The effects of photoionization on galaxy formation – I. Model and results at $z = 0$

A. J. Benson,[1]⋆ C. G. Lacey,[2] C. M. Baugh,[3] S. Cole[3] and C. S. Frenk[3]

[1]*California Institute of Technology, MC 105-24, Pasadena, CA 91125, USA*
[2]*SISSA, Astrophysics Sector, via Beirut 2-4, 34014 Trieste, Italy*
[3]*Physics Department, University of Durham, Durham DH1 3LE*

ABSTRACT

We develop a coupled model for the evolution of the global properties of the intergalactic medium (IGM) and the formation of galaxies, in the presence of a photoionizing background due to stars and quasars. We use this model to predict the thermodynamic history of the IGM when photoionized by galaxies forming in a cold dark matter (CDM) universe. The evolution of the galaxies is calculated using a semi-analytical model, including a detailed treatment of the effects of tidal stripping and dynamical friction on satellite galaxies orbiting inside larger dark matter haloes. We include in the model the negative feedback on galaxy formation from the photoionizing background. Photoionization inhibits galaxy formation in low-mass dark matter haloes in two ways: (i) heating of the IGM and inhibition of the collapse of gas into dark haloes by the IGM pressure, and (ii) reduction in the rate of radiative cooling of gas within haloes. The result of our method is a self-consistent model of galaxy formation and the IGM. The IGM is reheated twice (during reionization of H I and He II), and we find that the star formation rate per unit volume is slightly suppressed after each episode of reheating. We find that galaxies brighter than $L_\star$ are mostly unaffected by reionization, while the abundance of faint galaxies is significantly reduced, leading to present-day galaxy luminosity functions with shallow faint-end slopes, in good agreement with recent observational data. Reionization also affects other properties of these faint galaxies, in a readily understandable way.

Key words: galaxies: formation – intergalactic medium – cosmology: theory.

1 INTRODUCTION

It is now known that the hydrogen in the intergalactic medium (IGM), which became neutral at $z \sim 1000$ (Peebles 1968; Zeldovich, Kurt & Sunyaev 1968), must have been reionized somewhere between redshifts 6 and 30, the lower limit coming from the lack of a Gunn–Peterson trough in quasar spectra at that redshift (e.g. Fan et al. 2000), and the upper limit from the bound on the optical depth to the last scattering surface measured from the cosmic microwave background (CMB) (Netterfield et al. 2001). In fact, very recent results (Becker et al. 2001; Djorgovski et al. 2001) suggest that reionization may have occurred very close to the lower limit of this range. If there are large populations of galaxies or quasars at high redshifts, as is predicted by current structure formation models (e.g. Benson et al. 2001a) and as confirmed up to redshifts ≈ 6 observationally (Fan et al. 2000; Stern et al. 2000), then reionization is most likely to have occurred through photoionization, as both galaxies and quasars emit copious

quantities of ionizing photons (e.g. Couchman & Rees 1986). Several models of reionization have been developed in recent years (Gnedin & Ostriker 1997; Haiman & Loeb 1997; Valageas & Silk 1999; Chiu & Ostriker 2000; Ciardi et al. 2000; Gnedin 2000a; Miralda-Escudé, Haenhelt & Rees 2000; Benson et al. 2001a), many reaching the conclusion that reionization occurred at $z \approx 7$–12, although large systematic uncertainties remain due to uncertainties in the efficiency of galaxy formation, in the fraction of ionizing photons that escape a galaxy, and in the density distribution of ionized gas in the IGM (see, e.g., Benson et al. 2001a). If this picture of reionization is correct, then it is clear that the thermodynamic history of the IGM is determined by the formation and evolution of galaxies and quasars.

The photoionizing background responsible for reionizing the IGM may also act, directly and indirectly, to inhibit galaxy formation, as was first pointed out by Doroshkevich, Zeldovich & Novikov (1967), and first investigated in the context of CDM models by Couchman & Rees (1986). Galaxies are thought to form by a two-stage collapse process, in which gas first collapses into dark matter haloes along with the dark matter itself, and then

⋆E-mail: abenson@astro.caltech.edu

collapses relative to the dark matter within haloes if it is able to cool radiatively to below the halo virial temperature, thus losing its pressure support. The second stage of the collapse is necessary in order to increase the gas density to the point where it becomes self-gravitating relative to the dark matter, which is believed to be a necessary condition for the gas to be able to fragment to form stars. In the presence of an ionizing background, both stages of this collapse process are inhibited, particularly for low-mass haloes. First, the ionizing background heats the IGM to temperatures of around 10^4 K, and the resulting thermal pressure of the gas then prevents it from collapsing into low-mass haloes along with the dark matter. Secondly, the ionizing background reduces the rate of radiative cooling of gas inside haloes, mainly by reducing the abundance of neutral atoms which can be collisionally excited. Both of these mechanisms will strongly inhibit galaxy formation in haloes with virial temperatures less than $\sim 10^4$ K, and so may have important effects on the faint end of the galaxy luminosity function and also on the properties of the dwarf satellite galaxies of the Milky Way and other galaxies.

There have been many studies of the effects of an ionizing background on galaxy formation, both analytical (e.g. Babul & Rees 1992; Efstathiou 1992; Chiba & Nath 1994; Thoul & Weinberg 1996; Kepner, Babul & Spergel 1997; Nagashima, Gouda & Sugiura 1999) and using numerical simulations (e.g. Vedel, Hellsten & Sommer-Larsen 1994; Quinn, Katz & Efstathiou 1996; Navarro & Steinmetz 1997; Weinberg, Hernquist & Katz 1997), but in most of these the ionizing background was simply taken as an external input. A few studies have investigated the more difficult self-consistent problem, relating the ionizing background at any redshift to the fraction of baryons which had previously collapsed to form galaxies, and at the same time including the effect of the ionizing background in inhibiting further galaxy formation (e.g. Shapiro, Giroux & Babul 1994; Gnedin & Ostriker 1997; Valageas & Silk 1999). The analytical studies have used a wide variety of approaches and approximations, but have generally modelled galaxy formation and the effects of photoionization only in a very simplified or partial way (e.g., for photoionization, either considering only the suppression of collapse into dark haloes, or the suppression of cooling within dark haloes). On the other hand, the numerical studies were limited in the predictions they could make about properties of the present-day galaxy population by the range of physics included and by their limited dynamical range. In the present paper, we present a new model for the coupled evolution of the IGM, the ionizing background and galaxies, based on a semi-analytical model of galaxy formation, enabling us to determine in much more detail than in previous studies the effects of photoionization on observable galaxy properties. Compared to previous analytical studies (in particular, Valageas & Silk 1999), the main improvements are that we have a much more detailed model for galaxy formation through hierarchical clustering, including many different processes, and a more accurate model for how photoionization suppresses galaxy formation through the two mechanisms described above. In particular, the suppression of gas collapse into dark matter haloes due to the IGM pressure is modelled on the basis of the latest results from gas-dynamical simulations.

Our starting point is the semi-analytic model of galaxy formation described by Cole et al. (2000), which attempts to model the galaxy formation process *ab initio*, in the framework of structure formation through hierarchical clustering. We then modify this to include the new physics we are interested in here. We develop a model for the evolution of the thermodynamic

properties of the IGM in the presence of the ionizing radiation background produced by galaxies and quasars, the former predicted by the semi-analytic model, and the latter based on observational data. We are then able to predict the mean temperature of the IGM and the spectrum of the ionizing background as functions of cosmic time. We adapt the Cole et al. model to determine the mass of gas able to accrete on to each dark matter halo from the heated IGM, and to include the effects of heating by the ionizing background. Finally, we include a more detailed treatment of the dynamical evolution of satellites orbiting within larger dark matter haloes, including the effects of tidal stripping. The approach of investigating the effects of photoionization on galaxy formation by using a semi-analytic model was previously taken by Nagashima et al. (1999), but they considered only the heating of gas in haloes by the UV background, and so our current work represents a more thorough treatment of the problem, as well as being based on a much-improved galaxy formation model.

There are two parts to this paper. First, we describe how the physics of the IGM/galaxy interaction may be modelled in a simple way. Secondly, we present results from our model, focusing on the evolution of the IGM and ionizing background, and on the properties of the local galaxy population. We briefly comment on how high-redshift galaxies are affected. In a companion paper (Benson et al. 2001b) we will explore in detail the consequences of our model for the population of satellite galaxies seen in the Local Group.

The remainder of this paper is organized as follows. In Sections 2 and 3 we describe in detail our model for the evolution of the IGM and galaxy formation. In Section 4 we present results from this model for the evolution of the IGM and the population of galaxies at the present day in the currently favoured ΛCDM model of structure formation. Finally, in Section 5 we present our conclusions.

2 MODEL OF PHOTOIONIZATION AND IGM EVOLUTION

We use the semi-analytic model of galaxy formation developed by Cole et al. (2000) to determine the properties of galaxies in the Universe. The model includes formation and merging of dark matter haloes, shock-heating and radiative cooling of gas within haloes, collapse of cold gas to form galaxy discs, star formation from the cold gas, galaxy mergers within common dark matter haloes leading to formation of galaxy spheroids, chemical enrichment, and the luminosity evolution of stellar populations. The fiducial model presented by Cole et al. (for which $\Omega_0 = 0.3$, $\Lambda_0 = 0.7$, $\Omega_b = 0.02$ and $h = 0.7$)[1] has been shown to reproduce many of the properties of galaxies in the local Universe, such as their luminosity functions, the slope and scatter of the Tully–Fisher relation, colours, sizes and metallicities (Cole et al. 2000) and also the clustering of galaxies in real and redshift space (Benson et al. 2000a,b).

The Cole et al. (2000) model, like most other semi-analytic models of galaxy formation (e.g. Kauffmann, White & Guiderdoni 1993; Somerville & Primack 1999), includes a prescription for feedback due to energy input from supernovae and stellar winds. This is assumed to reheat cold gas and eject it from galaxies, thus inhibiting galaxy formation in low-mass dark matter haloes. Several studies of how this feedback may physically operate can be found in the literature (Dekel & Silk 1986; Mac Low & Ferrara

---

[1] We define Hubble's constant to be $H_0 = 100 \, h \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$.

1999; Goodwin, Pearce & Thomas 2001). This feedback is required in CDM models in order to produce a faint-end slope of the local galaxy luminosity function as shallow as that observed (White & Rees 1978; Cole 1991; White & Frenk 1991), and also to produce galactic discs of sizes comparable to those observed (Cole et al. 2000).

While ejection of gas by supernovae-driven outflows is undoubtedly an important process (e.g. Martin 1999), other processes may also inhibit galaxy formation, for example, preheating of the IGM (Evrard & Henry 1991; Kaiser 1991; Blanchard, Valls-Gabaud & Mamon 1992; Valageas & Silk 1999), heating of the gas inside galaxy and cluster haloes (Wu, Fabian & Nulsen 2000; Bower et al. 2001), and the effects of a photoionizing background. The last of these is perhaps the best studied (see, e.g., Efstathiou 1992, Katz, Weinberg & Hernquist 1996, Thoul & Weinberg 1996 and Bullock, Kravtsov & Weinberg 2000). A photoionizing background both supplies heat to the gas through ionization, and reduces the rate at which the gas can cool by reducing the abundance of neutral atomic species which can be collisionally excited. It thus raises the IGM temperature and so prevents it from collapsing into small haloes, and also reduces the cooling rate of gas within haloes and so reduces the fraction of baryons which can collapse to form a galaxy. For the formation of the very first objects at high redshift, cooling of the gas by molecular hydrogen is probably important, and one needs to consider the dissociation of these molecules by non-ionizing UV radiation (e.g. Ciardi et al. 2000), but these processes are important only well before the epoch of reionization in our model, since conversion of only a very tiny fraction of the baryons into stars is sufficient to produce enough UV radiation to dissociate all of the H_2 molecules.

In this section we describe how we modify the model of Cole et al. (2000) to calculate the evolution of the IGM temperature and ionizing background, the suppression of gas collapse into haloes by the IGM pressure, and the suppression of cooling within haloes by the ionizing background. Our modelling of the dynamical evolution of satellite galaxies within larger haloes is described in Section 3.

2.1 Evolution of the ionizing background and the IGM temperature

We will treat the IGM as a mixture of six species (e, H I, H II, He I, He II and He III) which interact with each other and with a uniform background of radiation emitted by stars and quasars. As we are here primarily interested in the properties of low-redshift galaxies, we will not include H_2 in our calculations, since it will be dissociated at high redshifts (e.g. Ciardi et al. 2000). We follow the evolution of the abundances of these species and the gas temperature for parcels of gas spanning a wide range in density contrast. The density contrast of each parcel is allowed to change with time as described in Section 2.1.1. Here, we treat all gas in the Universe as being part of the IGM. Since the fraction of the total gas content of the Universe which becomes part of a galaxy in our model is always small, this is a reasonable approximation. Some gas should, of course, fall into the potential wells of dark matter haloes (see Section 2.2). Since this gas typically occupies a small fraction of the volume of the Universe, we ignore it for calculating the properties of the IGM.

In the remainder of this section we describe in detail how we model the evolution of the ionizing background and IGM temperature.

2.1.1 Evolution of gas density

We wish to calculate the thermodynamic behaviour of gas in the IGM up until the point at which it falls into a virialized dark matter halo. The gas in the IGM will have a range of overdensities resulting from the growth of density fluctuations due to gravitational instability (we do not consider here the possibility of a multiphase medium, which may also produce variations in gas density). Since recombination rates, and consequently heating and cooling rates, depend on the gas density, it is necessary to take this evolving distribution of densities into account in our model.

We characterize the evolving distribution of gas densities via the probability distribution function (PDF), $P_V(\Delta, t)$, defined such that $P_V(\Delta, t)\, d\Delta$ is the fraction of volume in the Universe occupied by gas with a density contrast $\Delta = \rho/\bar{\rho}$ at time t, where ρ is the gas density at a point, and $\bar{\rho}$ is the mean gas density in the Universe. Normalization of this function to give the correct mean density and total mass requires that

$$\int_0^\infty P_V(\Delta, t)\, d\Delta = 1, \tag{1}$$

and

$$\int_0^\infty \Delta P_V(\Delta, t)\, d\Delta = 1. \tag{2}$$

The fraction of mass with density contrast $\leq \Delta$ is given by

$$F(\Delta, t) = \int_0^\Delta \Delta' P_V(\Delta', t)\, d\Delta'. \tag{3}$$

We assume that as the gas density field evolves, the ranking of gas elements by density remains the same. The density contrast at time t of a gas element which has density contrast Δ_0 at time t_0 is therefore given by the solution of

$$F(\Delta[t], t) = F(\Delta_0, t_0), \tag{4}$$

We can use equation (4) to calculate the evolution in overdensity $\Delta[t]$ of individual parcels of IGM gas having different values of Δ_0, once the functional form and evolution of $P_V(\Delta, t)$ have been specified. In our standard model, we assume that the PDF has a lognormal form, which has been found to provide a reasonable description of the density distribution produced by gravitational instability in the mildly non-linear regime (e.g. Coles & Jones 1991),

$$P_V(\Delta) = \left(\frac{A}{\Delta}\right) \exp\left[\frac{-(\ln\Delta - \overline{\ln\Delta})^2}{2\sigma_\Delta^2}\right], \tag{5}$$

Here, σ_Δ determines the width of the distribution, and the constants A and $\overline{\ln\Delta}$ are fixed from the normalization conditions (equations 1 and 2). The value of σ_Δ as a function of time can be chosen to reproduce a desired baryonic clumping factor

$$f_{\text{clump}} \equiv \frac{\overline{\rho^2}}{\bar{\rho}^2} = \int_0^\infty \Delta^2 P_V(\Delta, t)\, d\Delta, \tag{6}$$

where the overbar denotes a volume average. In particular, we will choose σ_Δ to reproduce the baryonic clumping factor, $f_{\text{clump}}^{\text{(variance)}}$, derived by Benson et al. (2001a). In their calculation, Benson et al. (2001a) assumed that gas in the IGM essentially traces the dark matter, except that pressure prevents the gas from falling into dark matter haloes with virial temperatures less than 10^4 K. They then calculated $f_{\text{clump}}^{\text{(variance)}} = 1 + \sigma^2$, where σ^2 is the variance of the dark

Table 1. The clumping factor, $f_{\text{clump}}^{(\text{variance})}$ obtained by Benson et al. (2001a) at several redshifts. Also shown are the corresponding parameters of the PDF defined in equation (5).

z	f_{clump}	A	$\overline{\ln\Delta}$	σ_Δ
0	286	0.168	−2.83	2.38
1	125	0.182	−2.42	2.20
2	94	0.187	−2.27	2.13
3	75	0.192	−2.16	2.08
5	47	0.203	−1.93	1.96
10	13	0.250	−1.28	1.60
20	2.1	0.461	−0.37	0.87
30	1.4	0.665	−0.18	0.60

matter density field in spheres of radius equal to the radius of a 10^4-K halo (σ^2 was calculated from the smoothed non-linear dark matter power spectrum obtained using the techniques of Peacock & Dodds 1996). In the present work, the halo mass below which gas accretion is negligible varies as a function of time. Nevertheless, our estimate of the clumping factor should still provide a reasonable approximation. We note that at the redshift appropriate for H I reionization in this work (see Section 4.1) the two different clumping factors considered by Benson et al. (2001a) are in fact very similar (see their fig. 9). Specific values of $f_{\text{clump}}^{(\text{variance})}$ and the parameters of the corresponding PDF are given in Table 1. (Note that here we define $f_{\text{clump}}^{(\text{variance})}$ relative to the mean gas density of the Universe, unlike Benson et al. 2001a who defined it relative to the mean density of gas remaining in the IGM.)

