are central galaxies, this may lead directly to the upturn in the slope of the $M_{\rm BH} - M_{\rm bulge}$ relation for the most massive bulges.

6.4 The evolution of the quasar luminosity function

We only consider black hole accretion during starbursts (whether driven by mergers or disc instabilities) when calculating the quasar luminosity function. We assume that accretion during the quasi-hydrostatic cooling phase does not contribute to the optical luminosity function.

We assume that mass is accreted onto the black hole at a constant rate over a timescale proportional to the bulge dynamical time. We calculate the quasar luminosity according to the following procedure:

(i) From the mass of stars formed during the burst ($\Delta M_{\rm stars,burst}$), calculate the mass accreted onto the black hole during the starburst, $\Delta M_{\rm BH}$:

$$\Delta M_{\rm BH} = F_{\rm BH} \times \Delta M_{\rm stars,burst} \tag{6.8}$$

(ii) Calculate the timescale for black hole accretion, $t_{\rm accretion}$:

$$t_{\rm accretion} = \text{constant} \times \mathrm{t}_{\rm dyn,bulge} \tag{6.9}$$

(iii) Calculate the black hole accretion rate, $\dot{M}_{\rm BH}$:

$$\dot{M}_{\rm BH} = \Delta M_{\rm BH}/t_{\rm accretion} \qquad \text{for} \qquad t_{\rm burst} < t_{\rm accretion} \tag{6.10}$$

(iv) Integrate the black hole accretion rate to find the black hole mass as a function of time since the start of the burst, t:

$$M_{\rm BH}(t) = M_{\rm BH,initial} + \Delta M \times t_{\rm burst}/t_{\rm accretion} \tag{6.11}$$

(v) Find the luminosity predicted without any consideration of Eddington limiting:

$$L_{\rm no\ Eddington\ limiting}(t) = \epsilon_2 \dot{M}_{\rm BH}(t) c^2 \tag{6.12}$$

We note that the parameter for accretion efficiency during starbursts, ϵ_2, is not identical to that defined in §6.2.5 for the accretion efficiency during accretion of cooling gas, ϵ_1.

(vi) Find the Eddington luminosity of the black hole

$$L_{\rm Edd}(t) = M_{\rm BH,initial} + (4\pi c G \mu_e/\sigma_{\rm T}) \times M_{\rm BH}(t) = (4\pi c G \mu_e/\sigma_{\rm T}) M_{\rm BH}(t) \tag{6.13}$$

(vii) If the luminosity exceeds the Eddington limit, then set it equal to the Eddington limit:

$$L_Q = \min(\epsilon_2 \dot{M}_{\rm BH}(t) c^2, L_{\rm Edd}(t)) \tag{6.14}$$

We set the accretion efficiency and the scaling of the quasar lifetime with the bulge dynamical time to achieve the best possible match to the observed quasar luminosity function. We use an accretion efficiency, $\epsilon_2 = 0.05$, and a quasar lifetime equal to $t_{\rm dyn,bulge}$. However, these parameters are far from uniquely determined. For example, we are also able to achieve an equally reasonable match to the observed luminosity function using $\alpha\,\epsilon_2$ and a quasar lifetime equal to $\alpha\, t_{\rm dyn,bulge}$ for various values of α not too far from one.

The reasonable, although far from perfect, match to the observational data suggests that our model for black hole growth may be in at least fair agreement with that in the observed Universe. In comparison to the data, the quasar luminosity function in the Bower et al. model has a break that seems rather too sharp. This is the case at all redshifts where we have been able to make the comparison. It may be that this is related to our very simple model for black hole growth, where the black hole accretes at a constant rate during one bulge dynamical time. In future work, we hope to explore different models for evolution of quasar luminosity during a starburst. For example, it is likely that if we used an exponentially decaying quasar lightcurve (following Kauffmann & Haehnelt 2000), this would smooth out the sharp break, and bring the predicted quasar

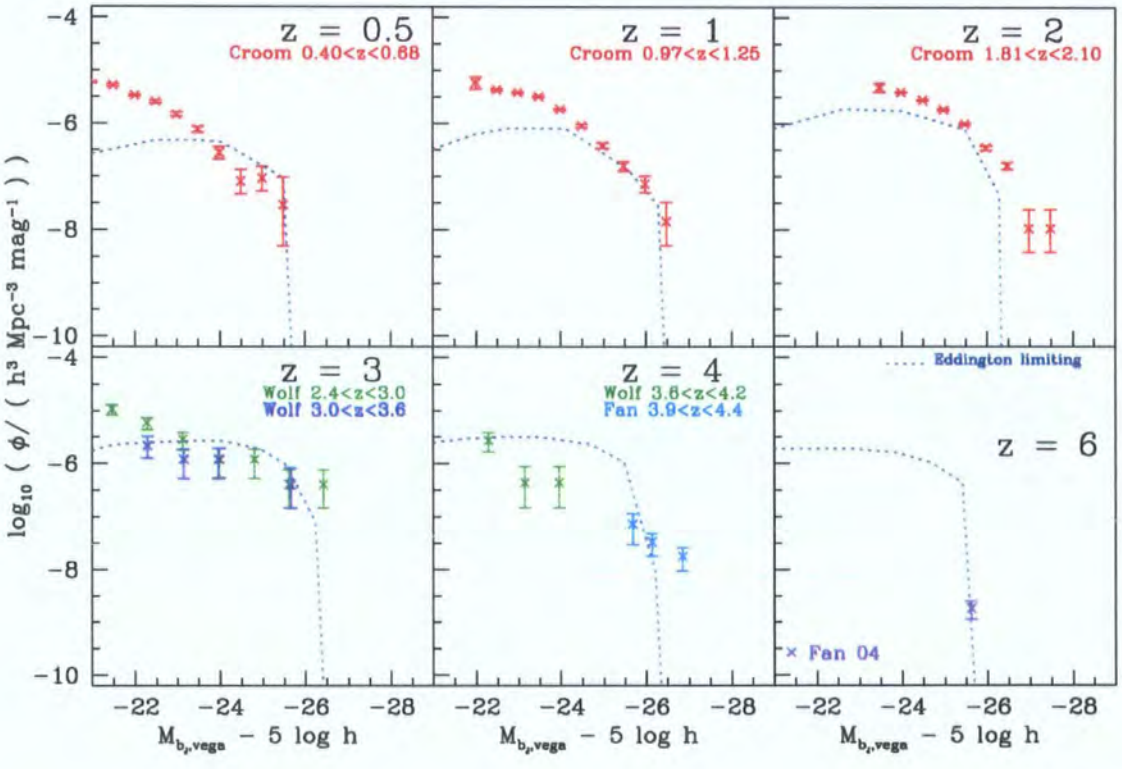

Figure 6.6: The evolution of the quasar luminosity function in the Bower et al. (2006) model. The model predictions are shown by lines and the data by symbols, with the source indicated in each panel. The model prediction includes Eddington limiting (as described in §6.4, which is identical to case (4) of §2.3.2). The data are taken from the following papers: Croom = Croom et al. 2004; Fan = Fan et al. 2001; Wolf = Wolf et al. 2003; Fan 04 = Fan et al. 2004.

luminosity function into better agreement with the observations. Ideally, we would like to try more realistic forms of the quasar lightcurve motivated by simulations of the fuel supply to black holes during galaxy mergers. In particular, we plan to try the shape of the quasar lightcurve proposed by Hopkins et al. (2005b) as a fit to the lightcurves they obtained from simulations.

The quasar luminosity functions we calculate only probe down to space densities of $\sim 10^{-8}h^{-3}\mathrm{Mpc}^3$, as can be seen in Fig. 6.6. This is due to the relatively small volume of the N-body simulation from which we obtained the dark matter halo merger trees. In future work, we hope to increase the effective volume of the calculation. Since large N-body volumes are expensive computationally, we may have to do this using extended Press-Schechter theory (c.f. Chapter 5).

We should clarify that we are using different efficiencies for AGN luminosity depending on whether the AGN is fuelled during a starburst (in which case we use $\epsilon_2 = 0.05$, as explained in §6.4) or whether it is fuelled by cooling gas in a quasi-hydrostatic cooling flow (in which case we use $\epsilon_2 = 0.016$, as explained in 6.2.5). Given that AGN in large groups and clusters are rarely seen as optical quasars, it is likely that our requirement of a lower AGN efficiency in AGN fuelled by cooling gas in quasi-hydrostatic cooling flows than in AGN fuelled by starbursts is indeed mirrored in the real Universe. Furthermore, higher AGN efficiency is more likely to be associated with accretion in the thin disc regime (Shakura & Sunyaev 1973) and hence with quasars observed optically, whilst lower AGN efficiency is more likely to be associated with accretion at a low fraction of the Eddington accretion rate, without much optical emission compared to X-ray and radio emission, as might be expected for AGN fuelled by cooling gas in quasi-hydrostatic cooling flows. It is therefore reasonable for us to calculate the optical luminosity function using only AGN fuelled by starbursts. In future work, we plan to extend our calculations of the quasar luminosity function to X-ray luminosities; X-ray observations are likely to provide a better census of all AGN activity in the Universe than optically selected quasar samples.

6.5 The evolution of the black hole mass function

At all redshifts, we see a significant knee in the black hole mass function at a mass of $10^8h^{-1}M_\odot$. Below this mass, the shape of the black hole mass function is fairly invariant with redshift. We believe that the main contribution to black hole mass for objects less massive than $10^8h^{-1}M_\odot$ is likely to be starbursts triggered by disc instabilities, although

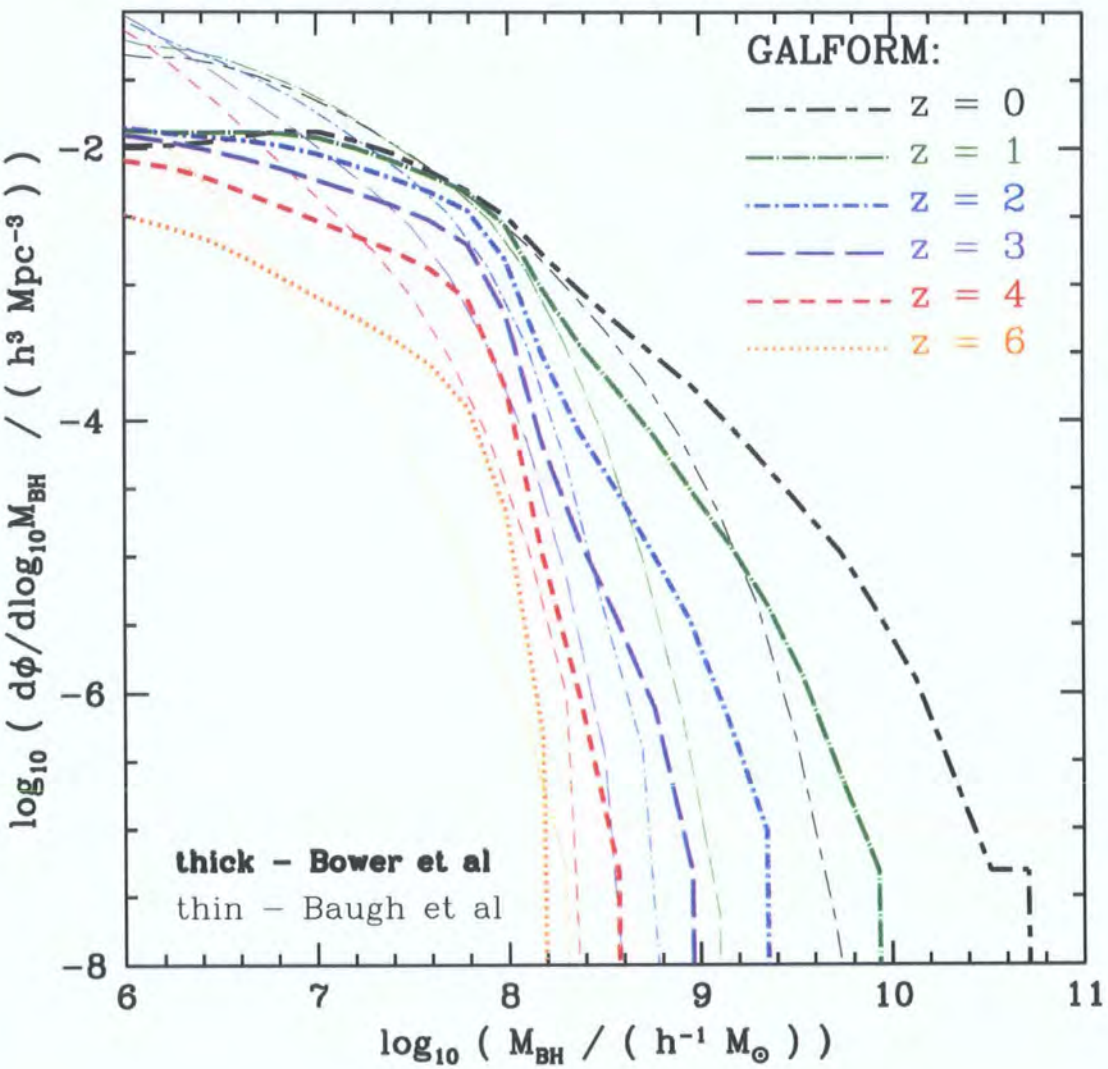

Figure 6.7: The evolution of the black hole mass function in the Bower et al. (2006) model (thick lines). Different colours and line types indicate different redshifts, as indicated in the key. Overplotted using thin lines is the mass function from the Baugh et al. (2005) model (which we showed in Chapter 3) for comparison.

we have yet to test this.

