How big were the first cosmological objects?

Article (Published Version)

Mon. Not. R. Astron. Soc. **330,** 927–936 (2002)

How big were the first cosmological objects?

Roger M. Hutchings,[1] F. Santoro,[1★] P. A. Thomas[1] and H. M. P. Couchman[2]
[1]*Astronomy Centre, CPES, University of Sussex, Falmer, Brighton BN1 9QJ*
[2]*Department of Physics and Astronomy, McMaster University, Hamilton, Ontario, L8S 4M1, Canada*

Accepted 2001 November 7. Received 2001 November 5; in original form 2001 February 9

ABSTRACT

We calculate the cooling times at constant density for haloes with virial temperatures from 100 K to 1×10^5 K that originate from a 3σ fluctuation of a CDM power spectrum in three different cosmologies. Our intention is to determine the first objects that can cool to low temperatures, but not to follow their dynamical evolution. We identify two generations of haloes: those with low virial temperatures, $T_{\rm vir} \lesssim 9000$ K that remain largely neutral, and those with larger virial temperatures that become ionized. The lower temperature, lower mass haloes are the first to cool to 75 per cent of their virial temperature. The precise temperature and mass of the first objects are dependent upon the molecular hydrogen (H_2) cooling function and the cosmological model. The higher mass haloes collapse later but, in this paradigm, cool much more efficiently once they have done so, first via electronic transitions and then via molecular cooling: in fact, a greater residual ionization once the haloes cool below 9000 K results in an enhanced H_2 production and hence a higher cooling rate at low temperatures than for the lower mass haloes, so that within our constant-density model it is the former that are the first to cool to really low temperatures. We discuss the possible significance of this result in the context of CDM models in which the shallow slope of the initial fluctuation spectrum on small scales leads to a wide range of halo masses (of differing overdensities) collapsing over a small redshift interval. This 'crosstalk' is sufficiently important that both high- and low-mass haloes collapse during the lifetimes of the massive stars which may be formed at these epochs. Further investigation is thus required to determine which generation of haloes plays the dominant role in early structure formation.

Key words: molecular processes – galaxies: formation – cosmology: theory – early Universe.

1 INTRODUCTION

The study of the first generation of objects in the Universe that are able to cool sufficiently to collapse and form luminous objects is a well-defined problem. By definition, there are no stars or other sources of ionizing radiation, and one does not have to consider feedback from supernovae and enrichment of the Universe with metals. The first objects to form arise from the collapse of high-sigma fluctuations in the background density field. These peaks will virialize and begin to cool. Objects with virial temperatures $T \lesssim 9000$ K are cooled by H_2 molecules: the molecules are excited by collisions with neutral hydrogen and then spontaneously de-excite with the emission of a photon. In the absence of metals the dominant cooling mechanisms for $9000 \lesssim T \lesssim 50000$ K are collisional excitation of neutral hydrogen and recombination of ionized hydrogen, and for $T \gtrsim 50000$ K the dominant mechanism is collisional excitation of ionized helium.

The question of what are the masses of the first objects to form in a standard CDM scenario was studied by Tegmark et al. (1997, hereafter T97) in a landmark paper entitled 'How small were the first cosmological objects?' They analytically tracked a top-hat collapse to the point of virialization, at which point the gas was cooled at constant density. They accepted an object as having cooled if it met the criterion $T(0.75z_{\rm vir}) \leq 0.75T_{\rm vir}$, where $T_{\rm vir}$ is the virial temperature, and $z_{\rm vir}$ the virialization redshift. They found that the first generation of objects that cooled in a standard CDM scenario, virialized at a redshift of 27, had a baryonic mass of about $10^5\,{\rm M}_\odot$. In a later paper, Abel et al. (1998) redid the calculation with a different H_2 cooling function, and estimated a very similar virialization redshift but a smaller baryonic mass, $7 \times 10^3\,{\rm M}_\odot$.

In the present paper we adopt a similar approach to T97, but consider the collapse of not only small objects with virial temperatures $\leq$9000 K, but also objects of virial temperatures up to 100 000 K. The motivation for this is that the high virial temperatures will partially ionize the gas. Since the gas will then cool more rapidly than it can recombine, the ionization level at temperatures $\leq$9000 K will be greater than it otherwise would have been if no re-ionization had taken place. This in turn accelerates production of H_2, ultimately resulting in enhanced cooling at lower

★E-mail: p.a.thomas@sussex.ac.uk (PAT); fernando@pact.cpes.susx.ac.uk (FS)

temperatures. A similar effect was noted by MacLow & Shull (1986) and Shapiro & Kang (1987) for the cooling of gas behind intergalactic shocks.

We confirm the earlier results on the masses of the *smallest* haloes that can cool, but show that these are not the *first* objects to do so in this constant-density model. A plot of cooling redshift versus halo mass shows two separate maxima corresponding to haloes with virial temperatures of about 4000 and 11 000 K. Thus there are two distinct generations of primordial haloes, depending upon whether or not their virial temperature is high enough to ionize the gas. Given our simplified model, each of these generations has a unique halo mass that will be able to cool first to low temperatures and form stars. If they virialized at the same time, then the higher mass haloes have the potential for more efficient cooling and, depending upon how the collapse occurs, could win out; however, in CDM cosmologies, the collapse of larger objects occurs slightly later than the collapse of smaller ones at the same overdensity relative to the rms at each scale. The enumeration of the difference in cooling times between the two generations of haloes is the main topic of this paper. We show that this difference is of order 10 Myr, comparable to the lifetimes of massive stars. It is thus possible that the second generation of massive haloes could collapse before feedback from the first stars affects their internal composition. The result of this paper, therefore, is to argue that more detailed modelling is required to establish the relative importance of the two generations for subsequent structure formation.

A second difference between our study and previous ones is that we consider more up-to-date cosmological models. In particular, we choose a density fluctuation spectrum with less power on small scales. In a critical-density universe, this has the effect of delaying collapse until a redshift of 11. Higher redshifts can be recovered by the introduction of a cosmological constant.