Once the evolution of the clumping factor has been chosen, our model results are insensitive to the particular functional form chosen for the PDF. For example, if instead of the lognormal distribution we use the form

$$P_V(\Delta) = \left(\frac{A}{\Delta}\right) \exp\left[\frac{-(|\ln\Delta - \overline{\ln\Delta}|)^3}{2\sigma_\Delta^3}\right], \qquad (7)$$

which falls off much more rapidly away from $\ln\Delta = \overline{\ln\Delta}$, this makes a negligible difference to the evolution of the mean IGM temperature, ionization state and the spectrum of the ionizing background. We truncate the distribution of gas densities above $\Delta = 300$, which is roughly the mean density contrast of haloes at $z = 0$ in our adopted cosmology, because reaction rates become extremely rapid for higher densities, making solution of the rate equations numerically difficult. Gas at higher overdensities accounts for only a small fraction of the total volume, and we have checked that moving the truncation point to larger Δ makes little difference to our results.

2.1.2 Background radiation

We follow the proper number density of photons per unit frequency, n_ν, which evolves with time as

$$\frac{\partial n_\nu}{\partial t} = \frac{\dot{a}}{a}\left[-3n_\nu + \frac{\partial}{\partial\nu}(\nu n_\nu)\right] + S_\nu - \sum_i\sum_j c\sigma_{\nu,i}f_{v,j}n_{i,j}n_\nu, \qquad (8)$$

where c is the speed of light, the term $-3(\dot{a}/a)n_\nu$ on the right-hand side represents the dilution of the number density by the Hubble expansion, and the term $(\dot{a}/a)\partial(\nu n_\nu)/\partial\nu$ describes the effect of the

redshifting of the photon frequencies. Here S_ν is the emissivity (i.e., number of photons emitted per unit volume, per unit time, per unit frequency), $\sigma_{\nu,i}$ is the photoionization cross-section for species i (H I, He I, He II), $f_{v,j}$ is the fraction of the volume of the Universe occupied by gas in density bin j, and $n_{i,j}$ is the abundance of species i in density bin j.

The photon number density is related to the background intensity by

$$J_\nu = \frac{ch_P\nu}{4\pi}n_\nu, \qquad (9)$$

where J_ν is the intensity per unit solid angle per unit frequency, and h_P is Planck's constant.

2.1.3 Rate equations

The evolution of the abundances of the different ionization states of H and He is described by equations of the form

$$\frac{dn_i}{dt} = [\alpha_i(T_{\text{IGM}})n_{i+1}n_e - \alpha_{i-1}(T_{\text{IGM}})n_in_e - \Gamma_{e,i}(T_{\text{IGM}})n_in_e$$

$$+ \Gamma_{e,i-1}(T_{\text{IGM}})n_{i-1}n_e - \Gamma_{\gamma,i}n_i + \Gamma_{\gamma,i-1}n_{i-1}]$$

$$+ \left[\frac{1}{\Delta(a)}\frac{d\Delta(a)}{dt} - 3\frac{\dot{a}}{a}\right]n_i, \qquad (10)$$

where for each atomic species H or He, i refers to the ionization state (i.e., $i = 1$ and 2 for H I and H II, and $i = 3$, 4 and 5 for He I, He II and He III), n_i is the proper number density, T_{IGM} is the temperature, α_i is the recombination rate coefficient to i, $\Gamma_{e,i}$ is the collisional ionization rate coefficient from i, and $\Gamma_{\gamma,i}$ is the photoionization rate coefficient from i. The evolution of the electron density then follows from the conservation of the total number of electrons.

We consider the evolution of a parcel of gas of density contrast $\Delta(t)$, which has a thermal energy per unit volume given by $E = \frac{3}{2}k_BT_{\text{IGM}}n_{\text{tot}}$, where n_{tot} is the total number of particles per unit volume. The energy changes due to adiabatic expansion/compression and atomic heating/cooling processes. Thus the evolution of E may be written as

$$\frac{dE}{dt} = \frac{5}{3}\left[\frac{1}{\Delta(a)}\frac{d\Delta(a)}{dt} - 3\frac{\dot{a}}{a}\right]E + (\Sigma^T - \Lambda^T), \qquad (11)$$

where the first term represents adiabatic expansion or compression, and the second represents atomic heating and cooling processes. Σ^T is the rate of heating per unit volume due to all heat sources (i.e., photoionization and Compton heating), and Λ^T is the rate of cooling per unit volume due to all heat sinks (i.e., Compton cooling and various atomic processes). We use the notation Σ^T and Λ^T to indicate rates of thermal energy gain/loss, as distinct from the usual radiative cooling function Λ which includes the entire energy of the photons emitted by recombinations. Note that the evolution of the gas density is entirely determined by the functional form of the PDF $P_V(\Delta)$ and the redshift evolution of f_{clump}, and is unaffected by any heating/cooling of the gas. In reality, the gas density distribution should respond to differences in gas pressure. However, the effects of pressure forces should be important only on scales smaller than the Jeans length (or, more precisely, the filtering length to be introduced in the next subsection), which always remains small ($\lesssim 1\,h^{-1}\,\text{Mpc}$) relative to the much larger scales over which we calculate volume averages.

From equation (11) and the definition of E, we obtain the

following equation for the evolution of the IGM temperature:

$$\frac{1}{T_{IGM}}\frac{dT_{IGM}}{dt} = -2\frac{\dot{a}}{a} + \frac{2}{3\Delta}\frac{d\Delta}{dt} + \frac{\Sigma^T - \Lambda^T}{\frac{3}{2}k_B T_{IGM} n_{tot}} - \frac{\Delta}{a^3 n_{tot}}\frac{d}{dt}\left(\frac{a^3 n_{tot}}{\Delta}\right).$$

(12)

The final term accounts for the effects of changes in the total particle number density due to ionization/recombination. For a homogeneous IGM ($\Delta = 1$) with no heating or cooling and no ionization or recombination, we have simply $n_{tot} \propto a^{-3}$ and $T_{IGM} \propto a^{-2}$.

Equations (8)–(12) describe the evolution of a parcel of gas of specified final density contrast Δ_0. These equations, along with those describing the evolution of the background radiation spectrum, are solved for a range of Δ_0 using a modified Bulirsch–Stoer method, which is applicable to this stiff set of equations (Bader & Deuflhard 1983). The matrix decomposition that must be carried out as part of this method is efficiently achieved using a suitable sparse matrix package.

The initial conditions for the abundance of each species and for the temperature are taken from the RECFAST code (Seager, Sasselov & Scott 2000), which accurately evolves the IGM through the recombination epoch (we typically begin our own calculations at $z = 200$, at which point recombination is essentially complete, but no significant sources of radiation have appeared in our model).

We take photoionization cross-sections from Verner et al. (1996), recombination rate coefficients from Verner & Ferland (1996) and Arnaud & Rothenflug (1985), and collisional ionization rates from Voronov (1997). The cooling rate due to collisional excitation of H I was taken from Scholz & Walters (1991), while that for He II was taken from Black (1981), with the modification introduced by Cen (1992) at high temperatures. The cooling rate due to free–free emission was computed using the Gaunt factors given by Sutherland (1998).

2.2 Critical mass for collapse

If the IGM has a non-zero temperature, then pressure forces will prevent gravitational collapse of the gas on small scales. In the absence of dark matter, the effects of pressure on the growth of density fluctuations in the gas due to their self-gravity are described by a simple Jeans criterion, such that density fluctuations on mass scales below the Jeans mass M_J are stable against collapse. However, this simple criterion needs to be modified in the case of non-linear collapse of the gas in the presence of a gravitationally dominant cold dark matter component which collapses to form dark matter haloes. Gnedin (2000b) has obtained an analytical description of the effects of gas pressure in this case, based on earlier work by Gnedin & Hui (1998). Using a linear perturbation analysis, Gnedin & Hui found that growth of density fluctuations in the gas is suppressed for comoving wavenumbers $k > k_F$, where the critical wavenumber k_F is related to the Jeans wavenumber k_J by

$$\frac{1}{k_F^2(t)} = \frac{1}{D(t)}\int_0^t dt'\, a^2(t')\frac{\ddot{D}(t') + 2H(t')\dot{D}(t')}{k_J^2(t')}\int_{t'}^t \frac{dt''}{a^2(t'')},$$

(13)

and k_J is defined as

$$k_J = a\left(4\pi G\bar{\rho}_{tot}\frac{3\mu m_H}{5k_B \bar{T}_{IGM}}\right)^{1/2}.$$

(14)

In the above, $\bar{\rho}_{tot}$ is the mean *total* mass density including dark

matter, $D(t)$ and $H(t)$ are the linear growth factor and Hubble constant respectively as functions of cosmic time t, and $\cdot$ represents a derivative with respect to t. This expression for k_F accounts for arbitrary thermal evolution of the IGM, through $k_J(t)$. Corresponding to the critical wavenumber k_F there is a critical mass M_F, which we will hereafter call the filtering mass, defined as

$$M_F = (4\pi/3)\bar{\rho}_{tot}(2\pi a/k_F)^3.$$

(15)

The Jeans mass M_J is defined analogously in terms of k_J. In the absence of pressure in the IGM, a halo of mass M_{tot} would be expected to accrete a mass $(\Omega_b/\Omega_0)M_{tot}$ in gas when it collapsed. Gnedin (2000b) found that in cosmological gas-dynamical simulations with a photoionized IGM, the average mass of gas M_{gas} which falls into haloes of mass M_{tot} can be fitted with the formula

$$M_{gas} = \frac{(\Omega_b/\Omega_0)M_{tot}}{[1 + (2^{1/3} - 1)M_F/M_{tot}]^3},$$

(16)

with the same value of M_F as given by equations (14) and (15). The denominator in the above expression thus gives the factor by which the accreted gas mass is reduced because of the IGM pressure. Specifically, M_F gives the halo mass for which the amount of gas accreted is reduced by a factor of 2 compared to the case of no IGM pressure.

In our model, we calculate the filtering mass $M_F(z)$ from equations (14) and (15), using for the IGM temperature the volume-averaged value $\bar{T}_{IGM} = \sum_j f_{v,j} T_{IGM,j}$ (where $T_{IGM,j}$ is the temperature of IGM gas in density bin j).

In the galaxy formation model of Cole et al. (2000) each dark matter halo would accrete, at the time of its formation, a mass of gas equal to $(\Omega_b/\Omega_0)M_{tot} - M_{gal}$, where M_{gal} is the mass of gas which had already formed into galaxies in progenitors of this halo. It is the formation history of a halo that fundamentally determines M_{gal}, and so two haloes of the same mass forming at the same time will have different M_{gal} due to their different formation paths. Gnedin (2000b) did not address the question of how the amount of gas accreted by the halo depended on its formation history. Therefore, to account for the effects of the filtering mass, we make the simplest possible modification, and assume that instead each halo accretes a mass of hot gas equal to

$$M_{hot} = \frac{(\Omega_b/\Omega_0)M_{tot} - M_{gal}}{[1 + (2^{1/3} - 1)M_F/M_{tot}]^3}.$$

(17)

In this way the amount of hot gas contained in the halo depends both upon its mass relative to the filtering mass and on its formation history. The dependence on formation history arises only through the M_{gal} term. In the absence of any cooling (in which case $M_{gal} = 0$), the amount of gas accreted by each halo would be independent of its formation history, and so the baryon fraction in a halo formed by the merging of several lower mass haloes would be higher than that of its progenitors (neglecting any time variation in the filtering mass).

The gas which would have been accreted in the absence of IGM pressure is assumed to remain in the IGM, presumably nearby the halo, resulting in regions with a baryon fraction higher than the Universal mean (although its exact location is not specified, nor required, by our model). This gas remains available for accretion later on in the merging process when another new halo is formed. The hot gas which does accrete is distributed within the halo as described by Cole et al. (2000). As well as measuring the mean mass of gas accreted by haloes of a given mass, Gnedin (2000b)

was able to approximately characterize the scatter in this relation. We typically do not include this scatter in our calculations, but find that its inclusion makes no significant difference to our results.

2.3 Cooling rate of gas in haloes

The cooling of gas within dark haloes, which controls how much of the gas can collapse to form galaxies, is also affected by the ionizing background. The gas within dark haloes is at much higher densities than in the IGM, so we assume that it is in ionization equilibrium under the combined effects of atomic collisions and the external photoionizing background. We also assume that the halo gas is optically thin to the ionizing background and to its own emission. While in our models the mean metallicity of the IGM remains low enough that it has negligible effect on the cooling, this is not true for all of the gas in haloes, some of which becomes significantly metal-enriched due to ejection of gas from galaxies by supernova feedback. We therefore use the publicly available photoionization code MAPPINGS III,[2] an updated version of the MAPPINGS II code used by Sutherland & Dopita (1993), to calculate the radiative cooling rate of gas in haloes in collisional and photoionization equilibrium, including the effects of metals. Using this code, we calculate the net cooling/heating rate of the gas as a function of density, temperature, metallicity and redshift, taking the photoionizing background predicted by our model at that redshift. We also include Compton cooling due to free electrons scattering off microwave background photons.

Fig. 1 shows the net cooling rate (i.e., the difference of heating and cooling rates) as a function of temperature, for gas in the presence of the ionizing background from our fiducial model (see Section 4.2), for a metallicity $Z = 0.3\,Z_\odot$, at three different redshifts, $z = 0$, 2 and 4. The cooling rates per unit volume are divided by n_H^2, and calculated at densities $n_H = 1.3 \times 10^{-3}$, 3.5×10^{-2} and 1.6×10^{-1} cm^{-3} at redshifts $z = 0$, 2 and 4 respectively, which correspond to the mean densities of gas in dark matter haloes at those redshifts. We also plot the cooling curve in the absence of an ionizing background (dotted line). For $z = 0$ and $z = 2$, gas cooler than $T \approx 2 \times 10^4$ K is actually heated rather than cooled in the presence of the ionizing background. We see that at $z = 4$, gas at the average density for virialized haloes is being heated only for $T < 10^4$ K, a consequence of the high gas density at this redshift. On the other hand, at the lower redshifts plotted, photoionization almost completely suppresses cooling at $T \lesssim 4 \times 10^4$ K. For gas at the halo virial temperature, the latter corresponds to a halo circular velocity of approximately 30 km s^{-1}. Barkana & Loeb (1999) described the photo-evaporation of gas in cool haloes after reionization. The consequences of this process for galaxy formation are accounted for in our model. Since the gas in these cool haloes is being heated (as described above), it will never cool to become part of a galaxy. (Note that we improve upon Barkana & Loeb's calculation by considering the metallicity of the evaporating gas, but at present we do not include self-shielding of the inner regions of the halo as they did.) We do not, however, account for any effects that this photo-evaporation may have on the surrounding IGM.