The slope of the black hole mass function for masses greater than $10^8 h^{-1} M_\odot$ rapidly becomes less steep as redshift decreases. This is because there is far more accretion onto higher mass black holes from gas in quasi-hydrostatic equilibrium, which must be prevented from cooling. More massive black holes are more likely to be found in higher mass dark matter haloes, which are more likely to be in the quasi-hydrostatic cooling regime. Furthermore, since the Eddington accretion rate increases in proportion to the black hole mass, the maximum black hole accretion rate allowed in the quasi-hydrostatic cooling regime increases in proportion to black hole mass (equation 6.1). Once black holes reach a mass of $10^8 h^{-1} M_\odot$, the major contribution to any further increases in mass is likely to be from the accretion of cooling gas.

We overplot with thin lines the black hole mass functions from the Baugh et al. model, which we also showed earlier in Fig. 3.3. In general, for black holes more massive than $10^8 h^{-1} M_\odot$, we find that black holes of any given space density tend to be more massive in the Bower et al. model than in the Baugh et al. model, and become increasingly more massive in comparison as redshift decreases. This is almost certainly due to the accretion of cooling gas which is included only in the Bower et al. model, and which becomes increasingly important towards low redshifts, especially for the most massive black holes.

When we specifically compare the $z = 6$ mass functions of the two models, we find that there is a higher abundance of massive black holes in the Bower et al. model, at least up to a mass of $10^8 h^{-1} M_\odot$. The abundance of black holes more massive than $10^8 h^{-1} M_\odot$ in the Bower et al. model because we are limited to a volume of $1.25 \times 10^8 h^{-3}\mathrm{Mpc}^3$, due to our use of N-body merger trees from the Millennium simulation. In future work, we intend to extend the analysis presented in Chapter 5 to the Bower et al. model, and to study how well we are able to reproduce the massive black holes at $z = 6$ implied by the Fan et al. (2001a, 2004) observations. In doing so, we will have to use semi-analytic merger trees to probe a sufficiently large volume.

6.6 The contribution to the black hole mass function from dark matter haloes of various masses

In Fig. 6.8, we show the contribution to the black hole mass function from dark matter haloes of various $z = 0$ masses, and from their progenitors at redshifts 1, 2, 3, 4 and 6. The shape of the black hole mass function is fairly independent of halo mass for black

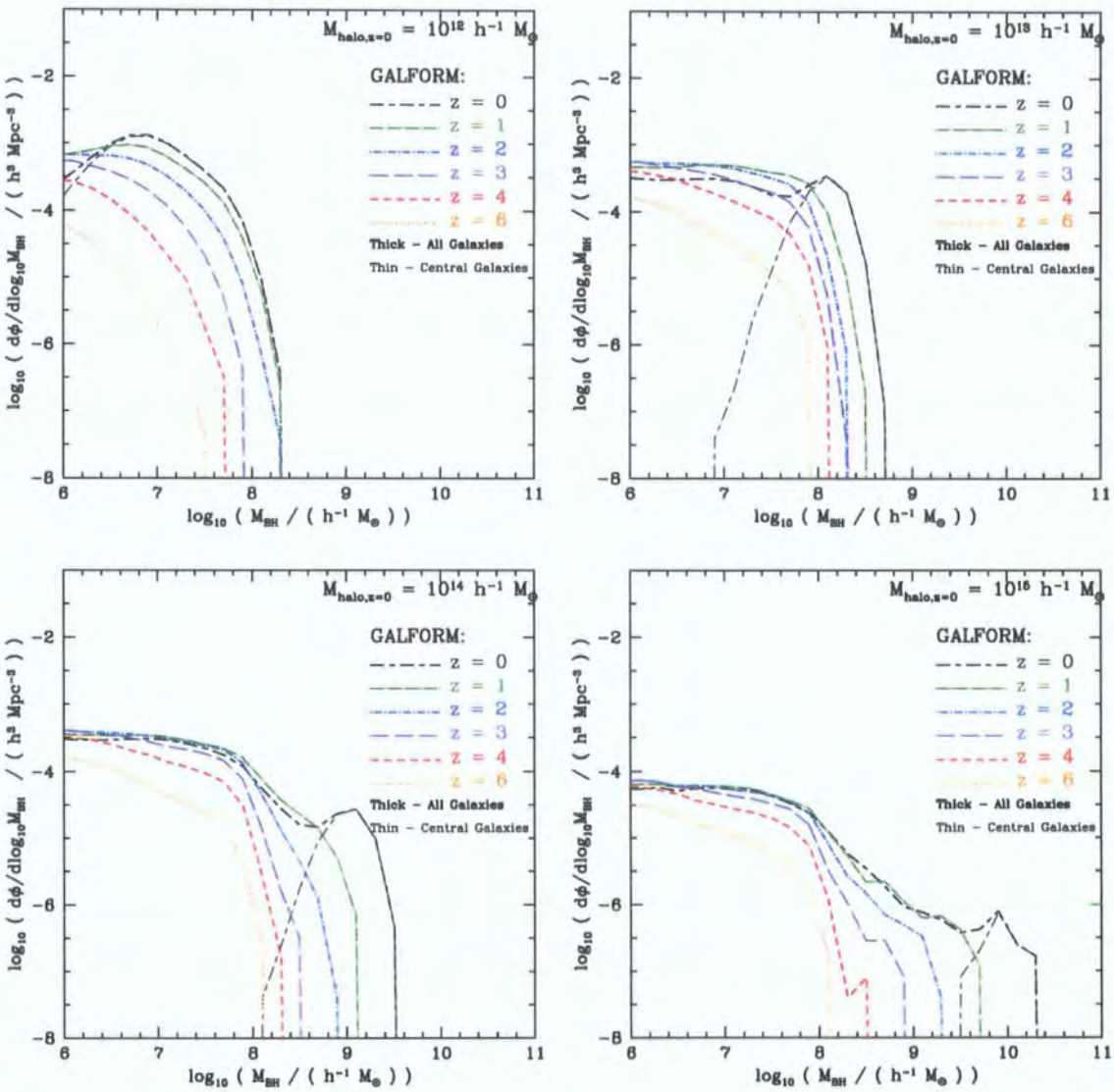

Figure 6.8: The contribution to the black hole mass function from dark matter haloes of various $z = 0$ masses, and from their progenitors at redshifts 1, 2, 3, 4 and 6. The redshift is shown by different colours and line types, as indicated in the keys. We show this for four different $z = 0$ dark matter halo masses: $10^{12}h^{-1}M_{\odot}$ (top-left), $10^{13}h^{-1}M_{\odot}$ (top-right), $10^{14}h^{-1}M_{\odot}$ (bottom-left) and $10^{15}h^{-1}M_{\odot}$ (bottom-right). In each case, the contribution to the black hole mass function is normalized to be the mass function per dex in $z = 0$ dark matter halo mass. For redshift zero, we also show (thin lines) the contribution to the mass function due to central galaxies only.

hole masses below $10^8 h^{-1} M_\odot$ (except at higher redshifts for the lowest dark matter halo mass shown). Presumably, this is because disc instabilities, which account for nearly all of the accretion onto black holes less massive than $10^8 h^{-1} M_\odot$, are fairly independent of halo mass. For black hole masses greater than $10^8 h^{-1} M_\odot$, the reduction in the slope of the black hole mass function as redshift decreases is not seen for the smallest dark matter halo mass shown ($10^{12} h^{-1} M_\odot$), but becomes increasingly obvious for larger masses of dark matter halo at $z = 0$, and is dramatic for the most massive $z = 0$ dark matter halo mass we consider ($10^{15} h^{-1} M_\odot$). This is because the growth of black hole mass due to accretion of cooling gas in quasi-hydrostatic haloes, which accounts for most of the black hole growth in black holes more massive than $10^8 h^{-1} M_\odot$, is much greater in higher mass dark matter haloes (and in their progenitors, which are also likely to be more massive). As expected, central galaxies contribute almost all of the objects at the high mass end of the black hole mass function for any given mass of dark matter halo.

6.7 Black hole growth through mergers and accretion

In Fig. 6.9 (top), we show the cumulative fraction of the mass assembled by mergers and accretion, as a function of final black hole mass at $z = 0$, for the Bower et al. model. In this Chapter, growth by accretion includes the contributions from disc instabilities, merger driven starbursts and accretion of cooling gas. We do not currently separate the growth by accretion into the contributions from these three channels, although it will be very interesting to do this. In the Bower et al. model, all masses of black hole grow primarily through direct accretion of gas onto their main progenitor. For all $z = 0$ black hole masses, there is a fairly small contribution to the final mass from mergers of pre-existing black holes onto the main progenitor. This is partly due to the poor mass resolution of the dark matter merger trees, which means that the contribution of mergers is reduced. For black holes more massive than $10^8 h^{-1} M_\odot$, the high contribution of direct accretion is largely due to the accretion of cooling gas in quasi-hydrostatically cooling haloes. This result can be compared to the same calculation for the Baugh et al. model, which was shown in Fig. 3.7 (top). Using the Baugh et al. model, black holes more massive than $10^8 h^{-1} M_\odot$ are built up predominantly from the mergers of pre-existing black holes. In both models, the least massive black holes are built up almost entirely by accretion onto a single progenitor.

In Fig. 6.9 (bottom), we show the merger rates of black holes exceeding various mass

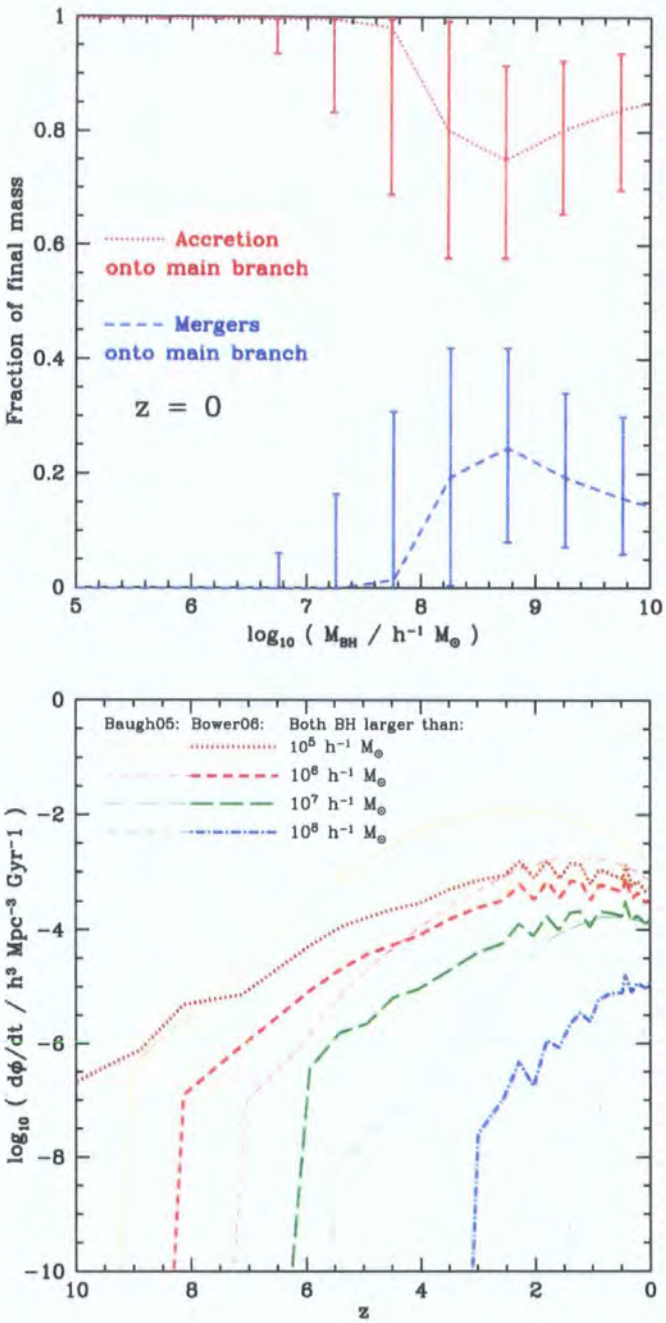

Figure 6.9: Top: The cumulative fraction of the mass assembled by mergers and accretion, as a function of final black hole mass at $z = 0$. This is shown for the Bower et al. model. The medians are connected by lines, and the 10–90 percentile spread of the distribution is shown as an errorbar for each black hole mass. Bottom: Black hole merger rate per unit time as a function of redshift. The merger rate is plotted for both the Bower et al. model and the Baugh et al. model, for 4 different mass thresholds which the (pre-starburst) masses of both black holes must exceed. The model used and the mass threshold are indicated in the key.

thresholds as a function of redshift. This is shown for both the Bower et al. and the Baugh et al. models. For the Bower et al. model, black hole merger rates increase as redshift decreases in general. However, particularly for lower mass black holes, there is a plateau or a slight decrease in the merger rates towards lower redshift. The increase in the merger rates towards lower redshift for the most massive black holes reflects the hierarchical assembly of structure; more massive objects are assembled later. For lower black hole masses ($M_{\rm BH} \sim 10^5 - 10^6 h^{-1} M_\odot$), we find that the merger rate is lower for the Bower et al. model than for the Baugh et al. model. Furthermore, as the redshift increases from redshift zero, then we see an increase in the BH-BH merger rate using the Baugh et al. model, but not when we use the Bower et al. model. These observations reflect the much greater importance in the Baugh et al. model of mergers of lower mass objects, including black holes, formed at high redshifts merging to form massive objects at low redshift. Furthermore, the merger rate of black holes of mass $\sim 10^5 h^{-1} M_\odot$ is comparatively low in the Bower et al. model since the resolution of the dark matter haloes is ~ 30 times poorer. We do not really trust that black holes less massive than $10^6 h^{-1} M_\odot$ are adequately resolved in the Bower et al. model.