We describe our chemical model in Section 2, and our numerical method in Section 3. The results are presented in Section 4, and discussed in Section 5. Finally, we summarize our conclusions in Section 6.

2 GAS CHEMISTRY AND COOLING FUNCTIONS

In this section we present a minimal model of the gas chemistry needed to accurately follow the temperature evolution of the gas. Our model is closely based on those of Abel et al. (1997) and Fuller & Couchman (2000), who both did a thorough search to identify the key reactions, but differs from theirs in the manner in which it handles the production and destruction of H^- and H_2^+. As a consequence, we end up with slightly different terms in our final equation for H_2 production. In principle, this could be quite important; however, it seems to make little difference to the results over the parameter ranges considered in this paper.

Table 1 lists the important reactions for the combinations of temperature and species abundances that we consider. The model is applicable to haloes with temperatures between 100 and 100 000 K, and redshifts up to 50. Note that hydrogen and helium do not interact chemically. The latter is included in order that the ionization level be correctly reproduced at high temperatures, but its omission would make very little difference to the results presented in this paper. We do not include photoionization from cosmic microwave background photons, which is important only at redshifts in excess of 100.

2.1 H_2 production and destruction

The equation for the rate of change of H_2 abundance is derived in the appendices:

$$\frac{\mathrm{d}n_{\mathrm{H_2}}}{\mathrm{d}t} = n_{\mathrm{H}}^2\left(n_{\mathrm{e^-}}\frac{R_3R_4}{R_{\mathrm{H^-}}} + n_{\mathrm{H^+}}\frac{R_6R_7}{R_{\mathrm{H_2^+}}}\right) - n_{\mathrm{H_2}}\left(n_{\mathrm{H^+}}n_{\mathrm{e^-}}\frac{R_9R_8}{R_{\mathrm{H_2^+}}} + n_{\mathrm{H}}R_{10} + n_{\mathrm{e^-}}R_{11}\right), \quad (1)$$

where

$$R_{\mathrm{H^-}} = n_{\mathrm{H}}R_4 + n_{\mathrm{H^+}}R_5, \quad (2)$$

$$R_{\mathrm{H_2^+}} = n_{\mathrm{H}}R_7 + n_{\mathrm{e^-}}R_8, \quad (3)$$

and R_n is the rate of reaction n from Table 1.

We next discuss the regimes under which each of these terms is important.

2.1.1 *Formation of H_2*

The H^- channel is the dominant H_2 formation path. The H_2^+

Table 1. This table summarizes the important reactions needed in order to calculate accurately the abundance of H_2. References are: HTL Haiman, Thoul & Loeb (1996); GP Galli & Palla (1998); FC Fuller & Couchman (2000); SLD Stancil, Lepp & Dalgarno (1998); AAZN Abel et al. (1997).

	Reaction	Rate/cm^3 s^{-1}	Reference
1	$H + e^- \mapsto H^+ + 2e^-$	$5.9\times10^{-11}T_0^{0.5}(1+T_5^{0.5})^{-1}\exp(-1.58/T_5)$	HTL
2	$H^+ + e^- \mapsto H + \gamma$	$3.3\times10^{-10}T_0^{-0.7}(1+T_6^{0.7})^{-1}$	HTL
3	$H + e^- \mapsto H^- + \gamma$	$1.4\times10^{-18}T_0^{0.93}\exp(-T_4/1.62)$	GP
4	$H^- + H \mapsto H_2 + e^-$	1.3×10^{-9}	FC
5	$H^- + H^+ \mapsto 2H$	$4.0\times10^{-6}T_0^{-0.5}$	FC
6	$H + H^+ \mapsto H_2^+ + \gamma$	$2.1\times10^{-23}T_0^{1.8}\exp(-2/T_1)$	SLD
7	$H_2^+ + H \mapsto H_2 + H^+$	6.4×10^{-10}	GP
8	$H_2^+ + e^- \mapsto 2H$	$1.2\times10^{-7}T_0^{-0.4}$	SLD
9	$H_2 + H^+ \mapsto H_2^+ + H$	$\min[3.0\times10^{-10}\exp(-2.11/T_4),\ 1.5\times10^{-10}\exp(-1.40/T_4)]$	GP
10	$H_2 + H \mapsto 3H$	$7.1\times10^{-19}T_0^{2.01}(1+2.13T_5)^{-3.51}\exp(-5.18/T_4)$	AAZN
11	$H_2 + e^- \mapsto 2H + e^-$	$4.4\times10^{-10}T_0^{0.35}\exp(-1.02/T_5)$	GP
12	$He + e^- \mapsto He^+ + 2e^-$	$2.4\times10^{-11}T_0^{0.5}(1+T_5^{0.5})^{-1}\exp(-2.85/T_5)$	HTL
13	$He^+ + e^- \mapsto He + \gamma$	$1.5\times10^{-10}T_0^{-0.64} + 1.9\times10^{-3}T_0^{-1.5}\exp(-5.64/T_5)(0.3+\exp(9.40/T_4)$	AAZN
14	$He^+ + e^- \mapsto He^{++} + 2e^-$	$5.7\times10^{-12}T_0^{0.5}(1+T_5^{0.5})^{-1}\exp(-6.32/T_5)$	HTL
15	$He^{++} + e^- \mapsto He^+ + \gamma$	$1.3\times10^{-9}T_0^{-0.7}(1+T_6^{0.7})^{-1}$	HTL

channel is important only at high redshifts ($z \gtrsim 200$ when the H^- channel is suppressed by photodestruction of H^-, which we have omitted here) or high temperatures. However, at high temperatures the H_2 is rapidly destroyed and, for all the models that we consider in this paper, omitting the H_2^+ formation channel makes no difference to the final abundance of H_2.