In our model of galaxy formation, the gas in a dark matter halo is assumed to be isothermal at the virial temperature T_{vir} of the halo, and to have a uniform metallicity Z_{halo}. The virial temperature is defined in terms of the circular velocity V_c at the virial radius of the
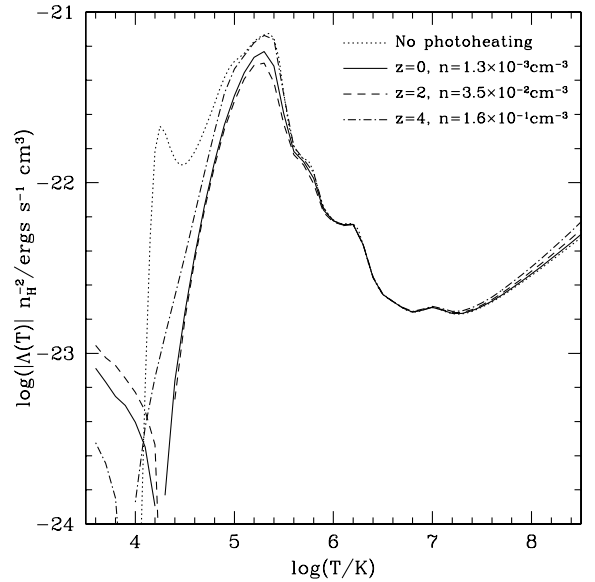
Figure 1. The net cooling/heating function for gas at different redshifts in the presence of the photoionizing background predicted in our fiducial model (Section 4.2). We plot the absolute value of the cooling–heating rate per unit volume, divided by n_H^2, for gas with metallicity $Z = 0.3\,Z_\odot$, at redshifts $z = 0$ (solid line), $z = 2$ (dashed) and $z = 4$ (dot-dashed). At each redshift, we choose the gas density corresponding to the mean density in virialized haloes at that redshift (thus $n_H = 1.3 \times 10^{-3}$, 3.5×10^{-2} and 1.6×10^{-1} cm^{-3} for $z = 0$, 2 and 4 respectively). The dotted line indicates the cooling curve when no photoionizing background is present. The $z = 4$ curve is almost indistinguishable from this case. At low temperatures ($T \approx 10^{4.3}$ K) the $z = 0$ and $z = 2$ curves show a discontinuity, below which there is net heating rather than cooling.

halo as

$$T_{vir} = \frac{1}{2} \frac{\mu m_H}{k_B} V_c^2. \tag{18}$$

At each time-step in our calculations we compute the age of the halo and the cooling time, defined as

$$\tau_{cool} = \frac{\frac{3}{2} n_{tot} k_B T_{vir}}{\Lambda(n_H, T_{vir}, Z_{halo}, z)}. \tag{19}$$

Equating τ_{cool} to the age of the halo, we solve for the density of the gas which is just able to cool, and hence for the cooling radius, using the assumed density profile of the halo gas. We then calculate the mass and angular momentum of gas cooling in that time-step in the way described by Cole et al. (2000).[3]

2.4 Comparison with numerical simulations of the IGM

Our model of the IGM is highly simplified, but we only require it to predict a few volume-averaged quantities, namely the IGM temperature and the spectrum of the ionizing background. The advantage of our approach is one of speed, allowing rapid exploration of many different models. The disadvantages, compared to N-body/gas-dynamical simulations, are that it does not include the effects of spatial variations in the ionizing background (no radiative transfer), and includes the effects of gas density variations in only a very approximate way. These

[3] Equation (4.3) of Cole et al. (2000) contains a typographical error – the factor μm_H should appear in the numerator, not the denominator.
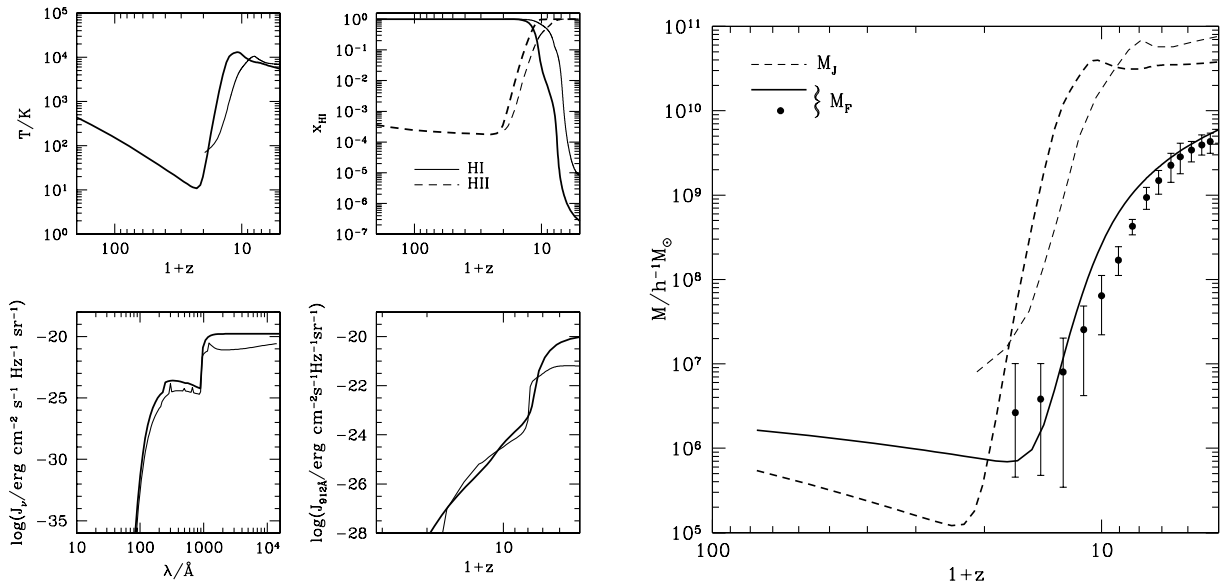
Figure 2. A comparison of properties of the IGM and the photoionizing background in our model (heavy lines) and in the numerical simulations of Gnedin (2000a) (thin lines). In our model we have assumed the same stellar emissivity as in the simulations. In the left-hand panel we show the volume-averaged temperature (top left), the fractions of neutral (solid lines) and ionized (dashed lines) hydrogen (top right), the spectrum of the ionizing background at $z \approx 9$ (bottom left) and the evolution of $J_\nu(912\,\text{Å})$ with redshift (bottom right). The right-hand panel compares the Jeans and filtering masses from our model with the simulation. Dashed lines show the Jeans mass, while points with error bars show the filtering mass measured from the simulations, and the solid line shows that predicted by our model.

limitations are likely to be most important just prior to full reionization, when there may be large spatial variations in the ionizing background, and in the ionization state and temperature of the IGM. However, in this paper we are interested chiefly in calculating how an ionizing background suppresses galaxy formation, and these suppression effects become strong only after the IGM has been reionized, when our approximation of a uniform ionizing background should be more accurate.

We have tested the effects of the approximations in our model for the evolution of the IGM and ionizing background by comparing it to the *N*-body/gas-dynamical numerical simulations of Gnedin (2000a), which include the effects of the detailed spatial distribution of gas and ionizing sources, as well as an approximate treatment of radiative transfer. To do this test, we input into our model the same volume-averaged stellar emissivity and spectrum as measured from the simulations, assuming also the same cosmological parameters. Fig. 2 compares predictions of our model with the same quantities measured from the simulations. Note that Gnedin's simulation was stopped at $z = 5$, so we cannot make any comparison at lower redshifts. The left-hand panel of Fig. 2 compares the volume-averaged IGM temperature, H I and H II fractions, and $J_\nu(912\,\text{Å})$ as functions of redshift, and also the background radiation spectrum at $z = 9$ (just prior to reionization for this model). The right-hand panel compares the Jeans and filtering masses predicted by our model with the values measured from the simulations by Gnedin (2000b). In the simulations, the filtering mass was determined by measuring the gas masses accreted by different haloes and fitting these with the formula (16).

Overall, the level of agreement between the two approaches is very good, although there are some differences in detail. The temperature of the IGM rises earlier in our model, and reionization occurs slightly earlier, presumably because the recombination rate in Gnedin's simulations is initially very high due to the ionizing sources forming in the highest density regions. The ionizing background at $z \approx 9$ is in reasonable agreement with that from the

simulation, although slightly higher. At wavelengths longwards of 912 Å our model predicts a significantly higher background. Here the gas is optically thin, so the details of absorption and the distribution of H I are unimportant. It seems therefore that the approximate radiative transfer used by Gnedin (2000a) somewhat underestimates the background in the optically thin case. This is also apparent in the bottom right panel, where we show $J_\nu(912\,\text{Å})$ as a function of redshift. Prior to reionization the two models predict very similar values, but afterwards our model reaches a significantly higher value than does Gnedin's. In the right-hand panel of Fig. 2 we compare the Jeans and filtering masses. The Jeans mass in our model begins to increase sooner than in Gnedin's simulations (as expected from the earlier temperature rise in our model), and this difference is reflected in the filtering mass. Nevertheless, our simple model of the IGM reproduces with reasonable accuracy the evolution of the filtering mass in the numerical simulation. For our purposes, this is the most important result of the comparison, because the largest effect of photo-ionization on galaxy formation is through the filtering mass, as we will see in Section 4.2.

3 MODEL FOR THE DYNAMICAL EVOLUTION OF SATELLITE GALAXIES

3.1 Model for dynamical friction and tidal stripping

When dark matter haloes merge, a new combined dark halo is formed. The largest of the galaxies they contained is assumed to become the central galaxy in the new combined halo, while the other galaxies become satellite galaxies in the new halo. These satellites evolve under the combined effects of dynamical friction, which makes their orbits sink towards the centre of the halo, and tidal stripping by the gravitational field of the host halo and central galaxy, both of the dark matter haloes originally surrounding the satellites and of the stars they contain. The Cole et al. (2000) model

included the effects of dynamical friction on the evolution of satellites, but did not include any treatment of tidal stripping. Since we are now interested in a more detailed study of the properties of the satellites around galaxies like the Milky Way (Benson et al. 2001b), we must improve our original model to include tidal effects on satellites. We do this by following the approach of Taylor & Babul (2001) (with a few modifications), following the orbits of satellites within host haloes and making simple analytical estimates of tidal effects (both 'static' tidal limitation and tidal 'shocks'). Taylor & Babul show that this simple model for the evolution of satellite haloes is able to reproduce well many of the results seen in high-resolution *N*-body simulations. We describe this part of our model briefly, referring the reader to Taylor & Babul for a detailed discussion, but will highlight the differences between our model and theirs.

We calculate the evolution of the orbit of each satellite galaxy in its host halo, under the influence of dynamical friction and tidal stripping. We specify the initial energy E and angular momentum J of the orbit (after the halo merger) in terms of the parameters $R_c^0/R_{vir,host}$ and $\epsilon = J/J_C$ respectively, where $R_c^0(E)$ is the radius of a circular orbit with energy E, $R_{vir,host}$ is the virial radius of the host halo, and $J_C(E)$ is the angular momentum of a circular orbit with energy E. We choose a constant value of $R_c^0/R_{vir,host}$ for all satellites. Our standard choice is $R_c^0/R_{vir,host} = 0.5$, which is representative of the median binding energy of satellite haloes seen in high-resolution *N*-body simulations (Ghigna et al. 1998) *at the output time of the simulation*. The typical value of $R_c^0/R_{vir,host}$ for satellites just entering their host halo should presumably be somewhat higher, since by the output time satellites will have lost some energy through dynamical friction. Lacking a direct measurement of the initial $R_c^0/R_{vir,host}$ from simulations, we will simply use 0.5 as a default, but will also explore other values to assess the impact of the uncertainty in this parameter on our final results. We select a value for the initial orbital circularity, $\epsilon = J/J_C$ by drawing a number at random in the range 0.1–1.0, which is a reasonable approximation to the distribution of circularities found by Ghigna et al. (1998). These choices for the initial orbital energy and angular momentum are the same as those of Bullock et al. (2000). Given the energy and angular momentum of the orbit, we determine the apocentric distance and begin integration of the orbit equations at that point, where tidal forces are weakest.

We model the dark matter in both the host and satellite haloes as an NFW density profile (Navarro, Frenk & White 1997), modified by the gravity of the galaxy which has condensed at the halo centre (the calculation of this adiabatic compression of the halo is described in detail by Cole et al. 2000). The galaxy at the centre of each halo is modelled as a combination of disc and spheroid. The disc has a density distribution given by

$$\rho_d(x, y, z) = \rho_{d,0} \exp\left[-\frac{(x^2 + y^2)^{1/2}}{r_d}\right] \text{sech}^2\left(\frac{z}{hr_d}\right), \quad (20)$$

where r_d is the disc radial scalelength, and h is the ratio of vertical to radial scalelength, which we take to be constant and equal to 0.1. The spheroid is modelled as a spherically symmetric $r^{1/4}$-law. The masses and sizes of these components are determined as described by Cole et al. (2000).

The satellite galaxy + halo moves under the influence of two forces. The first is just the net gravitational force due to the host halo and its central galaxy. The force due to the disc is calculated using the method of Kuijken & Gilmore (1989). The second force is that due to dynamical friction, which we estimate using

Chandrasekhar's formula (e.g. Binney & Tremaine 1987, section 7.1)

$$\boldsymbol{F}_{df,i} = -4\pi G^2 M_s^2 \ln\Lambda_i \rho_i B(x) \frac{\boldsymbol{v}_{rel,i}}{|\boldsymbol{v}_{rel,i}|^3}, \quad (21)$$

where $\ln\Lambda_i$ is the Coulomb logarithm, ρ_i is the local density, $B(x) = \text{erf}(x) - 2x \exp(-x^2)/\sqrt{\pi}$, $x = |\boldsymbol{v}_{rel}|/\sqrt{2}\sigma_i$, σ_i is the velocity dispersion, and $\boldsymbol{v}_{rel,i}$ is the relative velocity of the satellite and component i. We consider two components which contribute dynamical friction forces, namely the dark matter of the host halo and the spheroid of the central galaxy (which we treat together and indicate hereafter by a subscript 'h'), and the disc of the host halo galaxy (indicated by a subscript 'd'). We adopt the same values for the Coulomb logarithms as Taylor & Babul (2001) (namely 2.4 for the dark matter/spheroid and 0.5 for the disc), which fit the results of *N*-body simulations well. Taylor & Babul discuss in detail the possible choices for the Coulomb logarithms. While previous semi-analytic models have often used $\ln\Lambda_h = \ln M_h/M_s$, we prefer to use the same value as Taylor & Babul for this present work. The dynamical friction force depends upon the mass M_s of the satellite. We include in this mass that part of the satellite galaxy and its dark halo which has not yet been stripped by tidal forces.

For the disc velocity dispersion, we take $\sigma_d = V_c/\sqrt{2}$, where V_c is the rotation speed of the disc (computed for a spherically averaged disc), as did Taylor & Babul (2001). This results in an unrealistically high velocity dispersion when applied to the Milky Way (where the observed 1D velocity dispersion is approximately 30 to $40\,\text{km s}^{-1}$). We prefer to use the Taylor & Babul value at present, but find that using a lower value ($\sigma_d = 0.2V_c$) has almost no effect on the results presented in this paper (e.g., the galaxy luminosity functions of Fig. 9 are hardly affected by this change). For the dark matter/spheroid system we find σ_h by integration of the Jeans equation (assuming an isotropic velocity dispersion)

$$\frac{d(\rho_h \sigma_h^2)}{dr} = -\frac{GM_h(r)}{r^2} \rho_h(r), \quad (22)$$

where M_h is the total (i.e., dark plus baryonic) mass within radius r of the host halo. We assume that the dark matter follows the NFW profile for all radii outside of the virial radius. Cole & Lacey (1996) show that the velocity dispersion calculated in this way is in reasonable agreement with that measured in *N*-body simulations.

At each point in the orbit, we calculate the 'static' tidal limitation radius of the satellite galaxy + halo, r_t. This is the radius where the gravitational force of the satellite equals the sum of the tidal force from the host halo plus the pseudo-force due to the satellite's orbit,

$$\frac{GM_s(r_t)}{r_t^2} = \left[\omega^2 - \frac{d}{dR}\left(\frac{GM_h(R)}{R^2}\right)\right] r_t, \quad (23)$$

where R is the distance from the centre of the host halo, $M_s(r_t)$ is the total mass within radius r_t of the satellite, and ω is the instantaneous angular velocity of the satellite. Note that the factor of ω^2 is strictly accurate only for circular orbits. Here we follow Taylor & Babul (2001) and include this term for all orbits. For the purposes of this calculation and that of σ_h, the mass of the host halo disc is spherically averaged (the assumption under which equation 23 was derived). In all other calculations of satellite dynamics we use the density distribution of equation (20) to describe the disc. Equation (23) is valid under the assumption that the satellite is much smaller than the host halo, which is true for all but a very small fraction of satellites in our calculations.