We find that BH-BH merger rates in the Bower et al. model are comparable to those in the Baugh et al. model, and that in fact the merger rates of the most massive black holes ($M_{\rm BH} > 10^7 h^{-1} M_\odot$) in fact decline more slowly towards high redshift. This may at first seem counterintuitive given that a much higher fraction of the mass of the most massive black holes is assembled by mergers in the Baugh et al. model ($\sim 80 - 100\%$ – see Fig. 3.6) than in the the Bower et al. model ($\sim 5 - 40\%$ – see Fig 6.9). This apparent discrepancy can easily be explained by reference to the black hole mass functions in the two models (Fig. 6.7). In general, when we discuss 'assembly by accretion' and 'assembly by mergers' we have been referring to the *fraction* of the final black hole mass assembled by each of these channels. As shown in Fig. 6.7, black holes of any given space density are $\sim 1.5 - 10$ times more massive in the Bower et al. model. Thus assembly by mergers of a fairly small *fraction of the mass* of the most massive black holes in the Bower et al. model may in fact correspond to a very high *mass* assembled by mergers. Since, particularly for the most massive black holes, the space density is a very strong function of the black hole mass, black holes of any given mass have significantly higher space densities in the Bower et al. model. This gives a significant boost to the black hole merger rates, which we define as the merger rate of black holes which exceed some threshold in mass.

At low redshifts, we see 'saw-teeth' in the merger rates of black holes as a function

of redshift in the Bower et al. model, quite unlike the Baugh et al. model where the evolution of the BH-BH merger rate is smooth. We believe that this is probably some kind of numerical problem related to the very widely spaced timesteps of the N-body simulation from which we obtained the merger trees.

Ideally, would like to be able to make definitive predictions for the BH-BH merger rates for our alternative models of galaxy formation. It will then hopefully be possible to test these against the observed rate of BH-BH mergers in the Universe, derived in the future from gravitational wave observations from the LISA mission which should be launched in a few years.

The total black hole mass density at $z = 0$ is $\rho_{\rm BH} = 5.17 \times 10^5 h^3 M_\odot Mpc^{-3} = 1.77 \times 10^5 M_\odot Mpc^{-3}$. This is slightly less than that for the Baugh et al. model ($\rho_{\rm BH} = 2.83 \times 10^5 M_\odot Mpc^{-3}$, and slightly below the broad range spanned by the observational estimates. Observed values of $\rho_{\rm BH}/(10^5 M_\odot {\rm Mpc}^{-3})$, converted to $H_0 = 70 {\rm kms}^{-1}{\rm Mpc}^{-1}$, are : 2.9±0.5 (Yu & Tremaine 2002), 2.4±0.8 (Aller & Richstone 2002), 2.8±0.4 (McLure & Dunlop 2004), 4.2 ± 1.1 (Shankar et al. 2004) and $4.6^{+1.9}_{-1.4}$ (Marconi et al. 2004). Given the large errors in the observational estimates, our estimate of $\rho_{\rm BH}$ for the Bower et al. model may well be realistic. Our estimate is quite low largely because the space density of black holes less massive than $10^8 h^{-1} M_\odot$ is low, and we may well be able to increase it using merger trees of better mass resolution, which would allow us to simulate more realistically the low mass galaxies (or low mass bulges within more massive disc dominated galaxies) which would be likely to host low mass black holes. Furthermore, better mass resolution would allow us to increase the contribution to the black hole mass density from merger-driven starbursts (albeit at the expense of reducing the black hole mass density created during disc instabilities).

6.8 The evolution of the $M_{\rm BH}$ – bulge relations

In Fig. 6.10, we show the evolution with redshift of the relationships between black hole mass and various galaxy bulge properties: K-band and B-band bulge magnitude, bulge stellar mass and bulge velocity dispersion. In each case we plot the model predictions for the $M_{\rm BH}$ – bulge relationships at $z = 0$, 1, 2, 3, 4 and 6.

In Fig. 6.10 (a), we show the redshift evolution of the $M_{\rm BH} - M_{\rm B,bulge}$ relation, and in Fig. 6.10 (b), we show the redshift evolution of the $M_{\rm BH} - M_{\rm K,bulge}$ relation. Both show redshift evolution in the sense that, for a given black hole mass, the host bulge is more

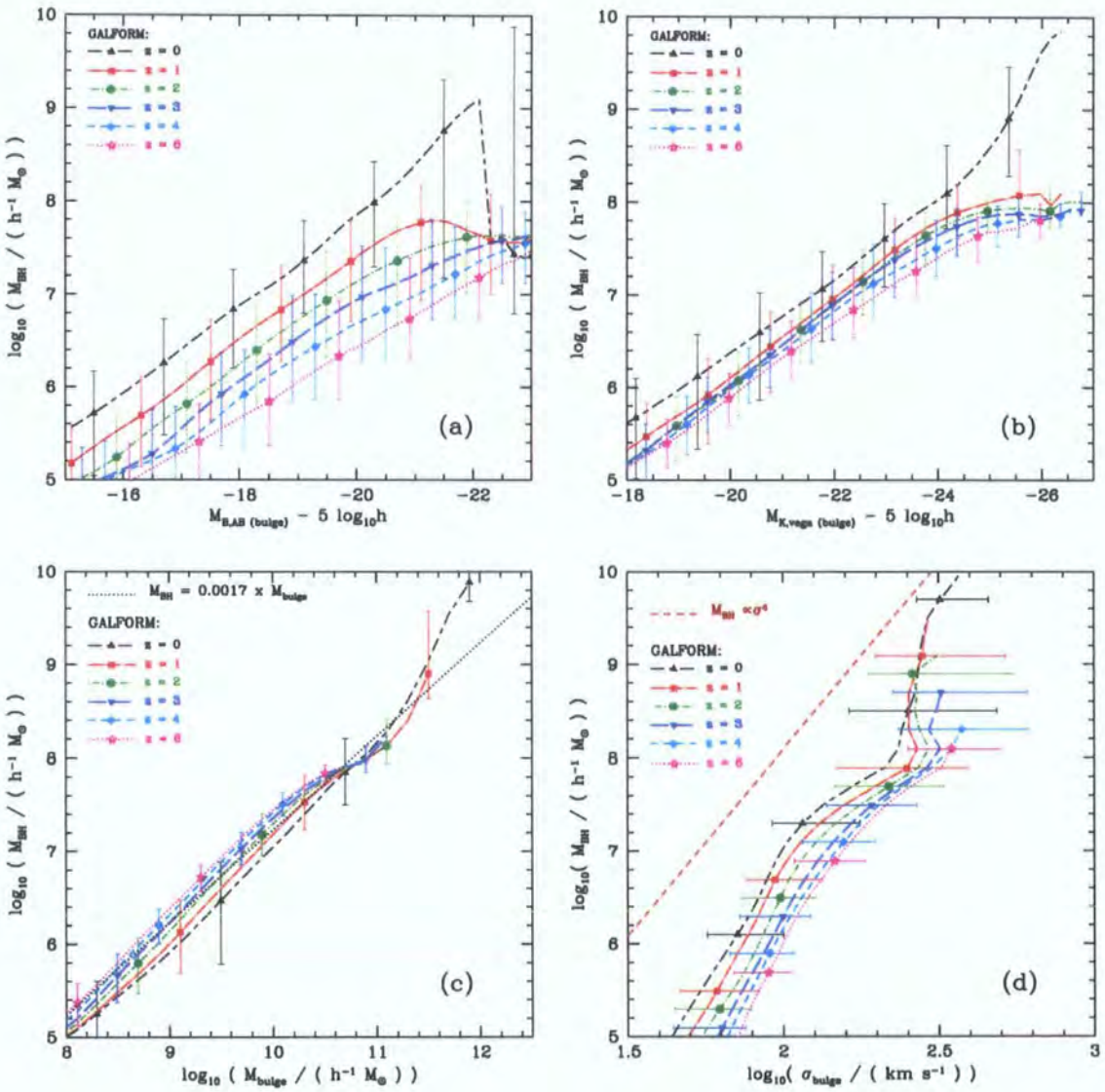

Figure 6.10: The redshift evolution of the relations between central black hole mass, $M_{\rm BH}$, and bulge properties for the Bower et al. (2006) model. Each panel shows the relationship between $M_{\rm BH}$ and a different property of the host spheroid : (a) the bulge rest-frame B-band magnitude; (b) the bulge rest-frame K-band magnitude; (c) the stellar mass of the bulge; (d) the velocity dispersion of the bulge. The model predictions are shown by the symbols with errorbars; the lines show the median relations and the errorbars the 10–90 percentile spread of the distributions. Redshifts 0, 1, 2, 3, 4 and 6 are shown in different line types, as indicated by the key.

luminous at high redshift. This trend is in the same direction as that found using the Baugh et al. model and simply reflects the dimming of stellar populations as they age. Stellar populations at low redshift tend to be older than those at high redshift. Compared to the predictions we made in Chapter 4 for the Baugh et al. model, we find that the evolution of the $M_{\rm BH} - M_{\rm B,bulge}$ relation is somewhat less pronounced in the Bower et al. model, and that the redshift evolution of the $M_{\rm BH} - M_{\rm K,bulge}$ relation is significantly less pronounced. For a given black hole mass, galactic bulges are only ~ 3 times more luminous in the K-band at $z = 6$ than at $z = 0$. To some extent, this reflects the reduced evolution in the $M_{\rm BH} - M_{\rm bulge}$ relation, which we will discuss next. Observationally, Peng et al. (2006), selecting high redshift quasars, find little trend in the $M_{\rm BH} - M_{\rm R,bulge,rest}$ relation with redshift. This suggests that our prediction of the evolution of the $M_{\rm BH} - M_{\rm K,bulge}$ relation may well be close to that in the real Universe (R-band and K-band magnitudes are very strongly correlated, especially for galaxies on the red sequence).

In Fig. 6.10 (c), we show the redshift evolution of the $M_{\rm BH} - M_{\rm bulge}$ relation for the Bower et al. model. There is very little evolution with redshift. In comparison, in the Baugh et al. model, there was greater evolution with redshift, particularly for lower mass bulges. Much more of the mass of the most massive galactic bulges is assembled via mergers of pre-existing stellar discs, leading to a shallower slope in the $M_{\rm BH} - M_{\rm bulge}$ relation. Thus there is a greater evolution with redshift of the normalization of the relation at lower bulge stellar masses in the Baugh et al. model than in the Bower et al. model, assuming that the zero-point is set using more massive bulges. Observationally, Peng et al. (2006) find that the ratio of $M_{\rm BH}/M_{\rm bulge}$ was 3–6 times larger at $z \gtrsim 2$ for AGNs than for quiescent galaxies at $z = 0$, whilst we find, using the Bower et al. model, that the $M_{\rm BH}/M_{\rm bulge}$ ratio is at most 1.5–2 times larger at $z \gtrsim 2$ than at $z = 0$. This may appear to conflict with the good agreement of the K-band/R-band relation from which Peng et al. estimated the $M_{\rm BH} - M_{\rm bulge}$ relation. The difference may well be simply one of conversion of galaxy magnitude to stellar mass. Peng et al. must assume an IMF and a star formation history for their galaxies in order to infer the stellar mass from the R-band magnitude, whereas the K-band magnitude we predict comes directly out of our model, and takes into account automatically both IMF and star formation history.

In Fig. 6.10 (d), we show the redshift evolution of the $M_{\rm BH} - \sigma_{\rm bulge}$ relation for the Bower et al. model. As was the case with the Baugh et al. model, we find that for a given black hole mass, bulges have a higher velocity dispersion at higher redshift. As with the Baugh et al. model, this to some extent reflects the expected variation in the

properties of dark matter haloes: at a given mass, the halo velocity dispersions scales as $\sigma \propto (z_{\rm form}+1)^{1/2}$. It can of course also be viewed as a reduction in the black hole mass with increasing redshift, for a fixed bulge velocity dispersion.

6.9 The redshift of black hole formation

In Fig. 6.11, we show the 'formation times' of black holes of various masses, for both the Bower et al. (left three panels) and the Baugh et al. (right three panels) models of galaxy formation. As discussed in detail in §3.6 of this Thesis, we define black hole 'formation' time/redshift in a number of ways. To recap, where the formation redshift is defined as the time when the *main progenitor* first exceeds a given fraction of the final mass, we refer to this as the *mass assembly* redshift (thin lines). Where the formation redshift is defined as the time when the *sum of all progenitors* first exceeds a given fraction of the final mass, we refer to this as the *mass transformation* redshift (thick lines). We consider three different mass fraction thresholds to define formation times: 0.01 (top), 0.5 (middle) and 0.95 (bottom). At low redshifts, we see 'saw-teeth' in the rate of formation of black holes as a function of redshift in the Bower et al. model. As with the black hole merger rates in the previous section, we believe this is a numerical problem related to the widely spaced outputs of the N-body merger trees.

We find, using the Bower et al. model, that the formation rates of black holes differ very little whether we consider assembly of the main progenitor, or whether we consider transformation of mass into black holes summed over all progenitors; the thin and thick lines in the plots nearly overlap, for all mass bins considered. This is unsurprising, given that we have already found that black holes grow mainly by accretion of gas onto the main progenitor, rather than by mergers of pre-existing black holes. In comparison, we find large differences in the formation rates in the Baugh et al. model for black hole masses above $10^7 h^{-1} M_{\odot}$ – massive black holes 'form' significantly earlier when we consider accretion onto any branch of the merger tree, than when we consider assembly of the main progenitor. This is to be expected since the most massive black holes at $z = 0$ grow largely through mergers of pre-existing black holes in this model.

We now examine in more detail the black hole formation rate in the Bower et al. model, where we consider the formation time[3] (middle-left panel of Fig. 6.11) of 50% of

[3]As we discussed earlier, it does not make much difference in the Bower et al. model whether we are considering mass transformation or mass assembly.