2.1.2 *Destruction of H_2*

It is the process of H_2 destruction where the present work differs slightly from that of Abel et al. (1997) and Fuller & Couchman (2000). In these two papers the term for destruction by H^+ is given as

$$n_{H_2} n_{H^+} R_9, \tag{4}$$

whereas we have

$$n_{H_2} n_{H^+} R_9 \frac{n_{e^-} R_8}{R_{H_2^+}} = n_{H_2} n_{H^+} R_9 \left(1 - \frac{n_H R_7}{R_{H_2^+}}\right). \tag{5}$$

This takes into account the fact that, at low ionization levels, much of the H_2^+ that is produced will be immediately converted back to H_2, so that there will be no net destruction. In some of the models that we consider, this can make transitory differences of a factor of 10 or more in the H_2 abundance. However, as H_2 production takes over from destruction at low temperatures, the effect on the final H_2 abundance is quite small (at most a few per cent).

With the correct rate for the H_2^+ destruction channel, there is a small parameter range for which H10 becomes the dominant destruction process. However, this is so fleeting that it makes a negligible difference to the results and can safely be omitted.

We note that the previous papers consider only low temperatures, $T < 6000$ K, for which H_2 destruction is relatively unimportant compared to its production, and the error in their results is negligible. Nevertheless, a minimal model that includes H_2 destruction should use our equation 1.

2.2 Ionization level

In principle, the equation for the rate of change of H^+ abundance is every bit as complicated as that for H_2. However, it turns out in practice that there are only two important terms:

$$\frac{dn_{H^+}}{dt} = n_{e^-} n_H R_1 - n_{e^-} n_{H^+} R_2. \tag{6}$$

Similarly, for the helium species,

$$\frac{dn_{He}}{dt} = n_{e^-} n_{He^+} R_{13} - n_{e^-} n_{He} R_{12}, \tag{7}$$

$$\frac{dn_{He^{++}}}{dt} = n_{e^-} n_{He^+} R_{14} - n_{e^-} n_{He^{++}} R_{15}. \tag{8}$$

The number density of electrons is given by charge conservation:

$$n_{e^-} = n_{H^+} + n_{He^+} + 2n_{He^{++}}. \tag{9}$$

2.3 Cooling terms

At low temperatures the main coolant is molecular hydrogen. We use the cooling rate given in Galli & Palla (1998, hereafter GP98) as summarized by Fuller & Couchman (2000). This gives a fit to the low-density limit of the calculations of Martin, Schwarz & Mandy (1996) and Forrey et al. (1997), which together cover a wide temperature range:

$$\log_{10}\left(\frac{\Lambda_{H_2}(T)}{n_H n_{H_2}\,\mathrm{erg\,cm^3\,s^{-1}}}\right) = -103.0 + 97.59T_{\log} - 48.05T_{\log}^2 + 10.80T_{\log}^3 - 0.9032T_{\log}^4, \tag{10}$$

where $T_{\log} = \log_{10}(T/\mathrm{K})$.

For all other cooling processes, we use the rates given in Haiman et al. (1996). The most important of these are as follows. At temperatures $T \gtrsim 10000K$, collisional excitation and (less importantly) ionization of atomic hydrogen take over from molecular hydrogen cooling:

$$\frac{\Lambda_{H,ce}(T)}{n_{e^-} n_H\,\mathrm{erg\,cm^3\,s^{-1}}} = 7.50 \times 10^{-19} \frac{1}{1 + T_5^{\frac{1}{2}}} e^{-\frac{1.183}{T_5}} \tag{11}$$

$$\frac{\Lambda_{H,ci}(T)}{n_{e^-} n_H\,\mathrm{erg\,cm^3\,s^{-1}}} = 4.02 \times 10^{-19} \frac{T_5^{\frac{1}{2}}}{1 + T_5^{\frac{1}{2}}} e^{-\frac{1.578}{T_5}} \tag{12}$$

where T_n is the temperature in units of 10^n K.

At even higher temperatures, $50000 \lesssim T \lesssim 100000$, collisional excitation of He^+ is the dominant coolant:

$$\frac{\Lambda_{He^+,ce}(T)}{n_{e^-} n_{He^+}\,\mathrm{erg\,cm^3\,s^{-1}}} = 5.54 \times 10^{-17} \frac{T_0^{-0.397}}{1 + T_5^{\frac{1}{2}}} e^{-\frac{4.737}{T_5}}. \tag{13}$$

We include other processes from Haiman et al. (1996) in our model, including inverse Compton cooling from cosmic microwave background photons, but at the modest redshifts ($z < 50$) and high densities of the collapsed haloes that we consider in this paper none contribute at more than the few per cent level.

3 NUMERICAL PROCEDURE

3.1 General approach

The approach that we have adopted for this paper is to specify an initial temperature, ionization level and H_2 fraction, and to follow their evolution at constant density, using the relevant cooling terms and reaction rates discussed in the previous section. The assumption of constant density is valid only for haloes whose cooling times are much longer than their dynamical times. It is adequate for the purposes of this paper to assess if a given halo at constant density can begin to cool efficiently, but not to follow the collapse of haloes once they do cool.

The equations are integrated using the RK4 integrator from Press et al. (1992), modified to use an adaptive time-step that allowed neither the abundances nor the temperature to vary by more than 0.1 per cent during a time-step. We also tried a Bulirsch–Stoer integrator, which gave identical results for a slightly poorer performance. For tests, we integrated simplified networks of equations for which a solution can be obtained analytically: for example, equations (15) and (16) of T97.