Weinberg (1994) has argued that mass loss may occur at smaller radii than suggested by the above expression, due to heating of the satellite by gravitational shocking as it passes near the centre of the host halo. We adopt the approach of Taylor & Babul (2001) to estimate the effect of this tidal shocking, and refer the reader to that paper for a complete description of the method. Briefly, during each fast shock (i.e., any shock for which the time-scale is less than the internal orbital period of the satellite at its half-mass radius), we calculate the rate of heating by tidal forces. The energy thereby deposited in the satellite causes the satellite to expand, pushing some material beyond the tidal radius, and so allowing more material to be removed by tidal forces. We define the *effective* tidal radius $r_t^{\rm eff}$ as the radius in the original satellite density profile beyond which material has been lost. When the satellite has been heated, this effective tidal radius will therefore be less than r_t as defined by equation (23). We remove matter from the satellite in spherical shells outside of the effective tidal radius in the heated satellite. Note that we leave the density profile of material inside the effective tidal radius unchanged, so that the maximum circular velocity in the satellite remains unchanged until the effective tidal radius is reduced below the position of the peak of the rotation curve (i.e., $2.16r_s$ for a pure NFW dark matter rotation curve, but some other value when the baryonic contribution is included).

Mass beyond the effective tidal radius of the satellite is removed gradually on the shorter of the angular orbital time-scale, $2\pi/\omega$ (which becomes the orbital period for circular orbits), and the radial infall time-scale R/v_R. Taylor & Babul (2001) considered only the angular orbital time-scale, which results in very low mass loss rates for satellites on nearly radial orbits (as sometimes occur if the dynamical friction force is strong). It is then simple to calculate the mass remaining in the satellite (including both dark matter and baryonic components), and to use this to calculate the dynamical friction force exerted on the satellite.

The orbit equations are integrated until one of three conditions is met.

(i) The final redshift (i.e., the redshift at which we are studying the galaxy population) is reached. In this case we calculate the remaining mass and luminosity of the satellite galaxy after tidal limitation.

(ii) The host halo merges to become part of a new halo. The satellite halo then becomes a satellite of the new halo, and is assigned a new orbit in that halo. We begin integration of the orbit equations again, but starting with the previous value of the effective r_t for the satellite.

(iii) The satellite merges with the central galaxy (which we assume happens when the orbital radius, R, first reaches $R_{\rm merge}$, which we take to be the sum of the half-mass radii of the host and satellite galaxies)[4]. In this case we add to the central galaxy of the host halo the remaining mass of satellite galaxy at the time when the pericentre of its orbit first passed within $R_{\rm merge}$ (even though some of this mass may have been stripped off since that time, its orbit will carry it into the central galaxy in any case), using the rules described by Cole et al. (2000) (and possibly triggering a burst of star formation).

Stellar mass stripped from satellite galaxies is added to a diffuse stellar component of the host halo, but is not considered further in our models. Any cold gas stripped from the satellites is added to the hot gas reservoir of the host halo, and so may be able to cool again

---

[4] The model results are insensitive to the exact definition of merger time, as once R reaches such small radii it decreases very rapidly to zero.

---

at a later time. Finally, as the satellite galaxy orbits in the host halo it continues to form stars, which causes some of the cold gas mass of the galaxy to be ejected according to the supernova feedback prescription of Cole et al. (2000). In the case of satellites with shallow potential wells, this can significantly alter the mass of the galaxy along its orbit. Therefore the mass of this reheated gas is removed from the satellite halo during the orbit.

In Fig. 3 we show an example of a satellite orbit calculated using the above model. The host halo has a mass of $2 \times 10^{13}\, h^{-1}\,{\rm M_\odot}$, a virial circular velocity of $440\,{\rm km\,s^{-1}}$, and a concentration (defined here as the ratio of virial radius to NFW scale radius) of 5.9. The same three quantities for the satellite when it was still a separate halo are $3.5 \times 10^{12}\, h^{-1}\,{\rm M_\odot}$, $340\,{\rm km\,s^{-1}}$ and 5.6, respectively. However, in this example, the satellite halo has already lost mass while being a satellite in a progenitor of the current host halo, and this is why in this plot it starts from a mass of $8 \times 10^{11}\, h^{-1}\,{\rm M_\odot}$. In this plot, the time t is measured from when the host halo formed, and the satellite orbit begins at the apocentre (approximately $250\,h^{-1}$ kpc from the centre of the host halo). The orbit decays rapidly due to the effects of dynamical friction, so that the satellite makes three orbits before merging with the galaxy at the centre of the host halo at $t \approx 2.5$ Gyr. The mass of the satellite is seen to decrease most rapidly when the satellite is close to pericentre. Note also that until just before the first passage through pericentre, the mass of the satellite is unchanging, as before this time tidal forces are simply not strong enough to strip any mass from the halo.

We do not attempt here to model changes in the density profile of the satellite galaxy + halo within the tidal radius – the profile of the unstripped material is assumed to remain as it was before any stripping occurred; nor do we account for any changes in the global properties of a galaxy which has lost mass to tidal forces (i.e., the galaxy keeps the same scalelengths, star formation time-scale, etc. as it had before any mass loss occurred). Numerical simulations of
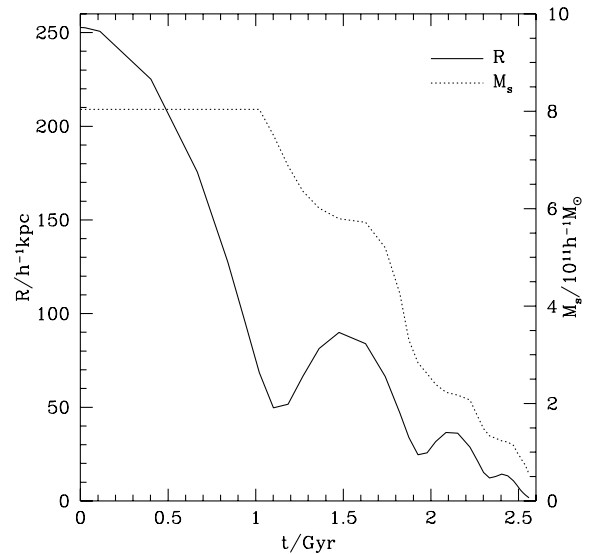
Figure 3. An example of the evolution a satellite galaxy orbit. The satellite enters the host halo at $t = 0$ and merges with the central galaxy of that halo after approximately 2.5 Gyr. The solid line shows the orbital radius of the satellite as a function of time, indicating that the orbit is decaying rapidly due to the effects of dynamical friction. The dotted line shows the remaining mass of the satellite. Note that the mass does not begin to decrease until the first passage through pericentre, as before this the tidal forces felt by the satellite are not strong enough to remove any mass. See text for more details.

satellites undergoing tidal interactions (e.g. Mayer et al. 2001) demonstrate that satellite mass profiles are affected by tidal interactions. Typically, they find a reduction in the amount of stellar mass within a given radius for radii within but comparable to the tidal radius. For low surface brightness galaxies (LSBs), this reduction can be up to a factor of around 2, but the effect is much weaker for high surface brightness galaxies (HSBs). At small radii, the stellar mass within a fixed radius is often increased by tidal interactions (through the production of a bar). The rotation curves of the galaxies are more seriously affected (presumably because the spherically distributed dark matter is less strongly bound than the stellar disc), often being reduced by a factor of 2 for LSBs (and somewhat less for HSBs).

Our calculation of merging times improves upon the simple estimates previously used in many semi-analytic models (which have often used results for satellites orbiting in isothermal haloes, with no tidal stripping). However, we find that on average our approach predicts comparable merging time-scales for satellite haloes to the simpler treatment in Cole et al. (2000), although some

fraction of satellites are predicted to have extremely long time-scales, as they lose so much mass through tidal stripping that dynamical friction forces become extremely weak.

3.2 Comparison with *N*-body simulations

For our present purposes, we are most interested in whether our model reproduces the abundance of satellite haloes, (or subhaloes) in a host halo typical of Milky Way-like galaxies. We therefore compare the predictions of our model for the number of satellite haloes with the results from high-resolution, dark-matter-only *N*-body simulations of Milky Way-like haloes in CDM models. Fig. 4 shows the comparisons with simulations of ΛCDM ($\Omega = 0.3$, $\Lambda = 0.7$) by Klypin et al. (1999) and of SCDM ($\Omega = 1$) by Moore et al. (1999). In the ΛCDM simulation, the host halo mass is $M_{\mathrm{halo}} = 1.1 \times 10^{12}\,\mathrm{M_{\odot}}$, and the minimum resolvable subhalo mass is $3.3 \times 10^8\,h^{-1}\,\mathrm{M_{\odot}}$ (corresponding to 20 particles). The corresponding quantities in the SCDM simulation are $1.0 \times 10^{12}\,h^{-1}\,\mathrm{M_{\odot}}$ and $1.6 \times 10^7\,h^{-1}\,\mathrm{M_{\odot}}$. We make the comparison in
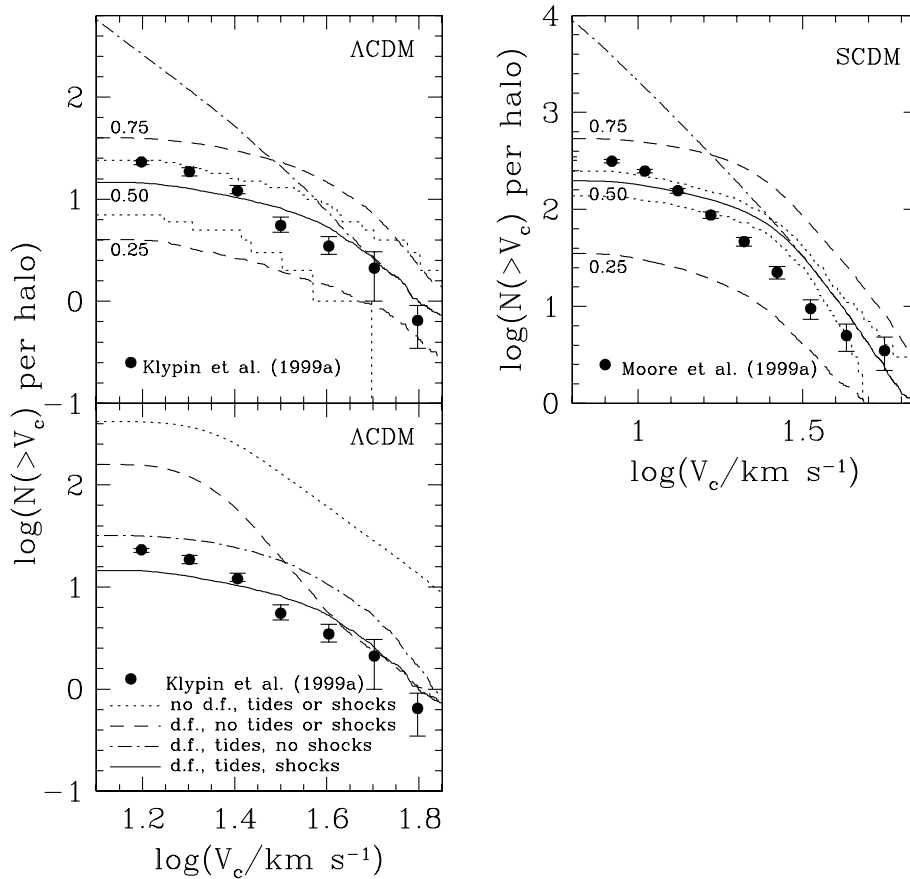

Figure 4. The number of satellite haloes as a function of circular velocity from the semi-analytic model compared to *N*-body simulations, for Milky Way-like haloes in CDM models. $N(> V_{\mathrm{c}})$ is the cumulative number of subhaloes per host halo, with V_{c} defined as the peak circular velocity of the subhalo. The simulations are of a ΛCDM model from Klypin et al. (1999) and of a SCDM model from Moore et al. (1999). In each panel, the solid points with error bars show the *N*-body simulation results, while the lines show the semi-analytic predictions for different assumptions. *Upper two panels:* the light solid lines show the semi-analytic prediction (averaged over 300 realizations), including in the semi-analytic model a cut on subhalo mass that mimics the limited resolution of the *N*-body simulation. The dotted lines on either side of the solid line show the 10–90 per cent range of the distribution around the mean value. The dot-dashed lines show the result from the semi-analytic model when no mass cut is applied. These results are for $R_{\mathrm{c}}^0/R_{\mathrm{vir,host}} = 0.5$. The dashed lines show the results if instead we assume $R_{\mathrm{c}}^0/R_{\mathrm{vir,host}} = 0.75$ or 0.25, including tidal stripping and the mass cut. *Lower panel:* this shows the contribution of different physical processes in the semi-analytic model, for the case of ΛCDM. The dotted line shows the predicted subhalo velocity distribution for the case of no dynamical friction, no static tidal limitation and no tidal shocking. Switching on dynamical friction produces the dashed line. Adding in static tidal limitation gives the dot-dashed line, and finally switching on tidal shocking produces the solid line. In this panel, all the curves are for $R_{\mathrm{c}}^0/R_{\mathrm{vir,host}} = 0.5$, and all include the same cut on subhalo mass as in the *N*-body simulation.

terms of the subhalo velocity function $N(> V_c)$, defined as the cumulative number of subhaloes per host halo with circular velocities greater than V_c, where V_c is defined as the peak circular velocity of the subhalo.

To compare our results to those of the N-body simulations, we run our semi-analytic model without baryons. Each dark matter halo then has a pure NFW profile. In the case of satellite haloes, the NFW profile is truncated beyond a radius r_t determined by the combined effects of static tidal limitation and tidal shocking. For an untruncated NFW profile, the circular velocity peaks at $r_m = 2.16 r_s$, where r_s is the NFW scale radius. We assume that the density profile of satellite haloes is unchanged within r_t, so the peak V_c is $V_{\mathrm{NFW}}(r_m)$ if $r_t > r_m$, and $V_{\mathrm{NFW}}(r_t)$ otherwise. We choose a host halo mass at $z = 0$ equal to the value in the simulation, run 300 different realizations of the halo merger tree, and then take the mean $N(> V_c)$ averaged over these realizations. In our model, subhaloes are completely destroyed only when they merge into the centre of the host halo. Tidal stripping reduces the mass of a halo, but is assumed never to destroy it completely. In our semi-analytic model, we can resolve much lower mass haloes than can be resolved in the N-body simulations. Since there can be a wide range of subhalo masses at a given value of the subhalo circular velocity V_c, it is essential to take into account this difference in mass resolution when we compare to the simulations. Therefore, to calculate $N(> V_c)$, we discard from the semi-analytic model all subhaloes with masses (within r_t) smaller than the minimum resolvable subhalo mass in the simulation. As Fig. 4 shows (compare the solid and dot-dashed curves), this mass cut produces a large reduction in the number of satellites below $V_c = 20$–$40\,\mathrm{km\,s}^{-1}$, and is very important for matching the simulation results.

The upper left and right panels of Fig. 4 show the comparison of our model with the ΛCDM and SCDM simulations respectively. The solid curves show the prediction for $R_c^0/R_{\mathrm{vir,host}} = 0.5$, including the mass resolution cut. The dotted lines on either side of the solid line show the 10–90 per cent range of the distribution seen among the different realizations. This range is larger for the ΛCDM than for the SCDM model, which mainly results from the smaller number of subhaloes per host halo in the former case. In the same panels, the upper and lower dashed lines show the effect of changing the assumed initial orbital energy to $R_c^0/R_{\mathrm{vir,host}} = 0.75$ and 0.25 respectively. We see that our standard value $R_c^0/R_{\mathrm{vir,host}} = 0.5$ gives significantly better agreement with the N-body simulations for both ΛCDM and SCDM.