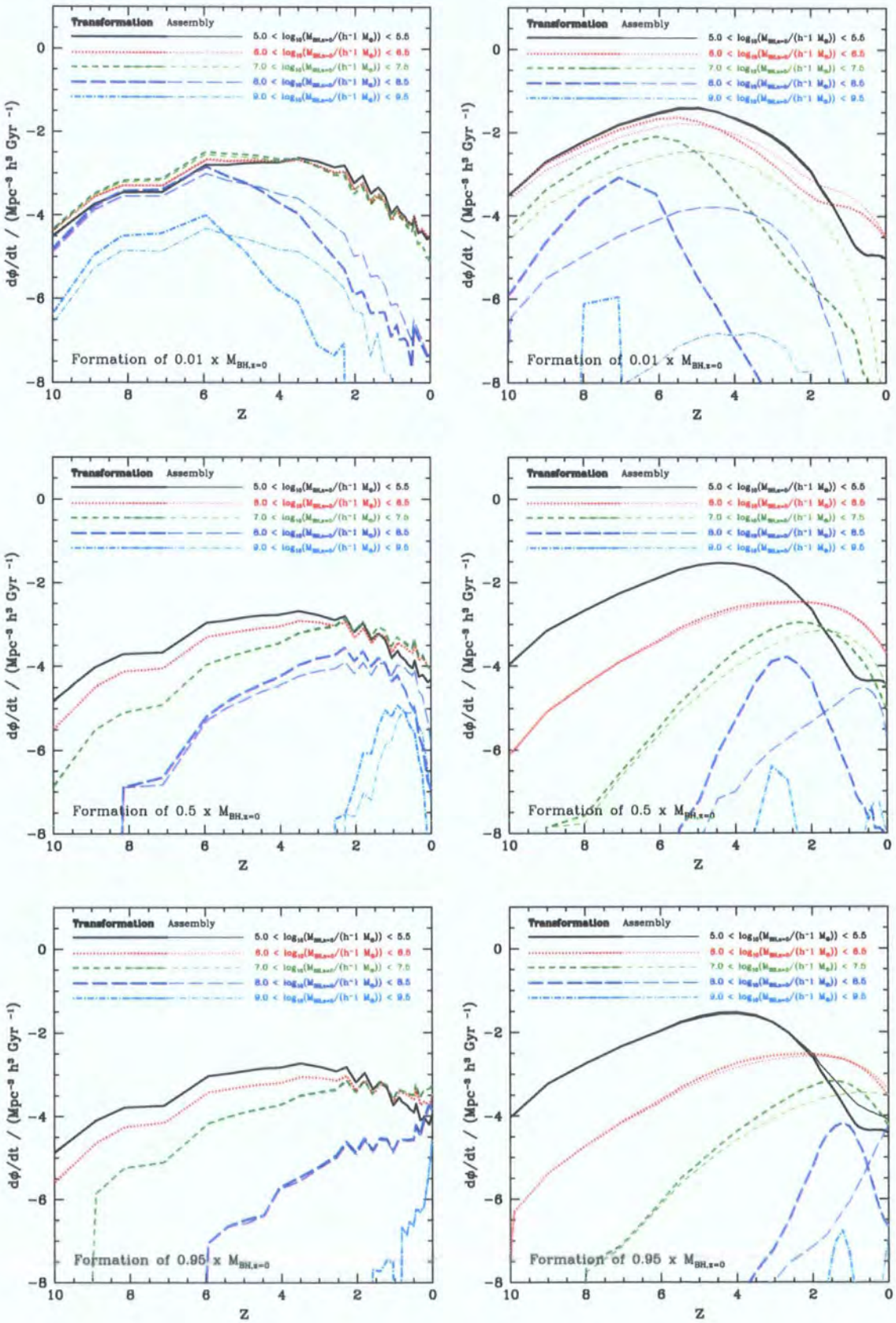

Figure 6.11: The distribution of formation redshifts of black holes in 5 different bins of $z = 0$ mass, as indicated by the key. The differing definitions of formation redshift are noted briefly on each plot and explained more fully in §6.9 of the text. We show results for the Bower et al. model (left), and the Baugh et al. model (right).

the final $z = 0$ black hole mass. When we examine the evolution from $z = 1$ to $z = 0$ of the black hole formation rate for black holes in various bins of $z = 0$ mass, we find that the formation rate of black holes in the mass range $10^9 - 10^{9.5} h^{-1} M_\odot$ declines far faster towards the present day than does the formation rate of black holes in the mass range $10^7 - 10^{7.5} h^{-1} M_\odot$. The evolution of the formation rate of $10^8 - 10^{8.5} h^{-1} M_\odot$ black holes is intermediate. This is an indirect indication of downsizing in the black hole population, at least for black holes of mass $10^7 < M_{\rm BH}/h^{-1} M_\odot < 10^{9.5}$.

6.10 Downsizing of black hole growth?

In Fig. 6.12, we show the ratio of the present accretion rate to the past average accretion rate ($< \dot{M} > \times t_{\rm age} / < M_{\rm BH} >$) as a function of black hole mass at $z = 0$, 1, 2, 3, 4 and 6. If this ratio exceeds unity, then the current mass accretion rate exceeds the average rate at which the black hole gained mass in the past (summed over all progenitors). We show this calculation for both the Bower et al. model (top) and the Baugh et al. model (bottom).

For the Bower et al. model, we see that the ratio of the present accretion rate to the past average accretion rate, at all redshifts, decreases as black hole mass increases. For any fixed black hole mass, this ratio decreases as redshift decreases. More massive black holes are growing more slowly than less massive ones, and low redshift black holes are growing more slowly than high redshift ones. The mass at which $< \dot{M} > \times t_{\rm age} / < M_{\rm BH} >$ crosses unity moves to progressively lower masses as redshift decreases – from $\sim 10^8 h^{-1} M_\odot$ at $z = 6$ to $\sim 10^6 h^{-1} M_\odot$ at $z = 0$. To some extent, this is a combination of the lower accretion rates of more massive black holes and of lower redshift black holes. We would like to factor out the overall decline in black hole formation rates as redshift decreases. Therefore, it is perhaps more telling that the break in $< \dot{M} > \times t_{\rm age} / < M_{\rm BH} >$ as a function of $M_{\rm BH}$ steadily moves to lower black hole masses as redshift decreases – from $\sim 3 \times 10^7 h^{-1} M_\odot$ at $z = 6$ to $\sim 10^7 h^{-1} M_\odot$ at $z = 0$. This is downsizing – black holes of progressively lower masses are being 'switched off' as redshift decreases.

In comparison, for the Baugh et al. model, we find that downsizing of black hole growth is in fact much more dramatic. Many recent papers (e.g. Bower et al. 2006; Cattaneo et al. 2006b; Croton et al. 2006; Kang et al. 2006) have made the claim that it is AGN feedback, in particular AGN feedback from low efficiency accretion, which prevents cooling from hot haloes of quasi-hydrostatic gas in massive dark matter haloes, which is the pivotal

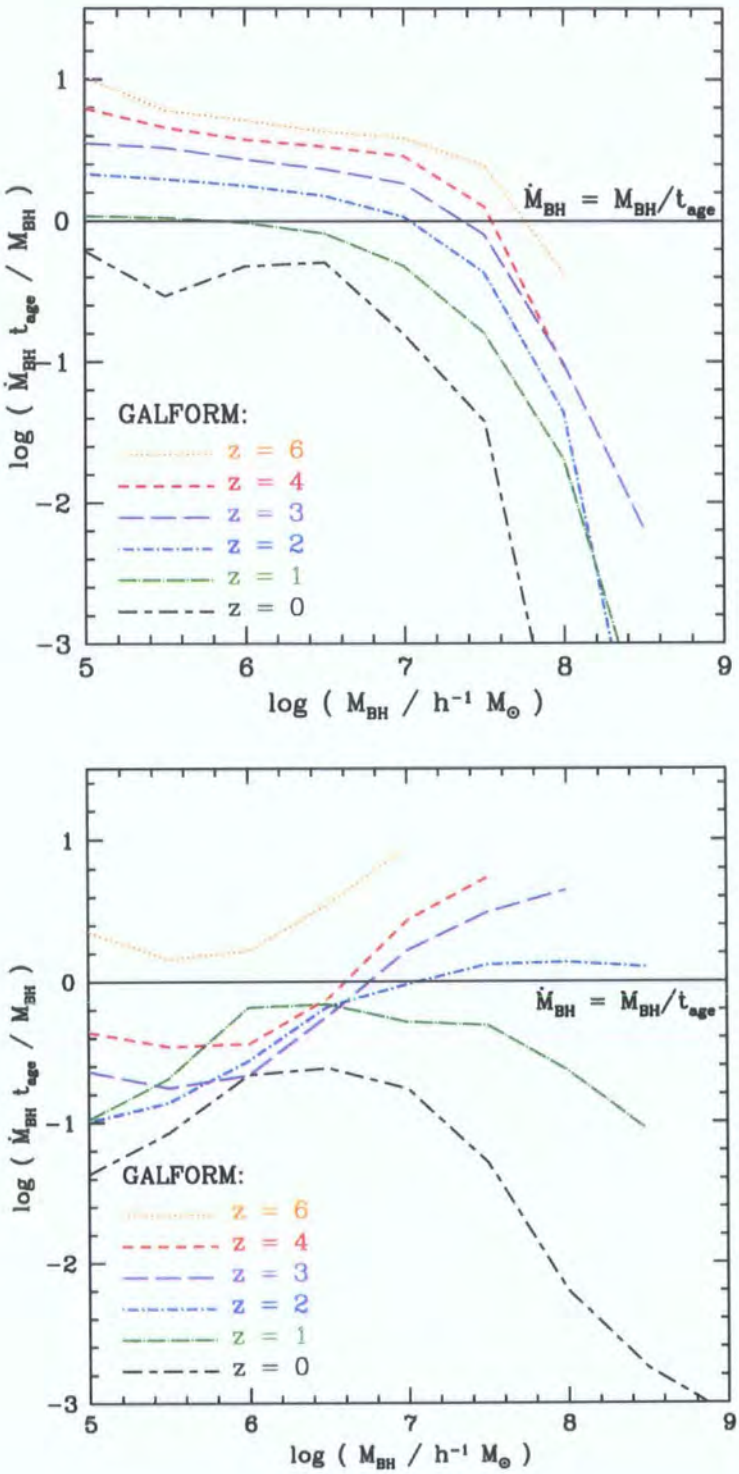

Figure 6.12: The current mean mass accretion rate normalized by the past average mass growth rate, plotted against black hole mass. This is shown for redshifts 0, 1, 2, 3, 4 and 6, as indicated in the keys, and for both the Bower et al. model (top) and the Baugh et al. model (bottom).

ingredient required to reproduce downsizing. This is obviously not true. We do not argue that AGN feedback is not required in our models[4], but we do argue that it is not the sole physical process which causes downsizing (e.g. see Mouri & Taniguchi 2006; Neistein et al. 2006).

However, it must be remembered that the sole arbiter of 'which model produces better predictions of downsizing' is the comparison of model predictions with observed data. We remind the reader that many of the black holes in our model (in particular, the Baugh et al. model) are not accreting at all, and so could not be observed as AGN – only accreting objects are included in observational inferences of downsizing. We must make a realistic determination of the luminosities which would be observed from accreting black holes, particularly their X-ray luminosities, since downsizing of black hole accretion has been best observed in the X-rays. This will be the most convincing way to test which of our models agrees best with the inferred trend of downsizing.

6.11 Discussion

The Baugh et al. and Bower et al. models of galaxy formation both have their strengths and weaknesses. As far as the black hole and AGN population goes, we are able to achieve a better match to the observed $M_{\rm BH}$ – bulge relations using the Bower et al. model, and a better match to the observed quasar luminosity function using the Baugh et al. model. As far as galaxy formation goes, both models reproduce the properties of $z = 0$ galaxies equally well, in particular the luminosity function. At higher redshift, the Bower et al. model reproduces the evolution of the stellar mass function and K-band luminosity function, whilst the Baugh et al. model better reproduces the properties of high redshift starbursts, in particular Lyman-break galaxies and sub-mm detected galaxies. We hope soon to combine the best aspects of both models.

In the future, the astronomy community should be very careful how it uses the terminology 'AGN feedback'. It seems very likely that there are two fairly distinct regimes of feedback from black holes. There is low efficiency accretion of gas cooling from quasi-hydrostatic haloes of hot gas, as was used in the Bower et al. model, and as was simulated by e.g. Dalla Vecchia et al. (2004). There is also AGN feedback during starbursts, as has

[4]e.g. we do believe it is required to prevent luminous galaxies being too blue, to make the shut-off in star formation when galaxies move from the blue sequence to the red sequence and quite probably to accentuate any downsizing which is in the model before addition of AGN feedback.

been simulated by e.g. Springel et al. (2005a), which we do not yet include in either the Bower et al. model or the Baugh et al. model.[5] The subject of AGN feedback in galaxy formation is rapidly evolving, and there is still a lot of confusion when astronomers talk about 'AGN feedback' – do they refer to AGN feedback due to the strong inflows of gas onto a central black hole triggered by a galaxy merger (e.g. Springel et al. 2005a), or to low efficiency accretion of cooling gas in massive quasi-hydrostatic hot gas haloes (e.g. Dalla Vecchia et al. 2004)? For example, astronomers have often referred to the simulations of Springel et al. of galaxy mergers in order to motivate the inclusion of AGN feedback due to low efficiency accretion in quasi-hydrostatic hot gas haloes. We hope in the future to clarify the differences between these two regimes of AGN feedback and their effects.

It is somewhat difficult to determine the effect of AGN feedback alone in a comparison of the Baugh et al. and Bower et al. models. Partly for historical reasons, the two models differ in many ways besides the simple inclusion of AGN feedback. It can be difficult to isolate the effect of a single ingredient in a semi-analytical model of galaxy formation, at least when we impose the requirement that a model accurately reproduce the properties of $z = 0$ galaxies. For example, Benson et al. (2003a) developed a feedback model which accounted very well for the $z = 0$ K-band luminosity function. They then showed the effect of switching off different types of feedback one by one, and found that the resulting models had a highly unrealistic $z = 0$ K-band luminosity function. We hope to create a model which explicitly includes both modes of AGN feedback discussed above, and, as far as possible, to disentangle their specific effects on the evolution of the galaxy and AGN population of the Universe.