The initial halo parameters are picked to represent the conditions of a virialized object which has collapsed from a 3σ peak in a CDM scenario, with virial temperatures, T_{vir}, in the range 100 to 100 000 K. We consider two different measures of cooling. First, we measure how long it takes haloes to cool to $T_{0.75} = 0.75T_{vir}$; T97 define an object to have cooled if its temperature decreases by 25 per cent or more in the time that redshift does likewise. We find two distinct populations of clouds: low-mass ones that cool via molecular hydrogen, and high-mass ones that cool via electronic

Table 2. Cosmological parameters for the three models: model name; density parameter; cosmological constant in units of $\lambda_0 = \Lambda/3H_0^2$; Hubble parameter in units of $h = H_0/100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$; root-mean-square dispersion of the density within spheres of radius $8\,h^{-1}\,\mathrm{Mpc}$; residual ionization fraction.

Name	Ω_0	λ_0	Ω_{b0}	h	Γ	σ_8	$n_{e,res}/n_{H,tot}$
SCDM	1.0	0.0	0.076	0.5	0.43	0.60	1.58×10^{-4}
τCDM	1.0	0.0	0.184	0.5	0.21	0.60	6.52×10^{-5}
ΛCDM	0.35	0.65	0.038	0.7	0.21	0.90	1.33×10^{-4}

transitions. Secondly, we look at the amount by which an object can cool in one dynamical time. Clouds with virial temperatures greater than 9000 K (i.e., those that have re-ionized) form later (at the same σ) but cool more effectively than lower mass clouds, and have a much greater H_2 fraction once they have cooled to low temperatures.

3.2 Cosmological models

We compute the collapse redshift of objects arising from 3σ peaks of a CDM power spectrum. The spectrum was calculated using a real-space top-hat window function, the transfer function of Bond & Efstathiou (1984) for scales above $1\,h^{-1}\,\mathrm{Mpc}$ and the transfer function of Bardeen et al. (1986, hereafter BBKS), for smaller scales. We choose this combination, as the Bond & Efstathiou transfer function is more accurate than BBKS, but makes no attempt to calculate the function accurately on scales below $1\,h^{-1}\,\mathrm{Mpc}$.

We present results for three different cosmological models, as listed in Table 2. For comparison with previous work, we first choose the standard cold dark matter cosmology (SCDM) with power spectrum shape parameter, $\Gamma = 0.43$. When normalized to the *COBE* results, this power spectrum is now known to have too much power on small scales, and so the other two models that we consider use a smaller value, $\Gamma = 0.21$.

The ΛCDM model is our best guess at the most favourable CDM model. The combination of Ω_0, Ω_{b0} and h naturally produces a power spectrum with the correct value of Γ, and the Hubble parameter and baryon fraction both lie close to currently preferred values (see, for example, Freedman et al. 2001 for the former, and Ettori & Fabian 1999 for the latter). The normalization of the power spectrum, σ_8, is chosen to reproduce the correct local abundance of rich clusters (Viana & Liddle 1996).

The τCDM model is an attempt to salvage a critical-density model in the post-*COBE* epoch. It has some motivation in decaying neutrino scenarios (e.g. Hansen & Villante 2000), but is mostly phenomenological in nature. For this reason, we have fixed Γ to have the same value as in the ΛCDM model, but have used a lower value of σ_8 to again reproduce the rich cluster abundance. In addition, we have had to take a lower value of the Hubble parameter in order to overcome the age problem, and a baryon density considerably higher than that predicted by primordial nucleosynthesis to give the correct baryon fraction in clusters.

3.3 Initial halo properties

We calculate the initial properties of haloes by assuming that they have settled down into virial equilibrium. This will be a good approximation for haloes whose cooling time substantially exceeds their dynamical time, and should therefore enable us to identify objects that are able to cool with reasonable accuracy. To correctly model the properties of haloes whose cooling time is shorter than their dynamical time would require a more sophisticated model.

In an isothermal sphere, the density is strongly peaked towards the centre of the halo. Here we follow the evolution of gas which has a density fixed at the mean value within the virialised halo,

$$\rho_{vir} = \left(\frac{\Delta_c}{\Omega}\right)\rho_{b0}(1+z_{vir})^3, \tag{14}$$

where z_{vir} is the virialization redshift; ρ_{b0} is the current mean density of baryons in the Universe, and Δ_c is the mean overdensity relative to the critical density within the virialized halo, which we take to be $18\pi^2$ for the SCDM and τCDM cosmologies, and $18\pi^2\Omega^{0.45}$ for the ΛCDM cosmology, where Ω is the density parameter at the time of virialization (Eke, Navarro & Frenk 1998). We take a hydrogen mass fraction (in all its forms: neutral, ionized and molecular) of $X = 0.755$. Thus

$$n_{H,tot} \equiv n_H + n_{H^+} + 2n_{H_2} = \rho_{vir}X. \tag{15}$$

Similarly,

$$n_{He,tot} \equiv n_{He} + n_{He^+} + n_{He^{++}} = \rho_{vir}Y, \tag{16}$$

where $Y = (1 - X) = 0.245$ is the helium mass fraction.

For the τCDM and ΛCDM cosmologies, the residual ionization level from the early Universe is taken from Peebles (1968, 1993), but divided by 2 as an attempt to compensate for the neglect of stimulated recombination as suggested by the results of Grachev & Dubrovich (1991) and Sasaki & Takahara (1993). Although dividing by 2 is somewhat arbitrary, it does result in a more accurate estimate of the initial ionization level than many previous papers without explicitly solving the time-dependent evolution of the ionization level with the inclusion of stimulated recombination. We take

$$\frac{n_{e,res}}{n_{H,tot}} = 6\times10^{-6}\Omega_0^{\frac{1}{2}}(\Omega_{b0}h)^{-1}, \tag{17}$$

which gives the residual ionization levels listed in Table 2.