The lower left-hand panel of Fig. 4 shows in more detail the separate effects of dynamical friction, static tidal limitation (i.e., r_t as defined by equation 23) and tidal shocks on the velocity distribution, for the ΛCDM case. All of the curves plotted there assume $R_c^0/R_{\mathrm{vir,host}} = 0.5$, and include the cut in subhalo mass corresponding to the resolution of the N-body simulation. The solid line includes all of the above processes, and is therefore identical to the solid line in the upper left-hand panel. The dotted line contains none of these processes, so subhaloes never merge and are never tidally stripped. Switching on dynamical friction results in the dashed line, which greatly reduces the number of high-V_c (relatively massive) subhaloes, but is much less important for the low-V_c haloes. Switching on static tidal limitation (and keeping dynamical friction switched on) results in the dot-dashed line. This greatly reduces the number of low-V_c haloes, as these are strongly affected by tidal forces, once the mass cut is included. The number of high-V_c haloes actually increases somewhat, since tidal limitation is able to reduce the mass of these haloes and so reduce

the strength of the dynamical friction forces which they experience. The remaining difference between the dot-dashed and solid lines is accounted for by tidal shocking. The overall effect of tidal stripping is to reduce the number of haloes at low V_c by a factor of ~ 10.

An important difference between our model and the similar calculation by Bullock et al. (2000) is we can resolve subhaloes within subhaloes (i.e., we record all branches of the merger tree, and so haloes merging with the final halo may have substructure of their own), whereas Bullock et al. considered only subhaloes (i.e., merging haloes were assumed to have no substructure of their own). This distinction is important, since, in our model of galaxy formation, every branch of the merger tree can potentially host a galaxy (providing its virial temperature exceeds $10^4\,\mathrm{K}$ and so is able to cool efficiently). Hence, when no tidal stripping is applied, we find many more satellites at a given V_c than did Bullock et al. However, when tidal stripping is included, both our model and that of Bullock et al. are in reasonable agreement with the N-body results (although Bullock et al. compared only with the ΛCDM simulations of Klypin et al. 1999).

We remind the reader that our choice of $R_c^0/R_{\mathrm{vir}} = 0.5$ was originally motivated by the measurement of the orbital energy distribution of all the satellites existing in a halo at the final output time of an N-body simulation. However, in our model, we use this R_c^0/R_{vir} as the initial value for each satellite after it joins the main halo. Plausibly, we should use a higher value, since when satellites first fall into a host halo they should be less bound than at any subsequent time. This would reduce the effectiveness of tidal limitation in our model (e.g., compare the curves for $R_c^0/R_{\mathrm{vir}} = 0.50$ and 0.75 in Fig. 4). Also, as noted above, we do not include any adjustment in the density profile of the material within the tidal radius in response to stripping of material from larger radii. This would be expected to make satellites less bound and to enhance the process of tidal stripping, and also to lower the circular velocity. These two effects work in opposite directions, but it is not clear which is the dominant process. N-body simulations suggest that the effect of the latter on the subhalo peak circular velocities V_c is in fact fairly small; Ghigna et al. (2000) find in their high-resolution simulations that for subhaloes where the tidal radius is larger than the initial peak-V_c radius, V_c typically changes by only ~ 20 per cent due to tidal effects. For now, we simply note that $R_c^0/R_{\mathrm{vir}} = 0.50$ does produce reasonable agreement with the numerical results, and so we adopt this throughout the remainder of this paper.

The model of Cole et al. (2000) calculates the concentration of dark matter haloes using the prescription proposed by Navarro et al. (1997). More recent work (Bullock et al. 2001; Eke, Navarro & Steinmetz 2001) has shown that the Navarro et al. method predicts concentrations which are too high at high redshifts. Since more concentrated haloes are harder to disrupt through tidal forces, this may have some impact on our results. We have carried out our calculations using the prescriptions of both Navarro et al. and Bullock et al. to determine halo concentrations. As expected, the Bullock et al. prescription results in a greater degree of tidal disruption for satellite haloes. However, the effect is quite small, reducing the model abundances in Fig. 4 at the 10–20 per cent level, and so does not affect our conclusions.

Our semi-analytic model includes the effects of baryonic collapse on the mass profiles of the host and satellite haloes, although this effect is turned off when we compare to pure dark matter N-body simulations. While baryonic dissipation makes satellite haloes more strongly bound, and so more resistant to tidal limitation, it also makes the tidal forces of the host halo and central

galaxy stronger. The central galaxy of the host halo also contributes to satellite destruction through its contribution to dynamical friction. In haloes of mass $\sim 10^{12} h^{-1} M_\odot$, the net result of including the baryonic components is to further reduce the number of satellite haloes (compared to a pure dark matter calculation). We find, for these haloes and with our standard galaxy formation model (see Section 4.2), that the number of satellites at a given V_c is reduced by around 40 per cent at $V_c = 60 \, \mathrm{km \, s^{-1}}$, and by about 60 per cent at $V_c = 15 \, \mathrm{km \, s^{-1}}$.

4   RESULTS

We are now able to explore in a self-consistent way the effects of a photoionizing background on the properties of galaxies, and also the effects of galaxies on the IGM. We will begin in Section 4.1 by obtaining a self-consistent model and exploring the evolution of the IGM and ionizing background. In Section 4.2 we examine the effects on the global properties of galaxies in the fiducial model of Cole et al. (2000). The properties of satellite galaxies will be explored in a separate paper (Benson et al. 2001b).

4.1   Properties of the IGM and ionizing background

4.1.1   Star formation history

Our starting point is the fiducial model of Cole et al. (2000), modified in the ways described in Sections 2 and 3. We use this to predict the star formation history and associated emissivity in ionizing photons as a function of redshift. We resolve all haloes that are able to cool in the redshift interval 0 to 25, to ensure that all ionizing photons are accounted for. To determine the spectrum of emission from these stars, we tabulate the mean star formation rate per unit volume from our model as a function of both cosmic time and metallicity, $\mathrm{d}^2 \rho_\star(t, Z)/\mathrm{d}t \, \mathrm{d}Z$. The stellar emissivity per unit volume at cosmic time t is then simply

$$F_\lambda(t) = \int_0^t \int_0^\infty \frac{\mathrm{d}^2 \rho_\star(t', Z)}{\mathrm{d}t' \mathrm{d}Z} \Phi(t - t', Z) \, \mathrm{d}Z \, \mathrm{d}t', \qquad (24)$$

where $\Phi(t, Z)$ is the spectral energy distribution of a stellar population of age t and metallicity Z, which we take from the models of Bruzual & Charlot (in preparation).

To account for the effects of absorption by dust and gas in galaxies on the ionizing emissivity, we multiply the above expression (24) at wavelengths $\lambda < 912 \, \text{Å}$ by a constant factor f_{esc}, defined as the fraction of ionizing photons produced by stars that escape through the dust and gas of the galaxy's interstellar medium (ISM). To calculate the effects of dust absorption for the non-ionizing radiation at $\lambda > 912 \, \text{Å}$, we use the same approach as in Cole et al. (2000). The value of f_{esc} for the ionizing photons is uncertain, both theoretically and observationally, so we will present results for two values, $f_{\mathrm{esc}} = 10$ per cent and $f_{\mathrm{esc}} = 100$ per cent, which we believe to bracket a reasonable range. The value $f_{\mathrm{esc}} = 100$ per cent results in our model in an emissivity in Lyman continuum (Lyc) ($\lambda < 912 \, \text{Å}$) photons from galaxies at $z = 3$ that agrees with the recent observational estimate for Lyman-break galaxies by Steidel, Pettini & Adelberger (2001) (after allowing for the differences in the assumed cosmological models), and predicts reionization of hydrogen at $z \approx 8$, compatible with measurements of the Gunn–Peterson effect in quasars, but perhaps too high, given recent observations of $z \sim 6$ quasars (Becker et al. 2001; Djorgovski et al. 2001). The observational estimate of $L_\nu(900 \, \text{Å})/L_\nu(1500 \, \text{Å})$ for the Lyman-break galaxies by Steidel

et al. implies $f_{\mathrm{esc}} \sim 10\text{–}40$ per cent, allowing for uncertainties in dust extinction and in the emission at $\lambda < 912 \, \text{Å}$ predicted by stellar models. The reason why we require a larger f_{esc} than Steidel et al. to produce the same net ionizing emissivity at $z = 3$ is that the Cole et al. (2000) model predicts too low a typical 1500-Å luminosity for Lyman-break galaxies, once we include dust. On the other hand, the value $f_{\mathrm{esc}} = 10$ per cent is more consistent with both observational (e.g. Leitherer et al. 1995; Steidel et al. 2001) and theoretical (e.g. Dove, Shull & Ferrara 2000; Benson et al. 2001a) estimates of the escape fraction at both low and high redshift. Our model with $f_{\mathrm{esc}} = 10$ per cent predicts an ionizing background that is in better agreement with observational estimates at $z < 4.5$, and predicts a redshift for H I reionization of $z \approx 5.5$, which is slightly lower than the current best estimates. These issues are discussed in more detail below. In the remainder of this work we will show results for both values of f_{esc} and/or comment on their differences.

We also include the contribution to the ionizing emissivity from quasars, according to the observational parametrization of Madau, Haardt & Rees (1999). Their parametrization is based on fitting observational data on numbers, magnitudes and redshifts of quasars at $z < 4.5$, assuming an Einstein–de Sitter cosmology. To obtain the emissivity in our chosen cosmology, we must allow for the dependence of the observationally inferred luminosities and number densities on the assumed cosmological model. We therefore use

$$\epsilon(z) = \epsilon_{\mathrm{MHR}}(z) \left[\frac{d_{\mathrm{L}}(z)}{d_{\mathrm{L}}^{(\mathrm{EdS})}(z)} \right]^2 \left[\frac{\mathrm{d}V(z)/\mathrm{d}z^{(\mathrm{EdS})}}{\mathrm{d}V(z)/\mathrm{d}z} \right], \qquad (25)$$

where $\epsilon_{\mathrm{MHR}}(z)$ is the emissivity from Madau et al. (measured from their fig. 2), $d_{\mathrm{L}}(z)$ is the luminosity distance, and $\mathrm{d}V(z)/\mathrm{d}z$ is the comoving volume per unit redshift. Functions with superscript (EdS) are calculated in the Einstein–de Sitter cosmology; those without superscript are calculated in the cosmology of our fiducial model. We use the same expression to extrapolate the quasar contribution to $z > 4.5$. We note that even at $z < 4.5$, the use of $\epsilon_{\mathrm{MHR}}(z)$ from Madau et al. involves a considerable extrapolation of the quasar luminosity function down to luminosities not directly observed.

We then use the total (i.e., stellar plus quasar) ionizing emissivity in calculating the thermal evolution of the IGM (and hence the filtering mass) and the ionizing background, both as functions of redshift. The stellar emissivity must be determined self-consistently with the feedback effects on galaxy formation from the IGM pressure and ionizing background, as described in Sections 2.2 and 2.3. We do this by means of an iterative procedure, starting from a galaxy formation model computed ignoring these feedback effects, calculating the ionizing background in this model, using this as input in calculating a revised model including the photoionization feedback effects, and repeating this cycle until we have a model whose star formation history is consistent with the photoionizing background that it produces. Cole et al. (2000) chose the parameters of their fiducial model to match certain observations of the local galaxy population, in particular the luminosity function of galaxies in the B and K bands. We find that if we keep the same parameter values as used by Cole et al. (2000), then when we include the photoionization feedback, our model still produces an acceptable fit to these luminosity functions. The only change is a small adjustment of Υ (which determines mass-to-light ratios) from 1.38 in Cole et al. (2000) to 1.32 (we adjust the recycled fraction in our chemical evolution model accordingly). As will be
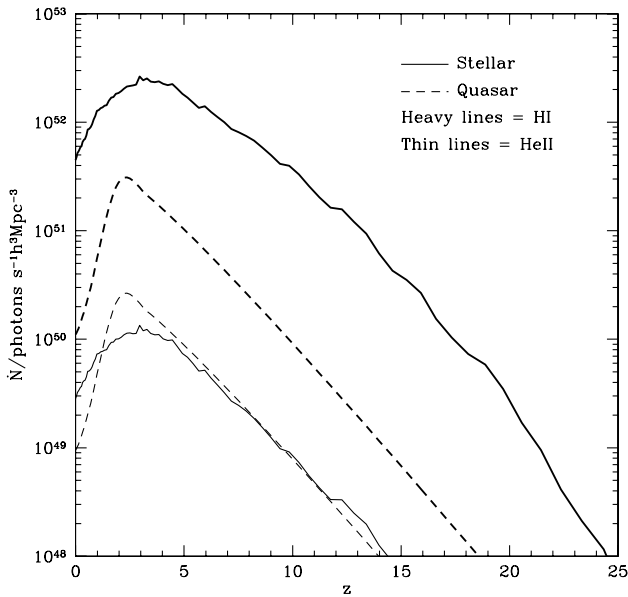
Figure 5. The emissivities in H I and He II ionizing photons (heavy and thin lines respectively) per comoving volume as a function of redshift. Solid lines show the emissivity from stars assuming $f_{esc} = 100$ per cent, while dashed lines show that from quasars.

Figure 6. The star formation rate per comoving volume in our $f_{esc} = 100$ per cent model as a function of redshift. The solid line represents the star formation rate used to compute the temperature of the IGM and the evolution of the ionizing background. The dashed line indicates the star formation rate in the penultimate iteration of the model, indicating that convergence has been reached over the range of interest. The dotted curve indicates the star formation rate in the model of Cole et al. (2000), scaled to the value of Y in our standard model for the purposes of this comparison. We also show a model with the effects of supernovae feedback switched off, both with (dot-dashed line) and without (short-dashed-long-dashed line) the effects of photoionization feedback included. The epochs of H I and He II reionization in the standard model are marked by vertical dotted lines. The small discontinuities in the star formation rate arise as we recompute our model at several intervals in redshift to ensure all haloes are resolved.

discussed in more detail in Section 4.2.1, the faint-end slopes are now somewhat flatter than before, but this is consistent with recent determinations of the luminosity functions.

The net emissivity in H I and He II ionizing photons from stars and quasars for our $f_{esc} = 100$ per cent model is shown in Fig. 5. (For $f_{esc} = 10$ per cent, the stellar emissivities evolve in a very similar way, only an order of magnitude lower in amplitude.) For $f_{esc} = 100$ per cent, stars always dominate the production of H I ionizing photons, but quasars dominate the production of He II ionizing photons until $z \approx 1$, when the rapidly falling quasar emissivity leads to stars becoming the dominant source. At very high redshifts ($z \gtrsim 10$) the shape of the stellar emissivity curve is essentially fixed by the rate at which dark matter haloes sufficiently hot to cool through atomic processes are forming (with an efficiency set by the star formation time-scale and feedback prescription of the model). It is interesting that this results in a redshift dependence which tracks quite closely that of the quasars. It should be noted, however, that the quasar emissivity at high redshifts ($z \gtrsim 4$) is highly uncertain, since it is based on an extrapolation of the quasar luminosity function and its evolution at lower redshifts. We will comment below on the effects of ignoring the quasar emissivity on our results.