[5]However, it is likely that AGN feedback during starbursts is a major contributor to the kinetic energy input which expels gas from dark matter haloes, i.e. superwinds, which the Baugh et al. model feedback scheme relies upon extensively.

Chapter 7 | *Conclusions*

7.1 Summary and Discussion

We have described an extension to the GALFORM semi-analytical model of galaxy formation in the ΛCDM cosmology to track the growth of black holes (BH). Our basic model for black hole growth, as described in Chapter 2 and in Malbon et al. (2006), has one free parameter, $f_{\rm BH}$. $f_{\rm BH}$ is the mass accreted onto the black hole expressed as a fraction of the stellar mass produced during a starburst. We set the value of $f_{\rm BH}$ so as to reproduce the zeropoint of the present day $M_{\rm BH}$ – bulge relations. The slope, scatter and evolution of the $M_{\rm BH}$ – bulge relations are model predictions.

In our basic model, black holes grow only during and following a galaxy merger. They grow through two distinct channels: mergers of pre-existing black holes and accretion of cold gas if a starburst is triggered by the merger. The importance of growth through black hole mergers increases with the mass of the black hole; at $z = 0$ the growth of black holes less massive than $5 \times 10^7 h^{-1} M_\odot$ is dominated by accretion, while the growth of more massive black holes is dominated by mergers. In general, the growth of black hole mass by mergers becomes more important at low redshifts as the supply of gas available for accretion is consumed by star formation.

Essentially all current observational estimates of the accumulation of black hole mass are only sensitive to *luminous* growth, i.e. mass accretion. However, we predict that the importance of growth through BH-BH mergers grows with *decreasing* redshift and with increasing black hole mass. BH-BH mergers represent a *dark* mode of growth that is more difficult to observe and confirm. The most obvious way to detect BH-BH mergers is through the emission of gravitational waves (e.g. Haehnelt 1994). This may be possible in ten years with the planned LISA gravitational wave interferometer. When gas is present during a black hole merger, a circumbinary accretion disc could form and the BH merger may produce high velocity gas outflows (Armitage & Natarajan 2002) followed by an X-ray afterglow which could be detected by the next generation of X-ray observatories (Milosavljević & Phinney 2005). Winged or X-shaped radio sources (Merritt & Ekers

2002) and cores in elliptical galaxies (Faber et al. 1997; Milosavljević et al. 2002) may be indirect evidence of gas-poor mergers.

Our model predicts that the most important growth mechanism for the most massive black holes is the 'dark' mode or mass assembly through BH-BH mergers. A testable prediction of our model is that a tail of black holes with masses above a few times $10^9 M_\odot$ should be found at $z = 0$ once high quality observations covering a large volume of the local Universe become available (§3.2). Furthermore, we expect that these black holes will be more massive than any found in quasars at high redshift. To date, only black holes less massive than $\sim 3 \times 10^9 M_\odot$ have been unambiguously observed in galaxies at $z = 0$ (Tremaine et al. 2002) and in luminous, optically-selected quasars over the redshift interval $0 < z < 2$ (McLure & Dunlop 2004). However, this implied limit on black hole mass is far from robust. The most massive black holes at $z = 0$ tend to reside in massive and hence rare elliptical galaxies, which could easily have been missed in existing surveys. Larger volumes ($\gtrsim$ a few times $10^6 h^{-3}\mathrm{Mpc}^3$) need to be surveyed to find such objects, which are therefore likely to lie at large distances. This, coupled with their expected low surface brightness (more massive ellipticals tend to have lower surface brightness cores), could make it difficult to measure their central mass using methods based on stellar dynamics (Kormendy & Gebhardt 2001). The high redshift quasar data do not give a complete census of the black hole populations, as quasar observations are only able to probe accreting black holes.

There is an important distinction to be made in our model between *mass transformation* and *mass assembly*. Mass transformation refers to the process of turning cold gas into black hole mass; at any one time in the formation history of a black hole, this phenomenon could be occuring across a number of progenitor black holes. Mass assembly refers to the accumulation of mass in a black hole's main progenitor, and may occur via both direct accretion of gas and merging of pre-existing black holes. In the Baugh et al. model, black hole mass *assembles* hierarchically; more massive black holes are assembled at lower redshifts than less massive black holes. However, if we choose to define the formation time of a black hole in terms of the mass *transformation* redshift when some fraction of its mass has been *accreted* onto *any* progenitor, we find that, for $M_{\mathrm{BH}} > 10^7 h^{-1} M_\odot$, more massive black holes form earlier. This dichotomy mirrors the growth of stellar mass in galactic spheroids in hierarchical models. In the semi-analytical models, galaxy mergers produce spheroids. At high redshift, the mergers tend to be gas rich and new stars are produced as a result of the merger event. At low redshift, galactic discs tend to be gas

poor and consist mainly of stars, with the result that the merger simply rearranges the pre-existing stars (Baugh et al. 1996; Kauffmann 1996; Bell et al. 2006; De Lucia et al. 2006).

While we find that black hole mass is assembled hierarchically, our model clearly exhibits a 'downsizing' in the mass of black holes which are undergoing luminous accretion. At the present day, we find that low mass black holes are accreting material at a higher proportion of their Eddington luminosity than high mass black holes. This distinction is less apparent at higher redshifts. Another way to demonstrate this downsizing is to examine the rate at which black holes are accreting mass, expressed as a fraction of the mass already in place: $< \dot{M}_{\rm BH} > / < M_{\rm BH} >$. At $z = 0$, $\dot{M}_{\rm BH}/M_{\rm BH}$ is largest for low mass black holes and drops rapidly with increasing mass. In the Baugh et al. model, there is a very strong trend with increasing redshift; accretion becomes an increasingly important mode of mass assembly for all masses of black hole at earlier epochs in the Universe. The trend with redshift is far less dramatic in the Bower et al. model, although we still see that accretion onto progressively lower mass black holes is shut-off as redshift decreases towards the present.

A number of authors have claimed that black holes grow in an 'anti-hierarchical' fashion (Marconi et al. 2004; Merloni 2004; Shankar et al. 2004). This conclusion is reached by comparing an inferred present day black hole mass function with the black hole mass function expected from AGN relics under the assumption that black holes grow exclusively by accretion. These calculations ignore any contribution to the mass of black holes arising from BH-BH mergers. Furthermore, the assumption that all black holes accrete at a constant fraction of their Eddington ratio (Marconi et al. 2004; Shankar et al. 2004), or the use by Merloni (2004) of the 'fundamental plane of black hole activity' (Merloni et al. 2003), which has a very large scatter, will introduce errors that may become cumulatively very large as the black hole mass function is integrated backwards in time.

While the zeropoints of the present day $M_{\rm BH}$ – bulge relations are set by adjusting a single free parameter, $f_{\rm BH}$. The slope, scatter and evolution of these relations are genuine predictions of the model. We find little evolution with redshift in the slope of any of the $M_{\rm BH}$ – bulge relations, although our model predicts differing evolution in their zeropoints, depending upon which particular bulge property is being considered. If we focus attention on a fixed black hole mass, we find that with increasing redshift, the typical host bulge is more luminous in the rest-frame B-band and K-band, shows little change in stellar mass (except in the case of low mass black holes, where the stellar mass is lower) and has a

somewhat higher velocity dispersion.

Our model predicts the presence of massive black holes at high redshift, and we have studied this extensively. Fan et al. have discovered quasars of magnitude $M_{1450} \sim -27$ at $z \sim 6$, albeit at a low space density, $\phi \approx 1.6 \pm 0.5 \times 10^{-10} h^3 \mathrm{Mpc}^{-3}$ (Fan et al. 2004). Assuming that these objects are radiating at the Eddington luminosity with an Elvis et al. (1994) spectrum, and that beaming and gravitational lensing are insignificant, Fan et al. (2001a) inferred that these quasars host black holes masses of $\sim 1 - 3 \times 10^9 h^{-1} M_\odot$. In this Thesis, we probe the mass function of $z = 6$ black holes down to a space density of $\phi \sim 1 \times 10^{-9} h^3 \mathrm{Mpc}^{-3}$, which is just about sparse enough to compare with the data of Fan et al. There is still some uncertainty in the background cosmology, and we find it much easier to match the observed space density of quasars in a model with a spectral index $n_s = 1$, a higher value of σ_8 and a higher value of Ω_m (Spergel et al. 2003) than an alternative cosmological model which also appears to fit the observational data with $n_s < 1$, a lower value of σ_8 and a lower value of Ω_m (Sánchez et al. 2006). If this second cosmological model is indeed closer to that of the real Universe, then we may need to make some major changes in our model in order to match the early growth of massive objects in the Universe which is observed.

It is often argued that the most massive black holes and the brightest quasars at $z \sim 6$ reside exclusively in the most massive black holes. We find that this is not the case. In our model, there is a strong trend of increasing black hole mass and increasing quasar luminosity, but this trend has a scatter of $\sim$ an order of magnitude. Since the dark matter mass function tails off extremely fast with increasing mass at the halo masses most relevant to quasar and black hole formation, this means that the typical dark matter halo hosts of the most massive black holes and of the brightest quasars have masses $\sim 10^{12} M_\odot$, rather than $\sim 10^{13} M_\odot$ as is commonly assumed.

The model we use throughout most of this Thesis (Chapters 2-5) and which we have referred to so far in this summary, the Baugh et al. (2005) model, does not consider the effect of feedback from black hole growth on galaxy formation. In Chapter 6, we examined black hole growth in a second model of galaxy formation, that of Bower et al. (2006). The Bower et al. model includes AGN feedback to stifle cooling in dark matter haloes with quasi-hydrostatic atmospheres of hot gas. This process may well impart the physical scale required to match the break in the observed luminosity function of galaxies. Two new channels of black hole growth are added – growth during starbursts triggered by disc instabilities, which dominates at higher redshifts and for black holes less massive than

$10^8 h^{-1} M_\odot$, and the growth required to fuel AGN feedback, which dominates at lower redshifts and for black holes more massive than $10^8 h^{-1} M_\odot$.

Contrasting the two models, we find a better match to the $M_{\rm BH}-$bulge relations using the Bower et al. model, and a better match to the evolution of the quasar luminosity function using the Baugh et al. model. In the Bower et al. model, black holes of all masses at $z = 0$ grow primarily by accretion onto a single progenitor, although there is still a noticeable contribution from mergers for black holes of mass at $z = 0$ greater than $3 \times 10^7 h^{-1} M_\odot$. Surprisingly perhaps, the Bower et al. model predicts BH-BH merger rates which are comparable to those expected in the Baugh et al. model, and merger rates which are actually higher both for $M_{\rm BH} > 10^8 h^{-1} M_\odot$ and, at higher redshifts, lower mass black holes. Comparing the predictions for black hole downsizing in the two models, we find a greater downsizing of black hole activity in the Baugh et al. model than we do in the Bower et al. model, although we have not yet determined which model is actually a better fit to the data. This belies the recent consensus in the community that black hole feedback in quasi-hydrostatic cooling flows is the pivotal ingredient required to produce downsizing in galaxy formation models, although it is undoubtedly still very important in solving many of the problems traditionally faced by hierarchical galaxy formation models.

In summary, we have presented a new model for the concurrent growth of galaxies and black holes in the ΛCDM cosmology. We have previously shown that this model can successfully account for many observed properties of the galaxy population over a large range of wavelengths, from the local optical and infrared galaxy luminosity function to the number counts of submillimeter galaxies and the UV luminosity function of Lyman-break galaxies at redshift $z \sim 3$ (Baugh et al. 2005). We have also briefly contrasted this model with that of Bower et al. (2006), which includes black hole feedback and we have previously shown to reproduce successfully the evolution of the stellar mass function and the K-band luminosity function. In this Thesis we have focussed on the properties of the black hole population that grows in unison with the spheroidal component of the galaxies. The models can both account for a variety of observables that involve black holes, such as the relationship between the mass of the central black hole in galaxies and the properties of the bulge, the quasar luminosity function, the presence of luminous quasars when the Universe was only 1Gyr old and the apparent 'anti-hierarchical' growth of black holes. The models may be tested by future observations of the evolution of the $M_{\rm BH}$ – bulge relations and, perhaps, by the detection of gravitational waves associated with the mergers of massive black holes that play a prominent role in both of our models.

7.2 Future work

We finally outline a number of exciting applications of our combined model of black hole and galaxy formation which we hope to pursue, especially now that we are able to include black hole feedback :

• We plan to make predictions of the AGN X-ray luminosity function. Particularly at faint luminosities, there is likely to be a significant contribution from low efficiency black hole accretion of cooling gas in quasi-hydrostatic hot haloes. The X-ray luminosity function is likely to provide a more complete census of AGN activity in the Universe, whereas the optical luminosity function is more affected by obscuration.

• We plan to make predictions of the radio luminosity function of quasars. The most luminous AGN in the radio tend to be in the centre of groups and clusters, where the feedback mode of accretion is most relevant. Indeed, the radio luminosity may be strongly related to the feedback luminosity required to prevent cooling in the intracluster or intragroup medium (e.g. Best et al. 2006).

• Whilst the spectrum of the X-ray background can be fairly well explained for energies less than the peak at 30keV by ensembles of black hole sources accreting at fairly high fractions of the Eddington rate, it is likely to be very difficult to match the shape of the spectra at higher energies without a significant fraction of the accretion in the Universe being at very low fractions of the Eddington rate (Chris Done, priv. comm.). Therefore, it is very likely that the feedback mode of accretion in the Bower et al. model will be of substantial help in matching the spectrum of the very hard X-ray background, and that the Baugh et al. model, where all accretion is triggered by starbursts, will struggle.