For haloes hotter than about 8000–9000 K, the gas will be ionized above the residual value by the process of virialization. For these we use the equilibrium values determined by solving equations (6), (7) and (8) for $dn_{H^+}/dt = dn_{He}/dt = dn_{He^{++}}/dt = 0$:

$$\frac{n_{H^+}}{n_{H,tot}} = \frac{R_1}{R_1 + R_2}, \tag{18}$$

$$\frac{n_{He^+}}{n_{He,tot}} = \frac{R_{12}R_{15}}{R_{13}R_{15} + R_{12}R_{14} + R_{12}R_{15}}, \tag{19}$$

and

$$\frac{n_{He^{++}}}{n_{He,tot}} = \frac{R_{12}R_{14}}{R_{13}R_{15} + R_{12}R_{14} + R_{12}R_{15}}. \tag{20}$$

The background level of H_2 in the Universe is taken to be $f_{H_{2,res}} = 1.1\times10^{-6}$ as calculated by GP98.

We take the initial temperature to be

$$T_{vir} = \frac{\mu m_H}{k_B}\frac{GM_{tot}}{2r_{vir}}$$

$$\approx 40.8\,\frac{\mu}{1.225}(1+z_{vir})\left(\frac{\Delta_c h^2}{18\pi^2\Omega_0}\right)^{\frac{1}{3}}\left(\frac{M_{tot}}{10^5\,M_\odot}\right)^{\frac{2}{3}}\mathrm{K}. \tag{21}$$

Here $M_{\rm tot}$ is the total mass (dark plus baryonic), which we assume to be distributed as an isothermal sphere within the virial radius, $r_{\rm vir}$; $m_{\rm H}$ is the mass of a hydrogen atom; $k_{\rm B}$ is the Boltzmann constant; G is the gravitational constant; and μ is the mean mass of particles in units of $m_{\rm H}$:

$$\mu = \frac{n_{\rm H,tot} + 4n_{\rm He,tot}}{n_{\rm H,tot} - n_{\rm H_2} + n_{\rm He,tot} + n_{\rm e^-}}. \quad (22)$$

When the molecular hydrogen abundance is low, then μ varies between 1.23 at low temperatures, $T \lesssim 10\,000$ K, and 0.59 at very high temperatures, $T \gtrsim 100\,000$ K.

We shall use the halo virialization redshift, $z_{\rm vir}$, as our ordinate in most of the plots that follow. This can be converted to temperature or mass in a cosmology-dependent way. We plot these relations in Fig. 1, and list examples of the conversion between temperature and other quantities in Table 3. The kink in the temperature profiles corresponds to the changing value of μ from equation (22) during hydrogen ionization (there is a second kink due to ionization of helium, but that is barely discernible). Note that haloes of a given temperature have similar total masses in each cosmology, although they may virialize at very different redshifts.

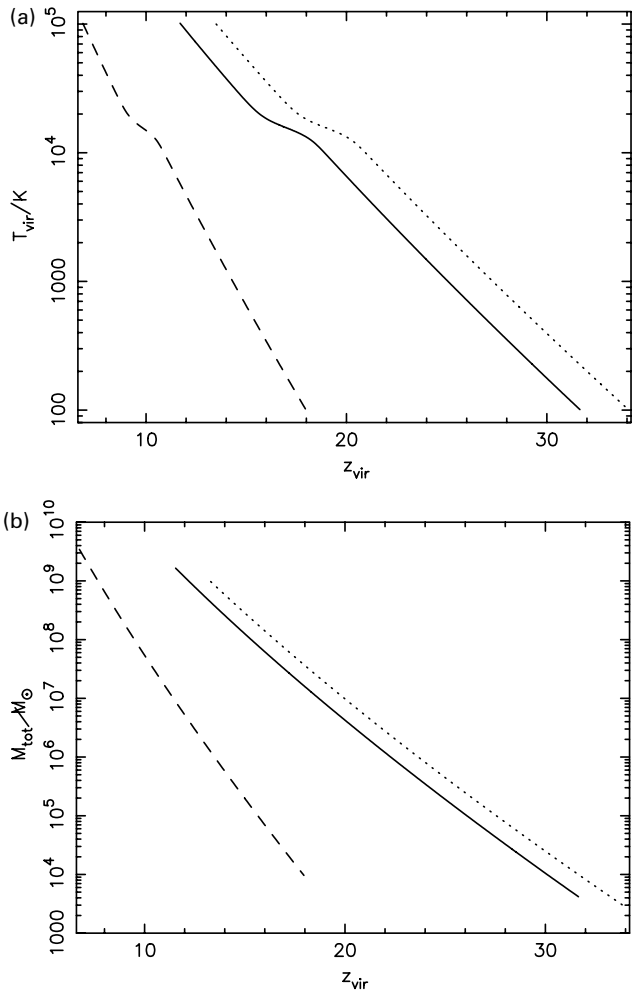

Figure 1. This figure shows the relation between virialization redshift and (a) temperature and (b) total mass for the three cosmologies: SCDM (solid line), τCDM (dashed line), and ΛCDM (dotted line).

Table 3. Conversion between virialization temperature, redshift and mass for the three models: temperature; model; virialization redshift; total mass; baryonic mass.

T/K	Model	$z_{\rm vir}$	$M_{\rm tot}/{\rm M}_\odot$	$M_{\rm bary}/{\rm M}_\odot$
100	SCDM	31.6	4.2×10^3	3.2×10^2
	τCDM	18.0	9.2×10^3	1.7×10^3
	ΛCDM	34.0	2.7×10^3	2.9×10^2
1000	SCDM	25.0	1.8×10^5	1.4×10^4
	τCDM	14.3	4.1×10^5	7.5×10^4
	ΛCDM	27.3	1.1×10^5	1.2×10^4
10 000	SCDM	18.9	8.7×10^6	6.6×10^5
	τCDM	10.9	1.9×10^7	3.5×10^6
	ΛCDM	21.0	5.4×10^6	5.9×10^5
100 000	SCDM	11.6	1.6×10^9	1.2×10^8
	τCDM	6.8	3.3×10^9	6.1×10^8
	ΛCDM	13.3	9.7×10^8	1.1×10^8

4 RESULTS

4.1 SCDM

To compare with previous work, we first consider the collapse of haloes in the SCDM cosmology. We do this for two different forms of the H_2 cooling function. The first, described above in equation (10), is from GP98; the second is from Lepp & Shull (1984, hereafter LS84). This latter form is the one used by Abel et al. (1998); it gives an order of magnitude more cooling at temperatures below 1000 K, and so favours the collapse of low-mass, low-virial temperature objects.