Fig. 6 shows the star formation rate per unit comoving volume in our $f_{esc} = 100$ per cent model as a function of redshift. The comparison of the solid line (final iteration) and dashed line (penultimate iteration) shows that the model has converged to a self-consistent star formation history in the presence of the photoionization feedback.[5] The dotted line shows for comparison the star formation history from the model of Cole et al. (2000), with no photoionization feedback and scaled to the value of Y used in

---

[5] We restart our calculation of the star formation rate at several intervals in redshift to ensure that all haloes are resolved. Since the merger trees used in our model do not reproduce exactly the Press–Schechter mass function at an earlier redshift, this results in small discontinuities in the star formation rate visible in Fig. 6.
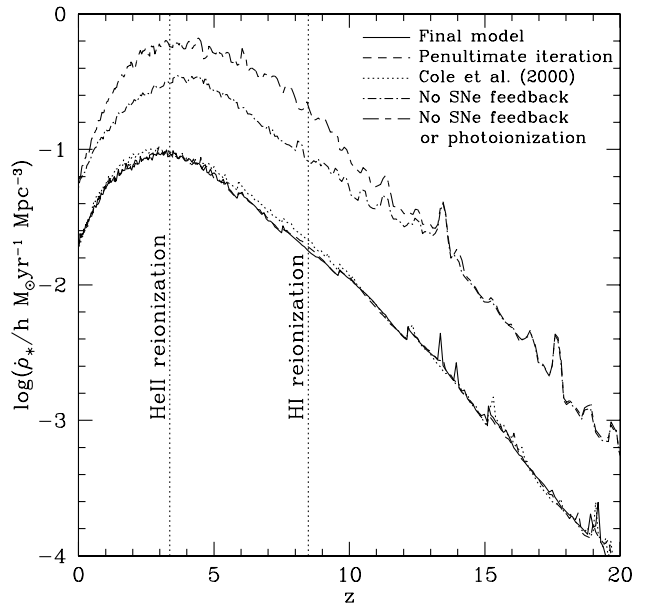
our standard model. For $z \gtrsim 10$ this is identical to that of our new model, a fact which is not surprising, since we use the same parameters as did Cole et al., and at these redshifts photoionization has yet to have much effect on the IGM. (Recall that gas in haloes with $T_{vir} \lesssim 10^4$ K is assumed to be unable to cool, even in the absence of an ionizing background.) Beginning just before H I reionization (as the IGM is being reheated) the star formation rate in our model falls below that of Cole et al. as the filtering mass rises. By $z \approx 4$ the star formation rate has recovered to the Cole et al. value as the continued formation of structure has created many haloes well above the filtering mass, and it is these which contribute most to the star formation rate. (The filtering mass is growing only rather slowly during this period.) The reionization of He II leads to a second episode of reheating, leading to an increase in the filtering mass which again suppresses star formation rates below the Cole et al. values. Once again, by $z = 0$ the differences have become very small, as star formation becomes dominated by galaxies in haloes well above the filtering mass. The effect is rather small, however, with star formation rates being reduced by around 25 per cent at most. The reason why the effects of photoionization feedback on the star formation history are quite modest in our model is that we also include supernova feedback according to the prescription of Cole et al. This greatly suppresses star formation in haloes with circular velocities $V_c \lesssim 200$ km s^{-1}, which includes the range of halo masses that are also affected by photoionization feedback. For $f_{esc} = 10$ per cent the star formation history is very similar, except that the

suppression of the star formation rate below the Cole et al. value at $z = 6-9$ is greatly reduced.

For comparison, we have also computed a model in which the feedback from supernovae is completely turned off. The other parameters in this model are identical to those in Cole et al. (2000), apart from Υ, which is reduced to 0.95 to match the bright end of the present-day galaxy luminosity function (see Section 4.2.1). The star formation rate as a function of redshift in this model, with $f_{esc} = 100$ per cent and photoionization feedback turned on, is shown by the dot-dashed line in Fig. 6. The absence of feedback from supernovae results in a much higher star formation rate than in our standard model at all redshifts, but especially at high redshift, where star formation is occurring mostly in small haloes which are the most strongly affected by supernova feedback. As a result, reionization occurs significantly earlier in this model, at $z = 11.5$ for H I and at $z = 4.5$ for He II. The short-dashed-long-dashed line shows the star formation rate when feedback from supernovae and from photoionization are both turned off. The reduction in the star formation rate after reionization due to photoionization feedback is seen to be much larger when there is no feedback from supernovae, than in our standard model which includes supernova feedback.

4.1.2 Thermal and ionization history of the IGM

Fig. 7 shows various properties of the IGM in our $f_{esc} = 100$ per cent model. We also show key properties for a model with a lower escape fraction, $f_{esc} = 10$ per cent, and for a model with a uniform IGM (i.e., $f_{clump} = 1$).

The upper left-hand panel shows the volume-averaged temperature of the IGM as a function of redshift. At around $z = 20$, ionizing photons from stars begin to heat the IGM [at higher redshifts the temperature scales as expected from adiabatic expansion, $T_{IGM} \propto (1 + z)^2$]. This results in the gas reaching a temperature of approximately 10^4 K at $z \approx 10$ (somewhat before reionization of H I), at which point atomic cooling processes balance the photoheating. The temperature then decreases until $z \approx 7$, when the photoionization of He II by emission from quasars leads to a second period of heating which lasts until $z \approx 5$. After this, the gas cools rapidly until $z = 0$, the cooling being due to adiabatic expansion (due both to Hubble expansion and the expansion of the gas in voids). The redshifts of reionization of H I and He II in the standard model are marked by vertical dotted lines. (We define the redshift of reionization somewhat arbitrarily as the point where 99 per cent of the species in question has been ionized. Since reionization takes place rapidly, the exact definition is not important.) With a uniform IGM, the temperature evolution is unchanged up to just before the reionization of H I, since the clumping in the standard model is relatively small at these high redshifts and the volume-weighted mean temperature shown here is dominated by the contribution from gas close to the mean density. At lower redshifts the clumpy IGM cools more rapidly as adiabatic expansion of gas in voids cools the gas (and these regions are strongly weighted in the volume-averaged temperature). The $f_{esc} = 10$ per cent model heats the Universe later, as expected. There is little difference in the peak temperature reached, which is essentially fixed by atomic physics, and the late-time temperatures are very similar to those of the standard model.

The upper right-hand panel shows the mean ionization state of hydrogen and helium in the IGM as a function of redshift. The quantity plotted is the average fraction of hydrogen or helium in each ionization state (e.g., for the H I fraction x_{HI}, we plot

$\langle n_{HII} \rangle / \langle [n_{HI} + n_{HII}] \rangle$, where $\langle \rangle$ denotes a volume average). Reionization is a much more rapid process than reheating (as has been noted previously by Gnedin & Ostriker 1997 and Valageas & Silk 1999, for example). For $f_{esc} = 100$ per cent, He I is ionized by stellar photons almost simultaneously with H I at a redshift of 8, but He II is not reionized until much later ($z \approx 4$) when the harder ionizing photons from quasars become abundant. Note that the initial decline in x_{HI} is similar in clumpy and uniform IGMs. As noted above, at these redshifts volume-averaged quantities in the clumpy case are dominated by gas close to the mean density, so we do not expect much difference from the uniform case. Once started, though, reionization is completed much more rapidly in the case of a uniform IGM. In the clumpy IGM the completion of reionization is delayed by the small fraction of high-density gas, which is reionized last. With an escape fraction of 10 per cent, reionization does not occur until much later, at $z \approx 5.5$ (which may be slightly too low to be consistent with recent measurements of the Gunn–Peterson effect in quasars at $z \approx 6$; Fan et al. 2000; Becker et al. 2001; Djorgovski et al. 2001), but otherwise proceeds in much the same way. If we ignore the quasar contribution to the emissivity, we find that H I reionization redshift is essentially unchanged for $f_{esc} = 100$ per cent and reduced to $z \approx 5$ for $f_{esc} = 10$ per cent. A larger effect occurs for the He II reionization redshift, which is significantly reduced by neglecting the quasar contribution. For $f_{esc} = 100$ per cent it falls to $z \approx 2$, while in the $f_{esc} = 10$ per cent case He II is not reionized even by $z = 0$ when the quasar contribution is ignored.

The lower left-hand panel shows the evolution of the Jeans and filtering masses with redshift for $f_{esc} = 100$ per cent. Note that both of these are defined as total (i.e., dark matter plus baryonic) masses, not baryonic masses. The Jeans mass simply tracks the temperature of the IGM, while the filtering mass approximately tracks the Jeans mass, but with a significant delay. As a result, the filtering mass can be up to 1000 times lower than the Jeans mass during the first episode of reheating. However, at lower redshifts the two are much more comparable, and by $z = 0$ the filtering mass is around 60 per cent of the Jeans mass. Note that in the period of cooling from $z = 9$ to $z = 7$ (during which time H I reionization has finished, but He II reionization has yet to begin) the Jeans mass decreases slightly with time, and as a consequence the filtering mass grows only slowly. Using a uniform rather than a clumpy IGM affects M_F only at low redshifts, where the lack of cool void gas in the uniform case results in a slightly larger filtering mass. Although the filtering mass does not begin to rise until later in a model with $f_{esc} = 10$ per cent, it actually rises above the $f_{esc} = 100$ per cent model at late times, since the IGM has actually been hotter in the recent past in this model.

The lower right-hand panel shows the values of the halo circular velocity at the virial radius, and corresponding halo virial temperature (equation 18), that correspond to the halo filtering mass. Also shown for comparison is the average IGM temperature as a function of redshift. It can be seen that, according to the filtering mass prescription of Gnedin (2000b) that we use, the critical halo virial temperature below which baryonic collapse into haloes is suppressed by 50 per cent in mass can be much greater than the IGM temperature (by a factor of 60 at $z = 0$ in our standard model). In our standard model, this temperature peaks at $T_{vir} \approx 10^5$ K, corresponding to $V_c \approx 60$ km s^{-1}, even though the IGM temperature is never significantly above 10^4 K. Clearly, it will be very important to test the accuracy of Gnedin's filtering mass prescription in greater detail using future high-resolution simulations. However, we note that similar results for the halo

circular velocity below which baryonic collapse is 50 per cent suppressed were also found by Quinn et al. (1996), from SPH simulations, and Thoul & Weinberg (1996), using a 1D hydro code (they both found $V_c \approx 50 \, \text{km s}^{-1}$ at $z \approx 2$). The thermal history of the IGM is somewhat sensitive to the inclusion of the quasar contribution to the emissivity. Leaving out this contribution reduces the filtering mass at $z = 0$ by factors of 1.4 and 2.4 for the $f_{esc} = 100$

and 10 per cent cases respectively (although in both cases most of this difference is due to the lack of quasar emissivity at $z \lesssim 4$, where the quasar luminosity function is reasonably well constrained). Thus the quasar emissivity contributes significantly to the filtering mass, and hence to the suppression of low-mass galaxy formation, but the majority of the contribution arises at low redshifts where the quasar emissivity is reasonably well known.
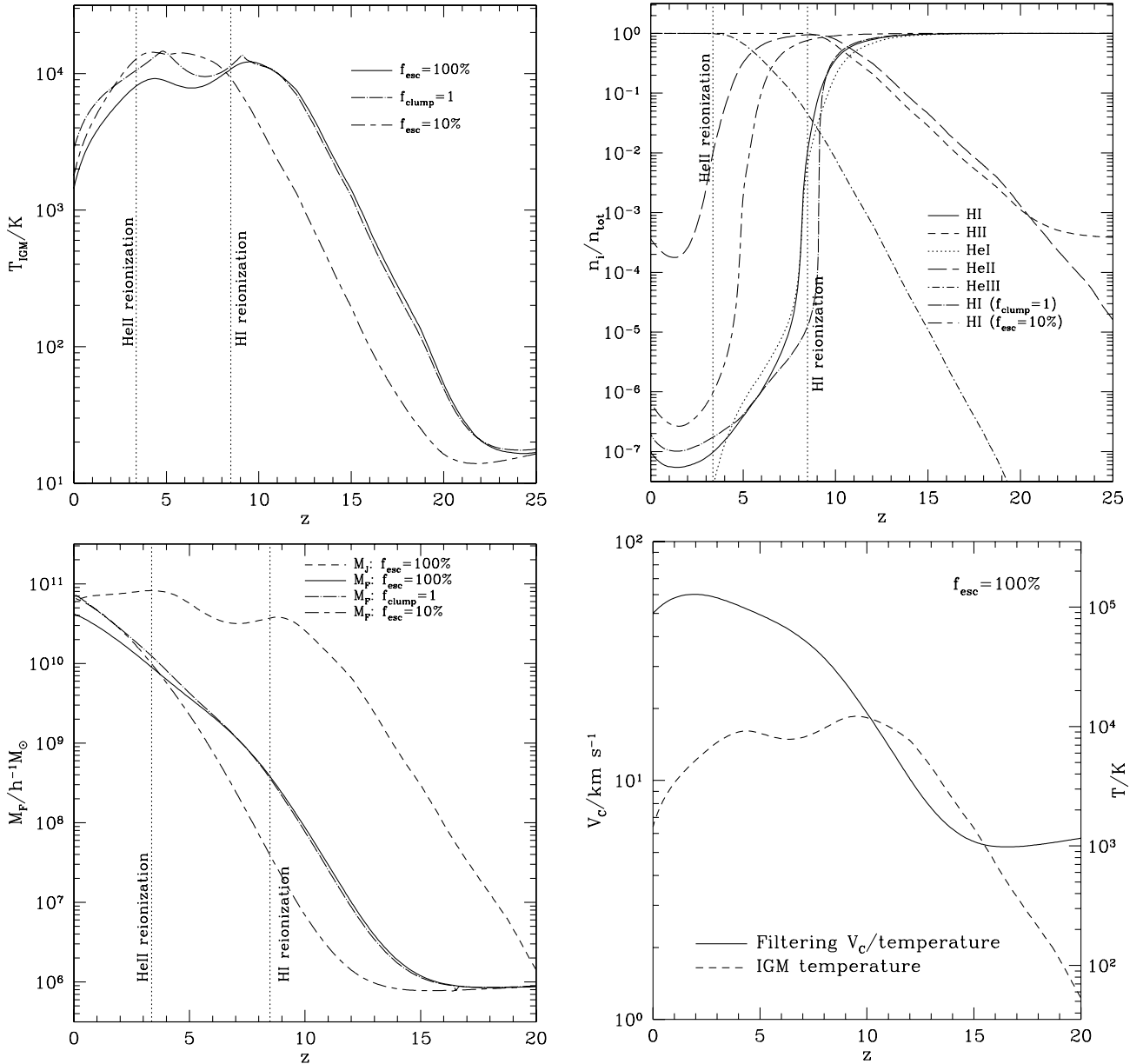
Figure 7. *Upper left-hand panel:* The volume-averaged temperature of the IGM as a function of redshift. The redshifts of reionization for H I and He II in our standard model are indicated by vertical dotted lines. The solid line shows the results for $f_{esc} = 100$ per cent, with the other lines showing the results for $f_{clump} = 1$ and $f_{esc} = 10$ per cent as indicated in the panel. *Upper right-hand panel:* The volume-averaged ionization state of hydrogen and helium as a function of redshift. The quantity shown is n_i/n_{tot}, where n_{tot} is the total abundance of the element in question in all ionization states. The epochs of H I and He II reionization in the $f_{esc} = 100$ per cent model are indicated by vertical dotted lines (we define these as the time at which n_i/n_{tot} reaches 0.99). We also show n_{HI}/n_{tot} for models with $f_{clump} = 1$ and $f_{esc} = 10$ per cent, as indicated in the panel. *Lower left-hand panel:* The Jeans mass (dashed line) and filtering mass (line types as defined in the panel) as a function of redshift for $f_{esc} = 100$ per cent. The filtering mass is also shown for models with $f_{clump} = 1$ and $f_{esc} = 10$ per cent. The redshifts of reionization for H I and He II in the $f_{esc} = 100$ per cent model are indicated by vertical dotted lines. *Lower right-hand panel:* The solid line shows the halo circular velocity (left-hand axis) or halo virial temperature (right-hand axis) that corresponds to the filtering mass in the $f_{esc} = 100$ per cent model. The dashed line shows the mean IGM temperature (repeated from the upper left panel).

4.1.3 The ionizing background

Fig. 8 shows the evolution with redshift of the predicted ionizing background at the Lyman limit, $J_\nu(912\,\text{Å})$. This is compared to observational estimates from the proximity effect in quasar spectra (e.g. Scott et al. 2000, and references therein), and upper limits from observational searches for $H\alpha$ fluorescence from extragalactic HI clouds at low redshift (Vogel et al. 1995) and for $Ly\alpha$ fluorescence from Ly-limit clouds at high redshift (Bunker, Marleau & Graham 1998). It can be seen that the background predicted for $f_{\text{esc}} = 10$ per cent is reasonably consistent with observational estimates, while that predicted for $f_{\text{esc}} = 100$ per cent is 5–10 times too high. While this comparison seems to favour the model with $f_{\text{esc}} = 10$ per cent, the estimated ionizing emissivity of Lyman-break galaxies at $z \approx 3$ is reproduced only in our model with $f_{\text{esc}} = 100$ per cent. This contradiction in part arises because our simple IGM model appears to underestimate the opacity of the IGM to ionizing photons at epochs when the IGM has been almost completely reionized. According to Madau et al. (1999), at $z \lesssim 5$, the Lyc opacity is dominated by the discrete absorbing clouds with neutral hydrogen column densities $N_{\text{HI}} \sim 10^{17}\,\text{cm}^{-2}$, which produce the Lyman-limit absorption features seen in quasar spectra. Madau et al. estimate the opacity as a function of redshift based on the observed statistics of quasar absorption lines, and find that the Universe becomes optically thin to Lyc photons only at $z \lesssim 1.6$. In contrast, our IGM model, which lacks these absorbing clouds, already becomes optically thin to ionizing photons at $z \sim 6$ for the case $f_{\text{esc}} = 100$ per cent. This explains why our standard model produces a much larger ionizing background at $z \approx 3$ than Steidel et al. (2001) calculate from combining their estimate of the ionizing emissivity of
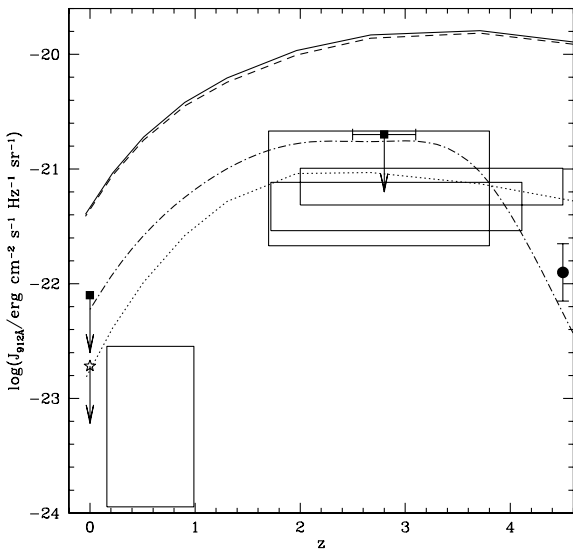
Lyman-break galaxies (which our model matches) with Madau et al.'s estimate of the Lyc opacity.