• The inclusion of black hole feedback at low redshifts in the Baugh et al. model will perhaps enable us to reproduce the $z = 0$ luminosity function, whilst reducing the ejection of gas from haloes by superwinds at high redshift. We may thus be able to produce more massive, gas-rich bursts at high redshift, which would allow us to reproduce the observed luminosity functions of Lyman break galaxies and number counts of sub-mm galaxies with a more realistic IMF in bursts than the extremely top-heavy one we are currently using. We may also introduce ingredients of the Baugh et al. model into the Bower et al. model. Our ultimate goal is to produce a model of galaxy formation which reproduces both the evolution of stellar mass, as does the Bower et al. model *and* the properties of high redshift starbursts, as does the Baugh et al. model.

In the end, we would like to find the model of galaxy formation which best describes galaxy formation in the Universe. As we showed earlier, the predictions we make of black hole merger rates will not enable us to distinguish reliably between the models on their own. It will therefore be necessary for us to extend the models to a wide variety of observable properties of galaxies and of AGN, and to compare these to the rapidly increasing selection of observations available, across the entire electomagnetic spectrum, as well as, we hope, of gravitational waves.

Bibliography

Abel, T., Bryan, G. L., & Norman, M. L.: 2002, *Science* **295**, 93

Abramowicz, M. A.: 2005, in A. Merloni, S. Nayakshin, & R. A. Sunyaev (eds.), *Growing Black Holes: Accretion in a Cosmological Context*, pp 257–273

Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M.: 2003, ApJ **584**, 45

Alexander, D. M., Bauer, F. E., Brandt, W. N., Hornschemeier, A. E., Vignali, C., Garmire, G. P., Schneider, D. P., Chartas, G., & Gallagher, S. C.: 2003, AJ **125**, 383

Alexander, D. M., Smail, I., Bauer, F. E., Chapman, S. C., Blain, A. W., Brandt, W. N., & Ivison, R. J.: 2005a, Nature **434**, 738

Alexander, D. M., Smail, I., Bauer, F. E., Chapman, S. C., Blain, A. W., Brandt, W. N., & Ivison, R. J.: 2005b, Nature **434**, 738

Aller, M. C. & Richstone, D.: 2002, AJ **124**, 3035

Almeida, C., Baugh, C. M., & Lacey, C. G.: 2006, *ArXiv Astrophysics e-prints*

Angulo, R., Baugh, C. M., & et al. in prep.: 2006

Archibald, E. N., Dunlop, J. S., Jimenez, R., Friaça, A. C. S., McLure, R. J., & Hughes, D. H.: 2002, MNRAS **336**, 353

Armitage, P. J. & Natarajan, P.: 2002, ApJ **567**, L9

Baker, J. G., Boggs, W. D., Centrella, J., Kelly, B. J., McWilliams, S. T., Miller, M. C., & van Meter, J. R.: 2007, *ArXiv Astrophysics e-prints*

Baker, J. G., Centrella, J., Choi, D.-I., Koppitz, M., van Meter, J. R., & Miller, M. C.: 2006, ApJ **653**, L93

Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., & Szalay, A. S.: 2004, ApJ **600**, 681

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S.: 1986, ApJ **304**, 15

Barger, A. J., Cowie, L. L., Mushotzky, R. F., Yang, Y., Wang, W.-H., Steffen, A. T., & Capak, P.: 2005, AJ **129**, 578

Barnes, J. E. & Hernquist, L. E.: 1991, ApJ **370**, L65

Barth, A. J., Martini, P., Nelson, C. H., & Ho, L. C.: 2003, ApJ **594**, L95

Baugh, C. M.: 2006, *Rep. Prog. Phys.*

Baugh, C. M., Cole, S., & Frenk, C. S.: 1996, MNRAS **283**, 1361

Baugh, C. M., Lacey, C. G., Frenk, C. S., Granato, G. L., Silva, L., Bressan, A., Benson, A. J., & Cole, S.: 2005, MNRAS **356**, 1191

Becker, R. H., Fan, X., White, R. L., Strauss, M. A., Narayanan, V. K., Lupton, R. H., Gunn, J. E., Annis, J., Bahcall, N. A., Brinkmann, J., Connolly, A. J., Csabai, I., Czarapata, P. C., Doi, M., Heckman, T. M., Hennessy, G. S., Ivezić, Ž., Knapp, G. R., Lamb, D. Q., McKay, T. A., Munn, J. A., Nash, T., Nichol, R., Pier, J. R., Richards, G. T., Schneider, D. P., Stoughton, C., Szalay, A. S., Thakar, A. R., & York, D. G.: 2001, AJ **122**, 2850

Begelman, M. C.: 1978, MNRAS **184**, 53

Begelman, M. C.: 2002, ApJ **568**, L97

Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D.: 2003, ApJS **149**, 289

Bell, E. F., Naab, T., McIntosh, D. H., Somerville, R. S., Caldwell, J. A. R., Barden, M., Wolf, C., Rix, H.-W., Beckwith, S. V., Borch, A., Häussler, B., Heymans, C., Jahnke, K., Jogee, S., Koposov, S., Meisenheimer, K., Peng, C. Y., Sanchez, S. F., & Wisotzki, L.: 2006, ApJ **640**, 241

Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Page, L., Spergel, D. N., Tucker, G. S., Wollack, E., Wright, E. L., Barnes, C., Greason, M. R., Hill, R. S., Komatsu, E., Nolta, M. R., Odegard, N., Peiris, H. V., Verde, L., & Weiland, J. L.: 2003, ApJS **148**, 1

Benson, A. J., Bower, R. G., Frenk, C. S., Lacey, C. G., Baugh, C. M., & Cole, S.: 2003a, ApJ **599**, 38

Benson, A. J., Cole, S., Frenk, C. S., Baugh, C. M., & Lacey, C. G.: 2000, MNRAS **311**, 793

Benson, A. J., Frenk, C. S., Baugh, C. M., Cole, S., & Lacey, C. G.: 2003b, MNRAS **343**, 679

Benson, A. J., Kamionkowski, M., & Hassani, S. H.: 2005, MNRAS **357**, 847

Benson, A. J., Lacey, C. G., Baugh, C. M., Cole, S., & Frenk, C. S.: 2002, MNRAS **333**, 156

Berlind, A. A. & Weinberg, D. H.: 2002, ApJ **575**, 587

Best, P. N., Kaiser, C. R., Heckman, T. M., & Kauffmann, G.: 2006, MNRAS **368**, L67

Bett, P., Eke, V., Frenk, C. S., Jenkins, A., Helly, J., & Navarro, J.: 2007, MNRAS pp 60–+

Binney, J.: 2004, MNRAS **347**, 1093

Binney, J. & Tremaine, S.: 1987, *Galactic dynamics*, Princeton, NJ, Princeton University Press, 1987, 747 p.

Birnboim, Y. & Dekel, A.: 2003, MNRAS **345**, 349

Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N.: 1991, ApJ **379**, 440

Borys, C., Smail, I., Chapman, S. C., Blain, A. W., Alexander, D. M., & Ivison, R. J.: 2005, ApJ **635**, 853

Bower, R. G.: 1991, MNRAS **248**, 332

Bower, R. G., Benson, A. J., Malbon, R., Helly, J. C., Frenk, C. S., Baugh, C. M., Cole, S., & Lacey, C. G.: 2006, MNRAS **370**, 645

Boyle, B. J. & Terlevich, R. J.: 1998, MNRAS **293**, L49

Bromley, J. M., Somerville, R. S., & Fabian, A. C.: 2004, MNRAS **350**, 456

Bromm, V. & Loeb, A.: 2003, ApJ **596**, 34

Bruzual, G. & Charlot, S.: 2003, MNRAS **344**, 1000

Bruzual A., G. & Charlot, S.: 1993, ApJ **405**, 538

Bryan, G. L. & Norman, M. L.: 1997, in D. A. Clarke & M. J. West (eds.), *ASP Conf. Ser. 123: Computational Astrophysics; 12th Kingston Meeting on Theoretical Astrophysics*, pp 363–+

Campanelli, M., Lousto, C. O., Zlochower, Y., & Merritt, D.: 2007a, *ArXiv General Relativity and Quantum Cosmology e-prints*

Campanelli, M., Lousto, C. O., Zlochower, Y., & Merritt, D.: 2007b, *ArXiv General Relativity and Quantum Cosmology e-prints*

Carr, B. J., Bond, J. R., & Arnett, W. D.: 1984, ApJ **277**, 445

Cattaneo, A.: 2001, MNRAS **324**, 128

Cattaneo, A., Blaizot, J., Weinberg, D. H., Colombi, S., Dave, R., Devriendt, J., Guiderdoni, B., Katz, N., & Keres, D.: 2006a, *ArXiv Astrophysics e-prints*

Cattaneo, A., Combes, F., Colombi, S., Bertin, E., & Melchior, A.-L.: 2005, MNRAS **359**, 1237

Cattaneo, A., Dekel, A., Devriendt, J., Guiderdoni, B., & Blaizot, J.: 2006b, MNRAS **370**, 1651

Chandrasekhar, S.: 1943, ApJ **97**, 255

Chapman, S. C., Smail, I., Windhorst, R., Muxlow, T., & Ivison, R. J.: 2004, ApJ **611**, 732

Cole, S., Aragon-Salamanca, A., Frenk, C. S., Navarro, J. F., & Zepf, S. E.: 1994, MNRAS **271**, 781

Cole, S. & Kaiser, N.: 1989, MNRAS **237**, 1127

Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S.: 2000, MNRAS **319**, 168

Cole, S., Norberg, P., Baugh, C. M., Frenk, C. S., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Cross, N., Dalton, G., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Glazebrook, K., Jackson, C., Lahav, O.,

Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Peacock, J. A., Peterson, B. A., Sutherland, W., & Taylor, K.: 2001, MNRAS **326**, 255

Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., Baugh, C. M., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Cross, N. J. G., Dalton, G., Eke, V. R., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Glazebrook, K., Jackson, C., Jenkins, A., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Peterson, B. A., Sutherland, W., & Taylor, K.: 2005, MNRAS **362**, 505

Colless, M., Dalton, G., Maddox, S., Sutherland, W., Norberg, P., Cole, S., Bland-Hawthorn, J., Bridges, T., Cannon, R., Collins, C., Couch, W., Cross, N., Deeley, K., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Jackson, C., Lahav, O., Lewis, I., Lumsden, S., Madgwick, D., Peacock, J. A., Peterson, B. A., Price, I., Seaborne, M., & Taylor, K.: 2001, MNRAS **328**, 1039

Collin, S. & Kawaguchi, T.: 2004, A&A **426**, 797

Comerford, J. M., Haiman, Z., & Schaye, J.: 2002, ApJ **580**, 63

Cooray, A. & Sheth, R.: 2002, Phys. Rep. **372**, 1

Cowie, L. L., Barger, A. J., Bautz, M. W., Brandt, W. N., & Garmire, G. P.: 2003, ApJ **584**, L57

Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G.: 1996, AJ **112**, 839

Croom, S. M., Smith, R. J., Boyle, B. J., Shanks, T., Miller, L., Outram, P. J., & Loaring, N. S.: 2004, MNRAS **349**, 1397

Croton, D. J.: 2006, MNRAS **369**, 1808

Croton, D. J., Gao, L., & White, S. D. M.: 2007, MNRAS **374**, 1303

Croton, D. J., Springel, V., White, S. D. M., De Lucia, G., Frenk, C. S., Gao, L., Jenkins, A., Kauffmann, G., Navarro, J. F., & Yoshida, N.: 2006, MNRAS **365**, 11

Cuoco, A., Iocco, F., Mangano, G., Miele, G., Pisanti, O., & Serpico, P. D.: 2004, *International Journal of Modern Physics A* **19**, 4431

Dalla Vecchia, C., Bower, R. G., Theuns, T., Balogh, M. L., Mazzotta, P., & Frenk, C. S.: 2004, MNRAS **355**, 995

de Jong, R. S. & Lacey, C.: 2000, ApJ **545**, 781

De Lucia, G., Springel, V., White, S. D. M., Croton, D., & Kauffmann, G.: 2006, MNRAS **366**, 499

Dekel, A. & Cox, T. J.: 2006, MNRAS pp 692–+

Di Matteo, T., Springel, V., & Hernquist, L.: 2005, Nature **433**, 604

Dicke, R. H., Peebles, P. J. E., Roll, P. G., & Wilkinson, D. T.: 1965, ApJ **142**, 414

Dunlop, J. S., McLure, R. J., Kukula, M. J., Baum, S. A., O'Dea, C. P., & Hughes, D. H.: 2003, MNRAS **340**, 1095

Efstathiou, G., Lake, G., & Negroponte, J.: 1982, MNRAS **199**, 1069

Efstathiou, G., Moody, S., Peacock, J. A., Percival, W. J., Baugh, C., Bland-Hawthorn, J., Bridges, T., Cannon, R., Cole, S., Colless, M., Collins, C., Couch, W., Dalton, G., de Propris, R., Driver, S. P., Ellis, R. S., Frenk, C. S., Glazebrook, K., Jackson, C., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Norberg, P., Peterson, B. A., Sutherland, W., & Taylor, K.: 2002, MNRAS **330**, L29

Efstathiou, G. & Rees, M. J.: 1988, MNRAS **230**, 5P

Einstein, A.: 1916, *Annalen der Physik* **49**, 769

Eke, V. R., Cole, S., & Frenk, C. S.: 1996, MNRAS **282**, 263

Eke, V. R., Frenk, C. S., Baugh, C. M., Cole, S., Norberg, P., Peacock, J. A., Baldry, I. K., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Dalton, G., de Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Glazebrook, K., Jackson, C. A., Lahav, O., Lewis, I., Lumsden, S., Maddox, S. J., Madgwick, D., Peterson, B. A., Sutherland, W., & Taylor, K.: 2004, MNRAS **355**, 769

Elvis, M., Wilkes, B. J., McDowell, J. C., Green, R. F., Bechtold, J., Willner, S. P., Oey, M. S., Polomski, E., & Cutri, R.: 1994, ApJS **95**, 1