Initially the gas within the haloes is assumed to be shocked to the virial temperature and to be pressure-supported. If it can cool significantly within one dynamical time, then it will contract towards the centre of the halo in order to maintain pressure support. Accordingly, we plot in Fig. 2 the redshift, $z_{0.75}$, at which the gas (if it maintains constant density) will lose 25 per cent of its initial energy. The solid curve was generated using the GP98 cooling function, and the dashed curve using that of LS84. The lower

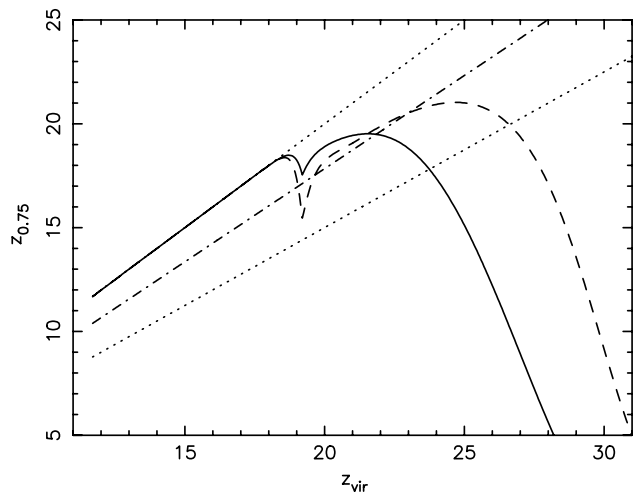

Figure 2. Redshift at which a gas halo will cool to 75 per cent of its initial temperature, $z_{0.75}$, versus virialization redshift, $z_{\rm vir}$, for cosmology SCDM for two different cooling functions: dashed curve, LS84; solid curve, GP98. The upper and lower dotted lines show the relations $z_{0.75} = z_{\rm vir}$ and $z_{0.75} = 0.75 z_{\rm vir}$, respectively; the dot-dashed line shows the redshift at one dynamical time after virialization.

dotted line shows the relation $z_{0.75} = 0.75 z_{\rm vir}$ which is the condition used by T97 to separate clouds which can cool from those which cannot. The dashed line shows the redshift at one dynamical time, $t_{\rm dyn}$, after the time of virialization, $t_{\rm vir}$, where we set $t_{\rm dyn} = t_{\rm vir}/4\sqrt{2}$.

First, concentrate on the GP98 curve. This intersects the line $z_{0.75} = 0.75 z_{\rm vir}$ at $z_{\rm vir} \approx 23.8$, which means that the smallest clouds that can cool according to the T97 criterion will have virial temperatures of $T_{\rm vir} = 1600\,{\rm K}$ and masses of $4 \times 10^5\,{\rm M}_\odot$. The most important quantity for comparison with the results of T97 is the temperature which agrees very well with that given in their fig. 5. For a virialization redshift of 23.8, they require a slightly higher temperature of 2500 K for collapse, but this difference is attributable to our having a slightly higher baryon density and H_2 cooling rate.

T97 assume, within the constraints of the constant-density model, that the smallest haloes that can cool will be the first to do so, but this is not the case. As one moves to lower $z_{\rm vir}$ and higher $T_{\rm vir}$, the efficiency of H_2 production increases and the cooling time decreases. Consequently, the first objects to cool to 75 per cent of their initial temperature virialize later at $z_{\rm vir} \approx 21.6$, and have higher virial temperatures (3700 K) and masses ($1.6 \times 10^6\,{\rm M}_\odot$). This maximum in $z_{0.75}$ corresponds roughly to the redshift at one dynamical time after virialisation.

At higher temperatures the efficiency of H_2 cooling continues to increase but cannot compensate for the later virialization redshift, and so $z_{0.75}$ decreases once more. Note the cusp in the curve at $z_{\rm vir} \approx 19.2$, corresponding to $T_{\rm vir} \approx 9000\,{\rm K}$. As this temperature is approached, H_2 starts to be destroyed, and so its abundance plummets and the cooling time increases. However, at higher temperatures collisional excitation and ionization take over, and these are much more efficient cooling processes, so the cooling time drops once more. In fact, for temperatures above 11 000 K the cooling is essentially instantaneous (i.e., the cooling time much less than the dynamical time).

Thus the two peaks in Fig. 2 correspond to two different classes of object: the first to collapse, at $z_{0.75} = 19.4$, are dominated by molecular cooling, and the second, at $z_{0.75} = 18.5$ by electronic transitions. As we are interested in the first objects to cool, it might be thought that the smaller haloes are the more important, but this is not entirely clear because the T97 criterion does not trace the cooling down to very low temperatures. In fact, the more massive haloes have a large residual ionization, which results in a greater H_2 production once they have cooled below 9000 K, and so they may be the first objects to cool to really low temperatures. We will discuss this further in Section 4.2.

Turning now to the LS84 curve, we see that this is qualitatively similar. However, it predicts shorter cooling times for low-mass clouds, because the H_2 cooling rate is much higher. Hence the smallest objects that can collapse have a lower mass and temperature than for the GP98 cooling function. The lowest mass objects that satisfy the T97 criterion virialize at a redshift of 26.6 and have virial temperatures of 570 K – this seems to agree reasonably well with the predictions from fig. 12 of Abel et al. (1998). The peak of the curve has moved to $z_{\rm vir} \approx 24.7$, and the temperature and mass of these first objects are significantly lower, 1100 K and $2.3 \times 10^5\,{\rm M}_\odot$, respectively. By contrast, the properties of the haloes corresponding to the higher mass peak are only slightly modified.