We would need to develop a much more sophisticated IGM model in order to include the effect of discrete clouds on the Lyc opacity in a way that was both self-consistent and agreed with observations of quasar absorption lines. According to our models, photoionization affects galaxy formation primarily through the effect of the IGM pressure (which mainly depends on the redshift of reionization) rather than on the cooling within haloes (which depends on the ionizing background at the redshift when the halo forms). Therefore we believe that the deficiencies of our model as regards predicting the ionizing background after reionization should not seriously affect the predictions that we make for galaxy formation.

4.2 Effects on galaxy properties

We now use our model, together with the properties of the IGM and ionizing background calculated in the previous subsection, to investigate the effects of photoionization feedback on the global properties of galaxies at $z = 0$, and to consider briefly the effects on galaxies at higher redshifts.

4.2.1 Luminosity functions

In Fig. 9 we present the B- and K-band luminosity functions for this model at $z = 0$, and compare them to a selection of observational data. Note that all model galaxy luminosities include extinction by dust, calculated using the model of Ferrara et al. (1999) as described by Cole et al. (2000). The heavy solid line shows the result from our $f_{\text{esc}} = 100$ per cent model, while the thin solid line shows that from the model of Cole et al. Brighter than $L_\star$, the galaxy luminosity function is mostly unaffected by the inclusion of the effects of photoionization. Fainter than this, differences become apparent, with the luminosity function being much flatter in our present model than in the Cole et al. model. At $M_B - 5\log h \approx -13$ the difference in amplitude of the B-band luminosity functions is about a factor of 4. Similar behaviour is seen in the K band. Compared to the Cole et al. model, our new model is in appreciably better agreement with recent observational determinations in the B and K bands from the 2dFGRS and 2MASS galaxy surveys, by Madgwick et al. (2001) and Cole et al. (2001) respectively, although in the K band the predicted slope is still slightly too steep.

Fig. 9 also shows the relative importance of the new effects we include compared to the Cole et al. (2000) model. The dot-dashed line shows the effect of including tidal stripping of satellite galaxies, but no photoionization feedback; this is seen to reduce the number of faint galaxies very slightly. The dashed line shows the effect of turning on the effect of IGM pressure (through the filtering mass), but not the effect of the ionizing background on cooling in haloes, while the dotted line has the modified cooling in haloes turned on, but not the filtering mass. Comparing these, we see that for the photoionization feedback it is primarily the effects of the IGM pressure which suppress the galaxy formation, while the reduction in cooling within haloes has a smaller effect. (Note that these three luminosity functions are calculated using the same value of Υ as for the standard model for the purposes of this comparison.) If instead of using Gnedin's (2000b) formula (equation 16) for the mass of gas accreted by haloes, we simply assume that accretion is completely efficient for haloes with $V_c > 60\,\text{km}\,\text{s}^{-1}$ but that no accretion occurs for lower mass haloes, we

Figure 8. Ionizing background versus redshift. The solid and dot-dashed lines are the predicted background intensity at the Lyman limit, $J_\nu(912\,\text{Å})$, for $f_{\text{esc}} = 100$ and 10 per cent respectively. The dashed and dotted lines show the separate contributions from stars and quasars in the model with $f_{\text{esc}} = 100$ per cent. The rectangular boxes and the data point at $z = 4.5$ are observational estimates based on the proximity effect (Batjlik, Duncan & Ostriker 1988; Kulkarni & Fall 1993; Williger et al. 1994; Cooke, Espey & Carswell 1997; Scott et al. 2000). The upper limits are based on searches for $H\alpha$ (Vogel et al. 1995, square; Weymann et al. 2001, star) or $Ly\alpha$ (Bunker, Marleau & Graham 1998) fluorescence; the last limit is somewhat model-dependent.
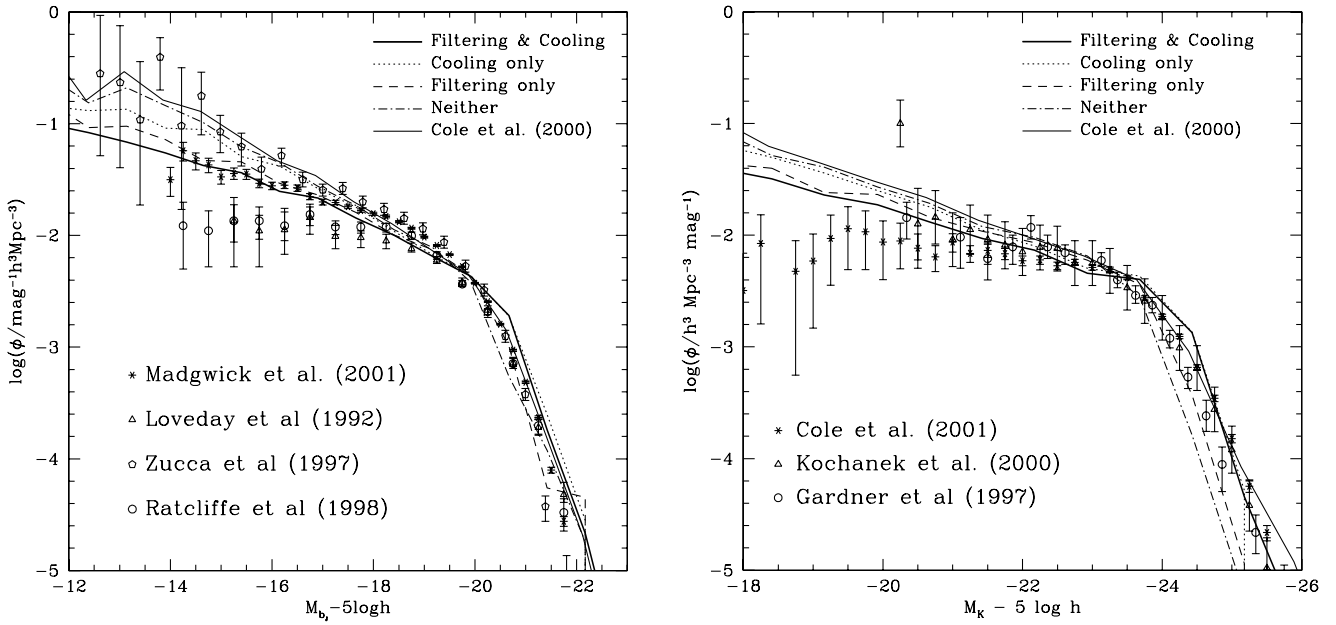
Figure 9. Galaxy luminosity functions at $z = 0$. The left-hand panel shows the luminosity function in the b_J band, while the right-hand panel shows that in the K band. In each panel, the heavy solid line shows the prediction of our $f_{esc} = 100$ per cent model with photoionization and supernovae feedback, and the thin solid line shows the model of Cole et al. (2000). Dotted lines show our standard model with the effects of the filtering mass switched off, dashed lines show the standard model with photoheating of gas in haloes switched off, and dot-dashed lines show the standard model with both of these effects switched off. All model luminosity functions include the effects of dust. The symbols show observational data.

find a luminosity function which is close to the 'Cooling only' line in Fig. 9 for around two magnitudes below L_*, but then flattens significantly and actually drops below the 'Filtering & Cooling' line at fainter magnitudes.

In Fig. 10 we show the B-band luminosity function in a model where photoionization feedback is included, with $f_{esc} = 100$ per cent, but feedback from supernovae is turned off. The star formation history for this same model was presented in Section 4.1.1. We choose $\Upsilon = 0.95$ to match the amplitude of the observed luminosity function at $L \sim L_\star$. The other parameters are the same as in Cole et al. (2000) and in our standard model. The low value of Υ is required because most gas which has cooled is locked up into small objects, leaving little to form bright galaxies. Strictly speaking, a value of $\Upsilon < 1$ is unphysical, because it requires a negative mass in brown dwarfs (defined here as objects with $m < 0.1\,M_\odot$). However, the same results as for $\Upsilon = 0.95$ could be obtained by small modifications to the IMF at $0.1 < m < 1\,M_\odot$, reducing the mass in low-mass stars which anyway contribute negligibly to the light from stellar populations. This 'no SNe feedback' model gives an acceptable match to the observed luminosity function at the bright end (except possibly at the highest luminosities). It predicts a faint-end slope that is much steeper than in our standard model, but only slightly steeper than the Cole et al. model, which had supernova feedback but no photoionization feedback, and also only slightly steeper than the measurement of Zucca et al. (1997). The faint-end slope is still much flatter than in a model with no feedback of any type.

We emphasize that the 'no SNe feedback' model we have presented here is by no means a 'best-fitting' model, since we have not varied other parameters to achieve a better match to the luminosity function, nor have we considered other observational constraints as Cole et al. did. (Preliminary analysis suggests that a model with only photoionization feedback has difficulties in matching the colours and sizes of present-day galaxies.) However,
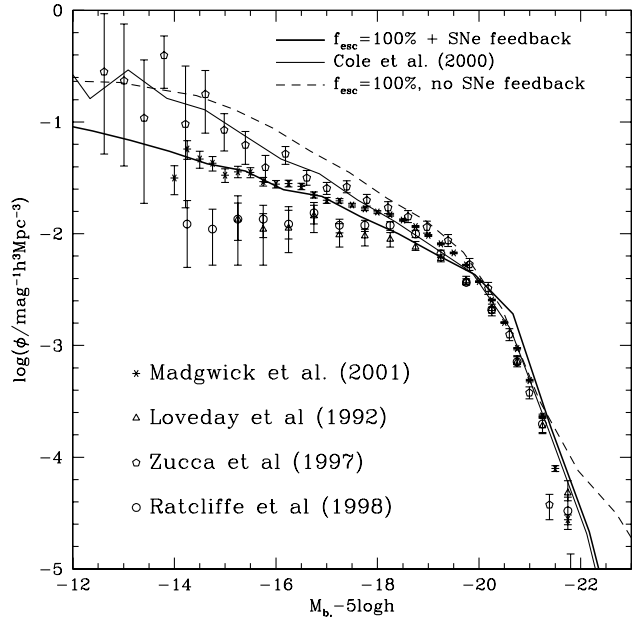

Figure 10. The b_J-band galaxy luminosity function at $z = 0$. The heavy solid line shows the prediction of our standard model with photoionization feedback, and the thin solid line shows the model of Cole et al. (2000). The dashed line shows a model with the effects of photoionization included, but without any feedback from supernovae. All model luminosity functions include the effects of dust. The symbols show observational data.

the prediction for the faint-end slope of the luminosity function is expected to be fairly robust, so we conclude that if the slope measured in the largest and most recent surveys (e.g. Madgwick et al. 2001) is correct, then photoionization feedback on its own does not produce a slope as flat as in the real Universe. We defer a

more detailed study of models without supernova feedback to a future paper.

4.2.2 Tully–Fisher relation

Fig. 11 shows the *I*-band Tully–Fisher relation of galaxies in our $f_{esc} = 100$ per cent model, compared to the observational data of Matthewson, Ford & Buchhorn (1992). For constructing the model relation, we select galaxies in the same way as in Cole et al. (2000), namely we select only star-forming spiral galaxies, but also select only those galaxies which have not been seriously disrupted by tidal forces (specifically we remove any galaxy which has lost more than 25 per cent of the mass of its disc through tidal stripping). These strongly tidally disrupted galaxies are unlikely to be recognizable as discs. If we do not remove these galaxies, then the Tully–Fisher relation in our model shows a scatter to very faint magnitudes at low circular velocities, but for circular velocities $V_c \gtrsim 100$ km s^{-1} the removal of these galaxies has little effect.

Fig. 11 also shows the model prediction of Cole et al. (2000), from which it can be seen that the differences from our new standard model are quite small. We also plot lines showing the result of switching off the effects of photoionization (as described in Section 4.2.1 and also in the figure caption). It can be seen from these curves that tidal stripping makes little difference to the Tully–Fisher relation, while the modified cooling in haloes has the larger effect at high luminosities and the IGM pressure the larger effect at low luminosities. In any case, photoionization does not help remove the offset in the predicted zero-point of the Tully–Fisher relation relative to the observed one. This offset persists to
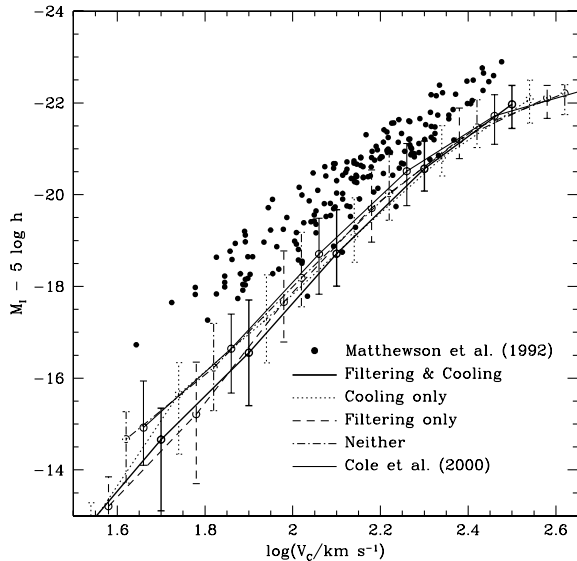

Figure 11. The Tully–Fisher relation in the *I* band at $z = 0$. The lines show the predicted median relation, with error bars indicating the 10 and 90 per cent intervals of the distribution, while the filled dots show observational data of Matthewson et al. (1992). The heavy solid line shows the $f_{esc} = 100$ per cent model of this paper, while the thin solid line shows the prediction from Cole et al. (2000). Only star-forming spiral galaxies are included in the model relation, with magnitudes corrected to their face-on value including the effects of dust. The velocities plotted are the circular velocity at the half-mass radius of the galaxy disc. Dotted lines show our standard model with the effects of the filtering mass switched off, dashed lines show the standard model with photoheating of gas in haloes switched off, and dot-dashed lines show the standard model with both of these effects switched off.

bright magnitudes, where photoionization has little effect on galaxy formation.

4.2.3 Further properties of the model at z = 0

We now briefly consider the effects of photoionization feedback on some other predicted properties of galaxies at $z = 0$. We consider the same properties as were compared with observational data in Cole et al. (2000). Most of these comparisons concerned fairly luminous galaxies, for which the properties in our new model are very similar to those of the Cole et al. fiducial model, so we just summarize the main results here.

Cole et al. (2000) computed the distribution of disc scalelengths of spiral galaxies at different luminosities, and compared to the observational data of de Jong & Lacey (2000), finding good agreement in the magnitude range they considered, $-19 > M_I - 5 \log h > -22$. Our model produces almost identical results, as may be expected for these bright galaxies.

Table 2 compares the fractions of S, S0 and E galaxies in our model brighter than $M_B - 5 \log h = -19.5$ (i.e., $L_\star$) with Cole et al. (2000) and with observational data. We assign morphological types to our model galaxies based on their bulge-to-total luminosity ratio in the *B* band, B/T_B, (including dust extinction). Galaxies having $B/T_B < 0.4$ are classed as S, those with $B/T_B > 0.6$ are classed as E, and those in between are classed as S0. Our model produces a slightly higher fraction of spirals than did that of Cole et al. (2000), a consequence of the more detailed calculation of merger times adopted here. This is in slightly better agreement with the observational data, but given the crude way in which morphological types are assigned in the models, these differences should not be over-emphasized.