Enoki, M., Nagashima, M., & Gouda, N.: 2003, PASJ **55**, 133

Faber, S. M., Tremaine, S., Ajhar, E. A., Byun, Y.-I., Dressler, A., Gebhardt, K., Grillmair, C., Kormendy, J., Lauer, T. R., & Richstone, D.: 1997, AJ **114**, 1771

Fall, S. M. & Efstathiou, G.: 1980, MNRAS **193**, 189

Fan, X., Carilli, C. L., & Keating, B.: 2006a, ARA&A **44**, 415

Fan, X., Hennawi, J. F., Richards, G. T., Strauss, M. A., Schneider, D. P., Donley, J. L., Young, J. E., Annis, J., Lin, H., Lampeitl, H., Lupton, R. H., Gunn, J. E., Knapp, G. R., Brandt, W. N., Anderson, S., Bahcall, N. A., Brinkmann, J., Brunner, R. J., Fukugita, M., Szalay, A. S., Szokoly, G. P., & York, D. G.: 2004, AJ **128**, 515

Fan, X., Narayanan, V. K., Lupton, R. H., Strauss, M. A., Knapp, G. R., Becker, R. H., White, R. L., Pentericci, L., Leggett, S. K., Haiman, Z., Gunn, J. E., Ivezić, Ž., Schneider, D. P., Anderson, S. F., Brinkmann, J., Bahcall, N. A., Connolly, A. J., Csabai, I., Doi, M., Fukugita, M., Geballe, T., Grebel, E. K., Harbeck, D., Hennessy, G., Lamb, D. Q., Miknaitis, G., Munn, J. A., Nichol, R., Okamura, S., Pier, J. R., Prada, F., Richards, G. T., Szalay, A., & York, D. G.: 2001a, AJ **122**, 2833

Fan, X., Narayanan, V. K., Strauss, M. A., White, R. L., Becker, R. H., Pentericci, L., & Rix, H.-W.: 2002, AJ **123**, 1247

Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., & Fukugita, M.: 2006b, AJ **132**, 117

Fan, X., Strauss, M. A., Richards, G. T., Newman, J. A., Becker, R. H., Schneider, D. P., Gunn, J. E., Davis, M., White, R. L., Lupton, R. H., Anderson, Jr., J. E., Annis, J., Bahcall, N. A., Brunner, R. J., Csabai, I., Doi, M., Fukugita, M., Hennessy, G. S., Hindsley, R. B., Ivezić, Ž., Knapp, G. R., McKay, T. A., Munn, J. A., Pier, J. R., Szalay, A. S., & York, D. G.: 2001b, AJ **121**, 31

Favata, M., Hughes, S. A., & Holz, D. E.: 2004, ApJ **607**, L5

Ferrarese, L. & Merritt, D.: 2000, ApJ **539**, L9

Fitchett, M. J.: 1983, MNRAS **203**, 1049

Freedman, W. L., Madore, B. F., Gibson, B. K., Ferrarese, L., Kelson, D. D., Sakai, S., Mould, J. R., Kennicutt, Jr., R. C., Ford, H. C., Graham, J. A., Huchra, J. P., Hughes, S. M. G., Illingworth, G. D., Macri, L. M., & Stetson, P. B.: 2001, ApJ **553**, 47

Frenk, C. S., White, S. D. M., & Davis, M.: 1983, ApJ **271**, 417

Fukugita, M., Hogan, C. J., & Peebles, P. J. E.: 1998, ApJ **503**, 518

Gallego, J., Zamorano, J., Aragon-Salamanca, A., & Rego, M.: 1995, ApJ **455**, L1+

Gao, L., Springel, V., & White, S. D. M.: 2005a, MNRAS **363**, L66

Gao, L., White, S. D. M., Jenkins, A., Frenk, C. S., & Springel, V.: 2005b, MNRAS **363**, 379

Gardner, J. P., Sharples, R. M., Frenk, C. S., & Carrasco, B. E.: 1997, ApJ **480**, L99+

Gebhardt, K., Bender, R., Bower, G., Dressler, A., Faber, S. M., Filippenko, A. V., Green, R., Grillmair, C., Ho, L. C., Kormendy, J., Lauer, T. R., Magorrian, J., Pinkney, J., Richstone, D., & Tremaine, S.: 2000, ApJ **539**, L13

Geller, M. J. & Huchra, J. P.: 1989, *Science* **246**, 897

Ghez, A. M., Salim, S., Hornstein, S. D., Tanner, A., Lu, J. R., Morris, M., Becklin, E. E., & Duchêne, G.: 2005, ApJ **620**, 744

Gingold, R. A. & Monaghan, J. J.: 1977, MNRAS **181**, 375

Glazebrook, K., Peacock, J. A., Miller, L., & Collins, C. A.: 1995, MNRAS **275**, 169

Gonzalez, J. A., Hannam, M. D., Sperhake, U., Brugmann, B., & Husa, S.: 2007, *ArXiv General Relativity and Quantum Cosmology e-prints*

Gonzalez, J. A., Sperhake, U., Bruegmann, B., Hannam, M., & Husa, S.: 2006, *ArXiv General Relativity and Quantum Cosmology e-prints*

Gott, J. R. I., Jurić, M., Schlegel, D., Hoyle, F., Vogeley, M., Tegmark, M., Bahcall, N., & Brinkmann, J.: 2005, ApJ **624**, 463

Governato, F., Mayer, L., Wadsley, J., Gardner, J. P., Willman, B., Hayashi, E., Quinn, T., Stadel, J., & Lake, G.: 2004, ApJ **607**, 688

Granato, G. L., De Zotti, G., Silva, L., Bressan, A., & Danese, L.: 2004, ApJ **600**, 580

Gunn, J. E. & Gott, J. R. I.: 1972, ApJ **176**, 1

Gunn, J. E. & Peterson, B. A.: 1965, ApJ **142**, 1633

Guth, A. H.: 1981, Phys. Rev. D **23**, 347

Haehnelt, M. G.: 1994, MNRAS **269**, 199

Haehnelt, M. G. & Rees, M. J.: 1993, MNRAS **263**, 168

Haiman, Z.: 2004, ApJ **613**, 36

Haiman, Z. & Loeb, A.: 1998, ApJ **503**, 505

Häring, N. & Rix, H.-W.: 2004, ApJ **604**, L89

Harker, G., Cole, S., Helly, J., Frenk, C., & Jenkins, A.: 2006, MNRAS **367**, 1039

Hasinger, G., Miyaji, T., & Schmidt, M.: 2005, A&A **441**, 417

Heckman, T. M., Kauffmann, G., Brinchmann, J., Charlot, S., Tremonti, C., & White, S. D. M.: 2004, ApJ **613**, 109

Helly, J. C., Cole, S., Frenk, C. S., Baugh, C. M., Benson, A., Lacey, C., & Pearce, F. R.: 2003, MNRAS **338**, 913

Herrmann, F., Hinder, I., Shoemaker, D., Laguna, P., & Matzner, R. A.: 2007, *ArXiv General Relativity and Quantum Cosmology e-prints*

Hopkins, P. F., Hernquist, L., Cox, T. J., Di Matteo, T., Martini, P., Robertson, B., & Springel, V.: 2005a, ApJ **630**, 705

Hopkins, P. F., Hernquist, L., Cox, T. J., Di Matteo, T., Robertson, B., & Springel, V.: 2006, ApJS **163**, 1

Hopkins, P. F., Hernquist, L., Martini, P., Cox, T. J., Robertson, B., Di Matteo, T., & Springel, V.: 2005b, ApJ **625**, L71

Hopkins, P. F., Strauss, M. A., Hall, P. B., Richards, G. T., Cooper, A. S., Schneider, D. P., Vanden Berk, D. E., Jester, S., Brinkmann, J., & Szokoly, G. P.: 2004, AJ **128**, 1112

Hoyle, F.: 1949, *IAU and International Union of Theoretical and Applied Mechanics Symposium pp. 195*

Hubble, E.: 1929, *Proceedings of the National Academy of Science* **15**, 168

Huchtmeier, W. K. & Richter, O.-G.: 1988, A&A **203**, 237

Islam, R. R., Taylor, J. E., & Silk, J.: 2003, MNRAS **340**, 647

Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., & Yoshida, N.: 2001, MNRAS **321**, 372

Kaiser, N.: 1984, ApJ **284**, L9

Kang, X., Jing, Y. P., Mo, H. J., & Börner, G.: 2005, ApJ **631**, 21

Kang, X., Jing, Y. P., & Silk, J.: 2006, ApJ **648**, 820

Kaspi, S., Smith, P. S., Netzer, H., Maoz, D., Jannuzi, B. T., & Giveon, U.: 2000, ApJ **533**, 631

Kauffmann, G.: 1995, MNRAS **274**, 153

Kauffmann, G.: 1996, MNRAS **281**, 487

Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M.: 1999, MNRAS **307**, 529

Kauffmann, G. & Haehnelt, M.: 2000, MNRAS **311**, 576

Kauffmann, G. & Haehnelt, M. G.: 2002, MNRAS **332**, 529

Kauffmann, G., Heckman, T. M., Tremonti, C., Brinchmann, J., Charlot, S., White, S. D. M., Ridgway, S. E., Brinkmann, J., Fukugita, M., Hall, P. B., Ivezić, Ž., Richards, G. T., & Schneider, D. P.: 2003, MNRAS **346**, 1055

Kauffmann, G., White, S. D. M., & Guiderdoni, B.: 1993, MNRAS **264**, 201

Kawakatu, N. & Umemura, M.: 2002, MNRAS **329**, 572

Keeton, C. R., Kuhlen, M., & Haiman, Z.: 2005, ApJ **621**, 559

Kereš, D., Katz, N., Weinberg, D. H., & Davé, R.: 2005, MNRAS **363**, 2

King, A. R.: 2002, MNRAS **335**, L13

Knapp, G. R., Turner, E. L., & Cunniffe, P. E.: 1985, AJ **90**, 454

Koppitz, M., Pollney, D., Reisswig, C., Rezzolla, L., Thornburg, J., Diener, P., & Schnetter, E.: 2007, *ArXiv General Relativity and Quantum Cosmology e-prints*

Kormendy, J. & Gebhardt, K.: 2001, in J. C. Wheeler & H. Martel (eds.), *AIP Conf. Proc. 586: 20th Texas Symposium on relativistic astrophysics*, pp 363–+

Kormendy, J. & Richstone, D.: 1995, ARA&A **33**, 581

Koushiappas, S. M., Bullock, J. S., & Dekel, A.: 2004, MNRAS **354**, 292

Lacey, C. & Cole, S.: 1993, MNRAS **262**, 627

Lacey, C. & Cole, S.: 1994, MNRAS **271**, 676

Lemson, G. & Kauffmann, G.: 1999, MNRAS **302**, 111

Li, Y., Hernquist, L., Robertson, B., Cox, T. J., Hopkins, P. F., Springel, V., Gao, L., Di Matteo, T., Zentner, A. R., Jenkins, A., & Yoshida, N.: 2006, *ArXiv Astrophysics e-prints*

Libeskind, N. I., Cole, S., Frenk, C. S., & Helly, J. C.: 2006, MNRAS **368**, 1381

Lilly, S. J., Le Fevre, O., Hammer, F., & Crampton, D.: 1996, ApJ **460**, L1+

Lodato, G. & Natarajan, P.: 2006, MNRAS **371**, 1813

Loveday, J., Peterson, B. A., Efstathiou, G., & Maddox, S. J.: 1992, ApJ **390**, 338

Lynden-Bell, D.: 1967, MNRAS **136**, 101

Lynden-Bell, D.: 1969, Nature **223**, 690

Madau, P., Ferguson, H. C., Dickinson, M. E., Giavalisco, M., Steidel, C. C., & Fruchter, A.: 1996, MNRAS **283**, 1388

Magorrian, J., Tremaine, S., Richstone, D., Bender, R., Bower, G., Dressler, A., Faber, S. M., Gebhardt, K., Green, R., Grillmair, C., Kormendy, J., & Lauer, T.: 1998, AJ **115**, 2285

Mahmood, A., Devriendt, J. E. G., & Silk, J.: 2005, MNRAS **359**, 1363

Malbon, R. K., Baugh, C. M., Frenk, C. S., & Lacey, C. G.: 2006, *ArXiv Astrophysics e-prints*

Maraston, C.: 1998, MNRAS **300**, 872

Maraston, C.: 2005, MNRAS **362**, 799

Marconi, A. & Hunt, L. K.: 2003, ApJ **589**, L21

Marconi, A., Risaliti, G., Gilli, R., Hunt, L. K., Maiolino, R., & Salvati, M.: 2004, MNRAS **351**, 169

Martini, P. & Weinberg, D. H.: 2001, ApJ **547**, 12

Mather, J. C., Cheng, E. S., Eplee, Jr., R. E., Isaacman, R. B., Meyer, S. S., Shafer, R. A., Weiss, R., Wright, E. L., Bennett, C. L., Boggess, N. W., Dwek, E., Gulkis, S., Hauser, M. G., Janssen, M., Kelsall, T., Lubin, P. M., Moseley, Jr., S. H., Murdock, T. L., Silverberg, R. F., Smoot, G. F., & Wilkinson, D. T.: 1990, ApJ **354**, L37

McLure, R. J. & Dunlop, J. S.: 2002, MNRAS **331**, 795

McLure, R. J. & Dunlop, J. S.: 2004, MNRAS **352**, 1390

McLure, R. J., Jarvis, M. J., Targett, T. A., Dunlop, J. S., & Best, P. N.: 2006, MNRAS **368**, 1395

Menci, N., Fiore, F., Perola, G. C., & Cavaliere, A.: 2004, ApJ **606**, 58

Menci, N., Fontana, A., Giallongo, E., Grazian, A., & Salimbeni, S.: 2006, ApJ **647**, 753