We use the more up-to-date GP98 cooling function throughout the rest of this paper.

4.2 τCDM

We now turn to the τCDM cosmology. Like the SCDM cosmology, this has a critical density of matter and a high baryon density, but it has a spectral shape that gives less power on small scales, and hence the first objects collapse at much lower redshift. This is shown in Fig. 3, which is the analogue of Fig. 2, but for the τCDM cosmology. We can see that gas haloes first manage to cool to $0.75\,T_{\rm vir}$ only at redshift $z_{0.75} = 10.8$: these have virial temperatures of 4500 K and total masses of $5.0 \times 10^6\,{\rm M}_\odot$.

In order to form stars, the gas has to cool to much less than $T_{0.75}$. Therefore we consider a second measure of cooling: Fig. 4 shows the final temperature of the halo after one dynamical time (once again assuming constant density – this will be a good assumption only for clouds which cool by a small amount in a dynamical time; for others it will underestimate their cooling rate). The coldest

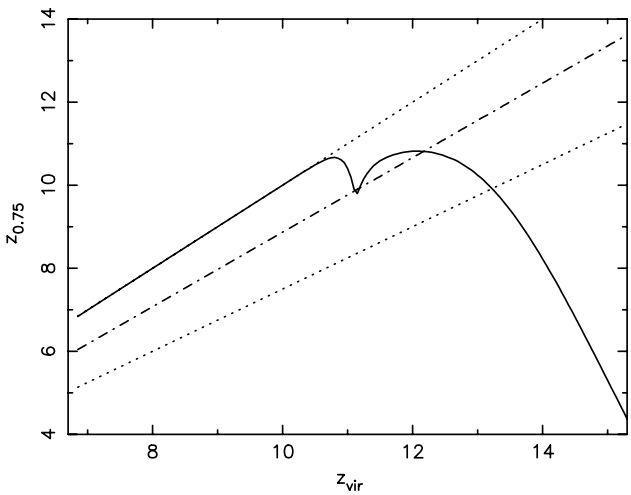

Figure 3. The solid line shows the redshift at which a gas halo will cool to 75 per cent of its initial temperature, $z_{0.75}$, versus virialization redshift, $z_{\rm vir}$, for cosmology τCDM using the GP98 cooling function. The upper and lower dotted lines show the relations $z_{0.75} = z_{\rm vir}$ and $z_{0.75} = 0.75 z_{\rm vir}$, respectively; the dot-dashed line shows the redshift at one dynamical time after virialization.

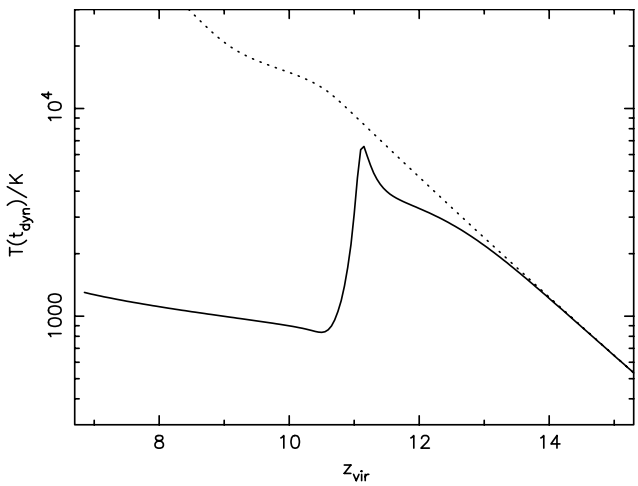

Figure 4. Temperature obtained after one dynamical time versus virialization redshift, for cosmology τCDM. The dotted line shows the initial, virial temperature of the halo. The cusp in the curve corresponds to a virial temperature of 9000 K.

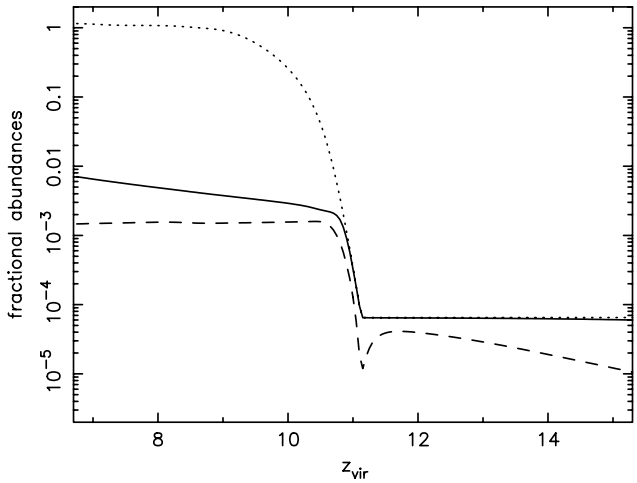

Figure 5. Ionization fraction (solid line) and H_2 abundance (dashed line) after one dynamical time, for cosmology τCDM. The dotted line shows the initial ionization level of the halo.

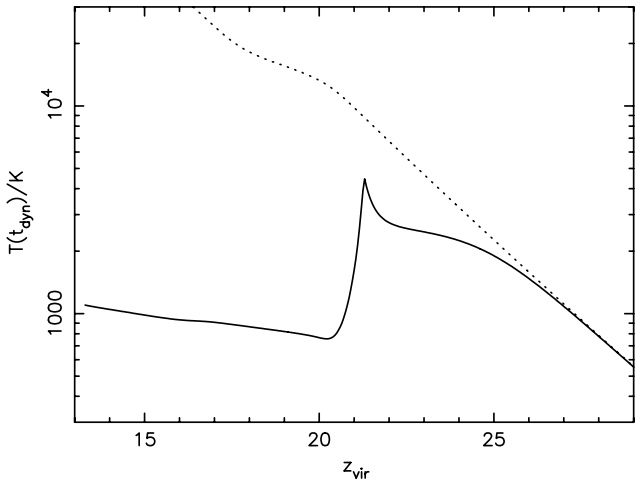

Figure 7. Temperature obtained after one dynamical time versus virialization redshift, for cosmology ΛCDM. The dotted line shows the initial, virial temperature of the halo.