The cold gas content of $L_\star$ spiral and irregular galaxies considered by Cole et al. (2000) is unchanged in our model, as are the metallicities of gas and stars in these galaxies. However, the metallicity of gas in spirals and irregulars does show a somewhat steeper dependence on luminosity than in Cole et al. (2000), resulting in slightly better agreement with observational data. This difference arises because of the effect of the filtering mass. In haloes only slightly more massive than the filtering mass, there can have been no enrichment of gas in smaller haloes in the merging hierarchy (as these haloes do not accrete gas). The faint central galaxies of these low-mass haloes are therefore accreting relatively metal-poor gas compared to those in the Cole et al. model, resulting in lower gas metallicities. The metallicity of stars in elliptical galaxies is also changed, but in a different way. Bright ellipticals are the same in our model as in that of Cole et al., but faint ones on average have somewhat higher metallicities than in Cole et al., which worsens the agreement with the observational data. Here,

Table 2. The morphological mix of galaxies brighter than $M_B - 5 \log h = -19.5$ from this work and from the model of Cole et al. (2000). Also shown is the morphological mix in the APM Bright Galaxy Catalogue (which is apparent-magnitude-limited) from Loveday (1996).

	S:S0:E
This work ($f_{esc} = 100\%$)	70:05:25
Cole et al. (2000)	61:08:31
Loveday (1996)	67:20:13

the main effect is that, with photoionization switched on, a galaxy of a given luminosity tends to be found in a halo with higher circular velocity (since the filtering mass reduces the amount of gas able to accrete into each halo). The higher circular velocity implies a deeper potential well, which makes the galaxy better at retaining metals (i.e., fewer are lost in the winds associated with supernovae feedback), increasing the effective yield and raising the metallicities of the low-luminosity ellipticals. Pre-processing of gas in lower mass haloes is not so important for the ellipticals, since the gas is processed right up to the effective yield very quickly in the burst of star formation which makes the elliptical.

In conclusion, the largest differences in galaxy properties between our new model and that of Cole et al. (2000) occur for low-luminosity galaxies, the differences beginning to be noticeable at around 1 mag faintwards of $L_\star$. The most important difference is a flattening of the faint end of the galaxy luminosity function. This slope is in reasonable agreement with the latest observational estimates from the 2dFGRS and 2MASS galaxy surveys.

4.2.4 *Properties of galaxies at high redshifts*

Semi-analytic models have been used extensively to investigate the populations of galaxies seen at high redshifts, such as Lyman-break galaxies (Baugh et al. 1998; Governato et al. 1998; Somerville, Primack & Faber 2001). We find that photoionization has very little effect on the properties of Lyman-break galaxies at $z = 3$, for the range of luminosities that is currently observed, because the filtering mass is well below the typical mass halo in which these galaxies reside. We defer a more detailed consideration of high-redshift galaxies to a future paper.

5 DISCUSSION

We have presented a coupled model for evolution of the ionization state and thermal properties of the IGM and the formation of galaxies. The IGM is photoionized by radiation from stars in galaxies and from quasars, and the photoionizing background in turn exerts a negative feedback effect on further galaxy formation. This photoionization feedback operates in two ways, by heating the IGM, and hence by the effects of gas pressure reducing the amount of gas which collapses into haloes, and by ionizing and heating gas within haloes, and hence reducing the amount of gas able to cool to form galaxies. The evolution of the ionizing luminosity of the galaxy population is calculated self-consistently with the effects of this photoionization feedback.

We calculate the formation of galaxies within the CDM model, by adapting the semi-analytical galaxy formation model of Cole et al. (2000), modified to include the photoionization feedback effects described above. This is coupled to a simple model for the evolution of a clumpy IGM, which, given the evolution of the ionizing emissivity of galaxies and quasars as an input, predicts the evolution of the mean ionized fractions of hydrogen and helium, the volume-averaged temperature of the IGM, and the ionizing background. We have tested the IGM model against the results from numerical simulations of the IGM, and find that the predictions for global properties agree reasonably well. In particular, we find that our simple IGM model accurately predicts the evolution of the characteristic halo mass below which accretion of baryonic matter is strongly suppressed, which is the most important quantity in our later study of galaxy formation.

In order to predict more accurately the properties of satellite galaxies within larger dark matter haloes, we have also improved the Cole et al. (2000) semi-analytical model by incorporating a detailed treatment of the dynamics of satellites, including the effects of dynamical friction, tidal stripping and heating by tidal shocks. We have compared this model in the pure dark matter case with the results from high-resolution *N*-body simulations on the amount of substructure in dark haloes, and find good agreement. This improved model predicts merging time-scales for galaxies that on average are comparable to those from the simple estimates used in previous work, although some satellites have their dark haloes heavily stripped by tidal forces, and these have much longer merging time-scales as a consequence of the weaker dynamical friction force resulting from the reduced satellite mass.

A significant uncertain parameter in our photoionization model is f_{esc}, the fraction of ionizing photons from stars able to escape from galaxies. In our model, we need to assume $f_{esc} = 100$ per cent in order to produce the emissivity of ionizing photons at $z = 3$ inferred observationally by Steidel et al. (2001). The model then predicts reionization of H I at $z \approx 8$ and reionization of He II at $z \approx 4$. Each reionization event is preceded by an episode of reheating. However, this model also produces an ionizing background at $z < 4$ that is higher than observational estimates based on the proximity effect. An escape fraction of 10 per cent gives much better agreement with observational data on the ionizing background, and produces a better (compared to recent observations), albeit somewhat too low, redshift of reionization. In any case, the choice of f_{esc} does not change our conclusions about the properties of galaxies at $z = 0$.

Applying our model to the evolution of the galaxy population, we find the following results.

(i) The global star formation rate in our model is suppressed slightly after each episode of reheating due to reionization of H I and He II. The suppression is quite small, with reductions of no more than 25 per cent compared to a model with no reionization. By $z = 0$, the star formation rate has recovered to the level predicted by our model with no reionization, as by then most star formation is occurring in haloes well above the masses and temperatures affected by photoionization feedback.

(ii) Galaxies brighter than $L_\star$ are mostly unaffected by photoionization. Faintwards of $L_\star$, photoionization becomes progressively more important, reducing the abundance of galaxies of given luminosity. Keeping the same prescription for supernovae feedback as used by Cole et al. (2000), we find that including photoionization feedback produces a much better fit to recent determinations of the faint end of the galaxy luminosity function at $z = 0$ (e.g. Madgwick et al. 2001; Cole et al. 2001). Most of the effect is due to the inability of hot IGM gas to accrete into low-mass dark matter haloes, but heating of gas in haloes by the ionizing background and tidal limitation of satellite galaxies also play a role.

(iii) Preliminary analysis of a model with no feedback from supernovae, but including the effects of photoionization indicates that such a model can produce a luminosity function with faint-end slope almost as flat as some observational estimates, and significantly flatter than a model without supernovae feedback or photoionization. Further work is needed to determine if such a model can be made consistent with other observational data.

(iv) Other properties of bright galaxies at the present day (e.g., sizes, Tully–Fisher relation, metallicities) are unaffected by photoionization. For faint galaxies, we find differences in the Tully–Fisher relation and in metallicities that are readily understood.

(v) Photoionization has little effect on the predicted properties of Lyman-break galaxies, over the range of redshifts and luminosities for which they are actually observed. These galaxies at $z = 3$ typically live in haloes significantly more massive than that at which photoionization feedback becomes important, so their properties are insensitive to the reionization history.

If the Universe was reionized through photoionization (and no convincing alternative has been proposed), then the mechanisms inhibiting galaxy formation which we have examined in this paper *must* operate. As such, no model of galaxy formation is complete without their inclusion. Although we have shown that the properties of bright galaxies are almost entirely unaffected, the properties of faint galaxies are strongly influence by photoionization. The methods described in this paper provide a flexible and computationally efficient way to assess the impact of photoionization on galaxy formation, and allow us to make definite predictions for the properties of faint galaxies.

As we have shown, photoionization feedback has the greatest effect on faint galaxies residing in low-mass dark matter haloes. As such, it will undoubtedly have important implications for predictions about the population of satellite galaxies found in the Local Group. In the second paper in this series, we will explore in detail the properties of these galaxies.

ACKNOWLEDGMENTS

REFERENCES

Arnaud M., Rothenflug R., 1985, A&AS, 60, 425
Babul A., Rees M. J., 1992, MNRAS, 255, 346
Bader G., Deuflhard P., 1983, Numerische Mathematik, 41, 373
Barkana R., Loeb A., 1999, ApJ, 523, 54
Batjlik S., Duncan R. C., Ostriker J. P., 1988, ApJ, 327, 570
Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 1998, ApJ, 498, 504
Becker R. H. et al., 2001, AJ, 122, 2850
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000a, MNRAS, 311, 793
Benson A. J., Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 2000b, MNRAS, 316, 107
Benson A. J., Nusser A., Sugiyama N., Lacey C. G., 2001a, MNRAS, 320, 153
Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2001b, MNRAS, 327, 1041
Binney J., Tremaine S., 1987, Galactic Dynamics. Princeton Univ. Press, Princeton
Black J. H., 1981, MNRAS, 197, 553
Blanchard A., Valls-Gabaud D., Mamon G. A., 1992, A&A, 264, 365
Bower R. G., Benson A. J., Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 2001, MNRAS, 325, 497
Bullock J. S., Kravtsov A. V., Weinberg D. H., 2000, ApJ, 539, 517
Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559

Bunker A. J., Marleau F. R., Graham J. R., 1998, AJ, 116, 2086
Cen R., 1992, ApJS, 78, 341
Chiba M., Nath B. B., 1994, ApJ, 436, 618
Chiu W. A., Ostriker J. P., 2000, ApJ, 534, 507
Ciardi B., Ferrara A., Governato F., Jenkins A., 2000, MNRAS, 314, 611
Cole S., 1991, ApJ, 367, 45
Cole S., Lacey C. G., 1996, MNRAS, 281, 716
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Cole S. et al. (The 2dFGRS Team), 2001, MNRAS, 326, 255
Coles P., Jones B., 1991, MNRAS, 248, 1
Cooke A. J., Espey B., Carswell B., 1997, MNRAS, 284, 552
Couchman H. M. P., Rees M. J., 1986, MNRAS, 221, 53
de Jong R. S., Lacey C. G., 2000, ApJ, 545, 781
Dekel A., Silk J., 1986, ApJ, 303, 39
Djorgovski S. G., Castro S. M., Stern D., Mahabal A., 2001, ApJ, 560, 5
Doroshkevich A. G., Zeldovich Ya. B., Novikov I. D., 1967, SvA, 11, 233
Dove J. B., Shull J. M., Ferrara A., 2000, ApJ, 531, 846
Efstathiou G., 1992, MNRAS, 256, 43
Eke V. R., Navarro J. F., Steinmetz M., 2001, ApJ, 554, 114
Evrard A. E., Henry J. P., 1991, ApJ, 383, 95
Fan X. et al. (The SDSS Collaboration), 2000, AJ, 120, 1167
Ferrara A., Bianchi S., Cimatti A., Giovanardi C., 1999, ApJS, 123, 437
Gardner J. P., Sharples R. M., Frenk C. S., Carrasco B. E., 1997, ApJ, 480, L99
Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 1998, MNRAS, 300, 146
Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 2000, ApJ, 544, 616
Gnedin N. Y., 2000a, ApJ, 535, 530
Gnedin N. Y., 2000b, ApJ, 542, 535
Gnedin N. Y., Hui L., 1998, MNRAS, 296, 44
Gnedin N. Y., Ostriker J. P., 1997, ApJ, 486, 581
Goodwin S., Pearce F. R., Thomas P., 2001, in Deiters S., et al., eds, ASP Conf. Ser. Vol. 228, Dynamics of Star Clusters and the Milky Way. Astron. Soc. Pac., San Francisco, p. 440
Governato F., Baugh C. M., Frenk C. S., Cole S., Lacey C. G., Quinn T., Stadel J., 1998, Nat, 392, 359
Haiman Z., Loeb A., 1997, ApJ, 483, 21
Kaiser N., 1991, ApJ, 383, 104
Katz N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Kepner J. V., Babul A., Spergel D. N., 1997, ApJ, 487, 61
Klypin A. A., Kravtsov A. V., Valenzuela O., Prada F., 1999, ApJ, 522, 82
Kuijken K., Gilmore G., 1989, MNRAS, 239, 571
Kulkarni V. P., Fall S. M., 1993, ApJ, 413, L63
Leitherer C., Fergusen H., Heckman T. M., Lowenthal J. D., 1995, ApJ, 454, 19
Loveday J., 1996, MNRAS, 278, 1025
Loveday J., Peterson B. A., Efstathiou G., Maddox S. J., 1992, ApJ, 390, 338
Mac Low M. M., Ferrara A., 1999, ApJ, 513, 142
Madau P., Haardt F., Rees M. J., 1999, ApJ, 514, 648
Madgwick D. S. et al., 2001, MNRAS, submitted (astro-ph/0107197)
Martin C. L., 1999, ApJ, 513, 156
Matthewson D. S., Ford V. L., Buchhorn M., 1992, ApJS, 81, 413
Mayer L., Governato F., Colpi M., Moore B., Quinn T., Wadsley J., Stadel J., Lake G., 2001, ApJ, 559, 754
Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, ApJ, 530, 1
Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, 19
Nagashima M., Gouda N., Sugiura N., 1999, MNRAS, 305, 449
Navarro J. F., Steinmetz M., 1997, ApJ, 478, 13
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Netterfield C. B. et al., 2001, ApJ, submitted (astro-ph/0104460)
Peacock J. A., Dodds S. J., 1996, MNRAS, 280, L19
Peebles P. J. E., 1968, ApJ, 153, 1
Quinn T., Katz N., Efstathiou G., 1996, MNRAS, 278, L49
Ratcliffe A., Shanks T., Parker Q. A., Fong R., 1998, MNRAS, 293, 197

Scholz T. T., Walters H. R. J., 1991, ApJ, 380, 302
Scott J., Bechtold J., Dobrzycki A., Kulkarni V. P., 2000, ApJS, 130, 67
Seager S., Sasselov D. D., Scott D., 2000, ApJS, 128, 407
Shapiro P. R., Giroux M. L., Babul A., 1994, ApJ, 427, 25
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Somerville R. S., Primack J. R., Faber S. M., 2001, MNRAS, 320, 504
Steidel C. C., Pettini M., Adelberger K. L., 2001, ApJ, 546, 665
Stern S., Spinrad H., Eisenhardt P., Bunker A. J., Dawson S., Stanford S. A.,
    Elston R., 2000, ApJ, 533, L75
Sutherland R. S., 1998, MNRAS, 300, 321
Sutherland R., Dopita M., 1993, ApJS, 88, 253
Taylor J. E., Babul A., 2001, ApJ, 559, 716
Thoul A. A., Weinberg D. H., 1996, ApJ, 465, 608
Valageas P., Silk J., 1999, A&A, 350, 725
Vedel H., Hellsten U., Sommer-Larsen J., 1994, MNRAS, 271, 743
Verner D. A., Ferland G. J., 1996, ApJS, 103, 467

Verner D. A., Ferland G. J., Korista K. T., Yakovlev D. G., 1996, ApJ, 465,
    487
Vogel S. N., Weymann R., Rauch M., Hamilton T., 1995, ApJ, 441, 162
Voronov G. S., 1997, Atomic Data and Nuclear Data Tables, 65, 1
Weinberg D. H., Hernquist L., Katz N., 1997, ApJ, 477, 8
Weinberg M. D., 1994, AJ, 108, 1398
Weymann R. J., Vogel S. N., Veilleux S., Epps H. W., 2001, ApJ, 561, 559
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Williger G. M. et al., 1994, ApJ, 428, 574
Wu K. K. S., Fabian A. C., Nulsen P. E. J., 2000, MNRAS, 318, 889
Zeldovich Ya. B., Kurt V. G., Sunyaev R. A., Zh. Eksp. Teor. Fiz., 55,1968
    278, (English trans., Soviet Phys.—JETP, 28, 146 [1969])
Zucca E. et al., 1997, A&A, 326, 477

This paper has been typeset from a T$_{\rm E}$X/L$^{\rm A}$T$_{\rm E}$X file prepared by the author.