Menou, K. & Haiman, Z.: 2004, ApJ **615**, 130

Merloni, A.: 2004, MNRAS **353**, 1035

Merloni, A., Heinz, S., & di Matteo, T.: 2003, MNRAS **345**, 1057

Merritt, D. & Ekers, R. D.: 2002, *Science* **297**, 1310

Merritt, D. & Milosavljević, M.: 2005, *Living Reviews in Relativity* **8**, 8

Mihos, J. C. & Hernquist, L.: 1994a, ApJ **425**, L13

Mihos, J. C. & Hernquist, L.: 1994b, ApJ **431**, L9

Milgrom, M.: 1983, ApJ **270**, 365

Milosavljević, M., Merritt, D., Rest, A., & van den Bosch, F. C.: 2002, MNRAS **331**, L51

Milosavljević, M. & Phinney, E. S.: 2005, ApJ **622**, L93

Mo, H. J., Mao, S., & White, S. D. M.: 1998, MNRAS **295**, 319

Mo, H. J. & White, S. D. M.: 1996, MNRAS **282**, 347

Mobasher, B., Sharples, R. M., & Ellis, R. S.: 1993, MNRAS **263**, 560

Monaco, P. & Fontanot, F.: 2005, MNRAS **359**, 283

Monaco, P., Fontanot, F., & Taffoni, G.: 2007, MNRAS **375**, 1189

Monaghan, J. J.: 1992, ARA&A **30**, 543

Mouri, H. & Taniguchi, Y.: 2006, A&A **459**, 371

Nagashima, M., Lacey, C. G., Baugh, C. M., Frenk, C. S., & Cole, S.: 2005a, MNRAS **358**, 1247

Nagashima, M., Lacey, C. G., Okamoto, T., Baugh, C. M., Frenk, C. S., & Cole, S.: 2005b, MNRAS **363**, L31

Nagashima, M., Yahagi, H., Enoki, M., Yoshii, Y., & Gouda, N.: 2005c, ApJ **634**, 26

Navarro, J. F., Frenk, C. S., & White, S. D. M.: 1997, ApJ **490**, 493

Neistein, E., van den Bosch, F. C., & Dekel, A.: 2006, MNRAS **372**, 933

Netzer, H. & Peterson, B. M.: 1997, in D. Maoz, A. Sternberg, & E. M. Leibowitz (eds.), *ASSL Vol. 218: Astronomical Time Series*, pp 85–+

Neyman, J. & Scott, E. L.: 1952, ApJ **116**, 144

Norberg, P., Cole, S., Baugh, C. M., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Cross, N. J. G., Dalton, G., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Glazebrook, K., Jackson, C., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Peacock, J. A., Peterson, B. A., Sutherland, W., & Taylor, K.: 2002, MNRAS **336**, 907

Norman, C. & Scoville, N.: 1988, ApJ **332**, 124

Okamoto, T., Eke, V. R., Frenk, C. S., & Jenkins, A.: 2005, MNRAS **363**, 1299

Okamoto, T., Jenkins, A., Eke, V. R., Quilis, V., & Frenk, C. S.: 2003, MNRAS **345**, 429

Peacock, J. A. & Smith, R. E.: 2000, MNRAS **318**, 1144

Pedersen, K., Rasmussen, J., Sommer-Larsen, J., Toft, S., Benson, A. J., & Bower, R. G.: 2006, *New Astronomy* **11**, 465

Peebles, P. J. E.: 1969, ApJ **155**, 393

Pei, Y. C.: 1992, ApJ **395**, 130

Peng, C. Y., Impey, C. D., Rix, H.-W., Kochanek, C. S., Keeton, C. R., Falco, E. E., Lehár, J., & McLeod, B. A.: 2006, ApJ **649**, 616

Penzias, A. A. & Wilson, R. W.: 1965, ApJ **142**, 419

Percival, W. & Miller, L.: 1999, MNRAS **309**, 823

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Nunes, N. J., Pain, R., Pennypacker, C. R., Quimby, R., Lidman, C., Ellis, R. S., Irwin, M., McMahon, R. G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B. J., Filippenko, A. V., Matheson, T., Fruchter, A. S., Panagia, N., Newberg, H. J. M., Couch, W. J., & The Supernova Cosmology Project: 1999, ApJ **517**, 565

Power, C., Navarro, J. F., Jenkins, A., Frenk, C. S., White, S. D. M., Springel, V., Stadel, J., & Quinn, T.: 2003, MNRAS **338**, 14

Press, W. H. & Schechter, P.: 1974, ApJ **187**, 425

Quilis, V.: 2004, MNRAS **352**, 1426

Ratcliffe, A., Shanks, T., Parker, Q. A., & Fong, R.: 1998, MNRAS **293**, 197

Reed, D., Gardner, J., Quinn, T., Stadel, J., Fardal, M., Lake, G., & Governato, F.: 2003, MNRAS **346**, 565

Reed, D. S., Bower, R., Frenk, C. S., Jenkins, A., & Theuns, T.: 2007, MNRAS **374**, 2

Rees, M. J. & Ostriker, J. P.: 1977, MNRAS **179**, 541

Richards, G. T., Hall, P. B., Vanden Berk, D. E., Strauss, M. A., Schneider, D. P., Weinstein, M. A., Reichard, T. A., York, D. G., Knapp, G. R., Fan, X., Ivezić, Ž., Brinkmann, J., Budavári, T., Csabai, I., & Nichol, R. C.: 2003, AJ **126**, 1131

Richards, G. T., Strauss, M. A., Pindor, B., Haiman, Z., Fan, X., Eisenstein, D., Schneider, D. P., Bahcall, N. A., Brinkmann, J., & Brunner, R.: 2004, AJ **127**, 1305

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., & Tonry, J.: 1998, AJ **116**, 1009

Robertson, B., Hernquist, L., Cox, T. J., Di Matteo, T., Hopkins, P. F., Martini, P., & Springel, V.: 2006, ApJ **641**, 90

Rubin, V. C., Thonnard, N., & Ford, Jr., W. K.: 1978, ApJ **225**, L107

Sage, L. J.: 1993, A&A **272**, 123

Salpeter, E. E.: 1955, ApJ **121**, 161

Salpeter, E. E.: 1964, ApJ **140**, 796

Sánchez, A. G., Baugh, C. M., Percival, W. J., Peacock, J. A., Padilla, N. D., Cole, S., Frenk, C. S., & Norberg, P.: 2006, MNRAS **366**, 189

Sanders, D. B. & Mirabel, I. F.: 1996, ARA&A **34**, 749

Saslaw, W. C., Valtonen, M. J., & Aarseth, S. J.: 1974, ApJ **190**, 253

Scannapieco, C., Tissera, P. B., White, S. D. M., & Springel, V.: 2006, MNRAS **371**, 1125

Schmidt, M.: 1963, Nature **197**, 1040

Schödel, R., Ott, T., Genzel, R., Eckart, A., Mouawad, N., & Alexander, T.: 2003, ApJ **596**, 1015

Seljak, U.: 2000, MNRAS **318**, 203

Shakura, N. I. & Sunyaev, R. A.: 1973, A&A **24**, 337

Shankar, F., Salucci, P., Granato, G. L., De Zotti, G., & Danese, L.: 2004, MNRAS **354**, 1020

Sheth, R. K., Mo, H. J., & Tormen, G.: 2001, MNRAS **323**, 1

Shields, G. A., Gebhardt, K., Salviander, S., Wills, B. J., Xie, B., Brotherton, M. S., Yuan, J., & Dietrich, M.: 2003, ApJ **583**, 124

Silk, J. & Rees, M. J.: 1998, A&A **331**, L1

Smoot, G. F., Bennett, C. L., Kogut, A., Wright, E. L., Aymon, J., Boggess, N. W., Cheng, E. S., de Amici, G., Gulkis, S., Hauser, M. G., Hinshaw, G., Jackson, P. D., Janssen, M., Kaita, E., Kelsall, T., Keegstra, P., Lineweaver, C., Loewenstein, K., Lubin, P., Mather, J., Meyer, S. S., Moseley, S. H., Murdock, T., Rokke, L., Silverberg, R. F., Tenorio, L., Weiss, R., & Wilkinson, D. T.: 1992, ApJ **396**, L1

Somerville, R. S., Lemson, G., Kolatt, T. S., & Dekel, A.: 2000, MNRAS **316**, 479

Somerville, R. S., Primack, J. R., & Faber, S. M.: 2001, MNRAS **320**, 504

Songaila, A.: 2004, AJ **127**, 2598

Spergel, D. N., Bean, R., Dore', O., Nolta, M. R., Bennett, C. L., Hinshaw, G., Jarosik, N., Komatsu, E., Page, L., Peiris, H. V., Verde, L., Barnes, C., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E., & Wright, E. L.: 2006, *ArXiv Astrophysics e-prints*

Spergel, D. N., Verde, L., Peiris, H. V., Komatsu, E., Nolta, M. R., Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Page, L., Tucker, G. S., Weiland, J. L., Wollack, E., & Wright, E. L.: 2003, ApJS **148**, 175

Springel, V., Di Matteo, T., & Hernquist, L.: 2005a, MNRAS **361**, 776

Springel, V. & Hernquist, L.: 2002, MNRAS **333**, 649

Springel, V. & Hernquist, L.: 2003, MNRAS **339**, 289

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., & Pearce, F.: 2005b, Nature **435**, 629

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., & Pearce, F.: 2005c, Nature **435**, 629

Steffen, A. T., Barger, A. J., Cowie, L. L., Mushotzky, R. F., & Yang, Y.: 2003, ApJ **596**, L23

Storrie-Lombardi, L. J. & Wolfe, A. M.: 2000, ApJ **543**, 552

Sutherland, R. S. & Dopita, M. A.: 1993, ApJS **88**, 253

Swinbank, A. M., Smail, I., Chapman, S. C., Blain, A. W., Ivison, R. J., & Keel, W. C.: 2004, ApJ **617**, 64

Teyssier, R.: 2002, A&A **385**, 337

Tinsley, B. M.: 1972, A&A **20**, 383

Tinsley, B. M.: 1980, *Fundamentals of Cosmic Physics* **5**, 287

Toomre, A. & Toomre, J.: 1972, ApJ **178**, 623

Tremaine, S., Gebhardt, K., Bender, R., Bower, G., Dressler, A., Faber, S. M., Filippenko, A. V., Green, R., Grillmair, C., Ho, L. C., Kormendy, J., Lauer, T. R., Magorrian, J., Pinkney, J., & Richstone, D.: 2002, ApJ **574**, 740

Tully, R. B. & Fisher, J. R.: 1977, A&A **54**, 661

Turnshek, D. A.: 1997, in P. Petitjean & S. Charlot (eds.), *Structure and Evolution of the Intergalactic Medium from QSO Absorption Line System*, pp 263–+

Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T.: 2003, ApJ **598**, 886

Vestergaard, M.: 2002, ApJ **571**, 733

Volonteri, M.: 2006, *ArXiv Astrophysics e-prints*

Volonteri, M., Haardt, F., & Madau, P.: 2003, ApJ **582**, 559

Volonteri, M. & Rees, M. J.: 2005a, ApJ **633**, 624

Volonteri, M. & Rees, M. J.: 2005b, ApJ **633**, 624

Volonteri, M. & Rees, M. J.: 2006, ApJ **650**, 669

Walter, F., Bertoldi, F., Carilli, C., Cox, P., Lo, K. Y., Neri, R., Fan, X., Omont, A., Strauss, M. A., & Menten, K. M.: 2003, Nature **424**, 406

Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B.: 2006, ApJ **652**, 71

Weil, M. L., Eke, V. R., & Efstathiou, G.: 1998, MNRAS **300**, 773

White, S. D. M. & Frenk, C. S.: 1991, ApJ **379**, 52

White, S. D. M. & Rees, M. J.: 1978, MNRAS **183**, 341

Willott, C. J., McLure, R. J., & Jarvis, M. J.: 2003, ApJ **587**, L15

Willott, C. J., Percival, W. J., McLure, R. J., Crampton, D., Hutchings, J. B., Jarvis, M. J., Sawicki, M., & Simard, L.: 2005, ApJ **626**, 657

Wilman, R. J., Gerssen, J., Bower, R. G., Morris, S. L., Bacon, R., de Zeeuw, P. T., & Davies, R. L.: 2005, Nature **436**, 227

Wolf, C., Wisotzki, L., Borch, A., Dye, S., Kleinheinrich, M., & Meisenheimer, K.: 2003, A&A **408**, 499

Woo, J.-H., Treu, T., Malkan, M. A., & Blandford, R. D.: 2006, ApJ **645**, 900

Wyithe, J. S. B.: 2006, MNRAS **365**, 1082

Wyithe, J. S. B. & Loeb, A.: 2003, ApJ **595**, 614

Yang, X., Mo, H. J., Jing, Y. P., & van den Bosch, F. C.: 2005, MNRAS **358**, 217

Yoo, J. & Miralda-Escudé, J.: 2004, ApJ **614**, L25

Yoshida, N., Stoehr, F., Springel, V., & White, S. D. M.: 2002, MNRAS **335**, 762

Yu, Q. & Tremaine, S.: 2002, MNRAS **335**, 965

Zaritsky, D., Kennicutt, Jr., R. C., & Huchra, J. P.: 1994, ApJ **420**, 87

Zheng, Z., Berlind, A. A., Weinberg, D. H., Benson, A. J., Baugh, C. M., Cole, S., Davé, R., Frenk, C. S., Katz, N., & Lacey, C. G.: 2005, ApJ **633**, 791

Zucca, E., Zamorani, G., Vettolani, G., Cappi, A., Merighi, R., Mignoli, M., Stirpe, G. M., MacGillivray, H., Collins, C., Balkowski, C., Cayatte, V., Maurogordato, S., Proust, D., Chincarini, G., Guzzo, L., Maccagni, D., Scaramella, R., Blanchard, A., & Ramella, M.: 1997, A&A **326**, 477

Zwaan, M. A., Briggs, F. H., Sprayberry, D., & Sorar, E.: 1997, ApJ **490**, 173