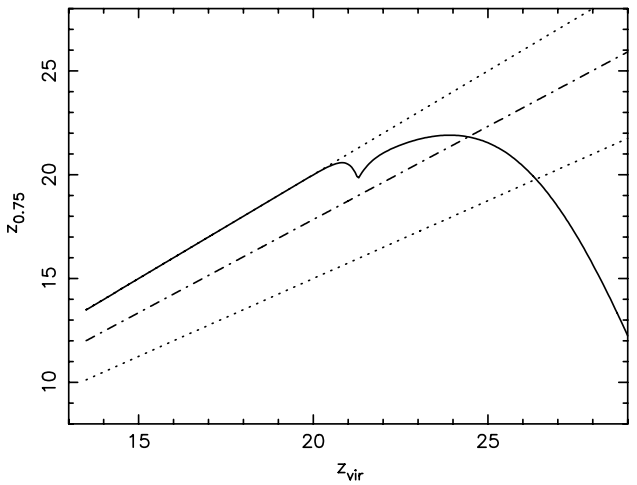

Figure 6. The solid line shows the redshift at which a gas halo will cool to 75 per cent of its initial temperature, $z_{0.75}$, versus virialization redshift, z_{vir}, for cosmology ΛCDM using the GP98 cooling function. The upper and lower dotted lines show the relations $z_{0.75} = z_{vir}$ and $z_{0.75} = 0.75 z_{vir}$, respectively; the dot-dashed line shows the redshift at one dynamical time after virialization.

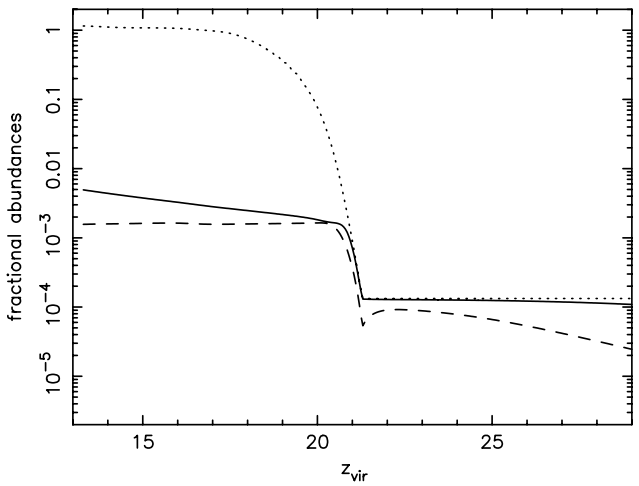

Figure 8. Ionization fraction (solid line) and H_2 abundance (dashed line) after one dynamical time, for cosmology ΛCDM. The dotted line shows the initial ionization level of the halo.

clouds are those to the right-hand side of the figure, but this is only because they were born with low temperatures: they have cooled very little. The final temperature is an increasing function of virial temperature up to about 9000 K when it suddenly plummets, so that, somewhat surprisingly, the larger, higher virial temperature haloes have cooled to a lower temperature than the smaller ones. The reason for this is shown in Fig. 5 which shows that the H_2 abundance after one dynamical time is larger in high-mass haloes than low-mass ones. This is because they become highly ionized, and their residual ionization, once they have cooled back down below 9000 K, is greater than that of low-mass haloes, which in turn results in a greater production of H_2. Note that the cooling rate of these high-mass haloes will continue to be greater than that of low-mass haloes even after one dynamical time. It is therefore conceivable that they will be the first to cool to really low temperatures. Because the gas is free to move around on this time-scale, however, realistic dynamical simulations are required to model the problem accurately.

4.3 ΛCDM

This cosmology is currently the most popular CDM model. Haloes virialize at a higher redshift than in τCDM, but qualitatively their behaviour is just the same. Figs 6–8 mimic those of Figs 3–5, and show all the same features. In this cosmology the first haloes to cool to $T_{0.75}$ have virial temperatures of 3400 K, masses of 8.9×10^5 $M_\odot$, and cool to 75 per cent of their virial temperature at $z_{0.75} \approx 21.8$; the second generation of higher mass haloes cool at $z \approx 20.6$, and have virial temperatures and masses of 10 600 K and 5.9×10^6 $M_\odot$.

5 DISCUSSION

The properties of the haloes corresponding to the two peaks in Figs 2, 3 and 6 are summarized in Table 4.

Note that although the redshift at which the first objects cool is quite different in different cosmologies, their virial temperatures are very similar, reflecting the similar chemical processes that are occurring within them.

Table 4. Properties of the two generations of haloes to form in each of the cosmologies (corresponding to the molecular hydrogen and electron-cooling peaks of the curves in Figs 2, 3 and 6): cosmological model; redshift at which the halo cools to 75 per cent of the virial temperature; virial temperature; total mass of the halo; baryonic mass of the halo.

Model	$z_{0.75}$	$T_{\rm vir}$/K	$M_{\rm tot}/{\rm M}_\odot$	$M_{\rm bary}/{\rm M}_\odot$
Generation 1				
SCDM	19.4	3700	1.6×10^6	1.2×10^5
τCDM	10.8	4500	5.0×10^6	9.1×10^5
ΛCDM	21.8	3400	8.9×10^5	9.7×10^4
Generation 2				
SCDM	18.5	10 800	9.9×10^6	7.6×10^5
τCDM	10.7	10 600	2.1×10^7	3.8×10^6
ΛCDM	20.6	10 600	5.9×10^6	6.4×10^5

5.1 The effect of cosmology

Although it is often not stated explicitly, previous work seems to concentrate on the SCDM cosmology, which is now known to be a poor model of the Universe, the currently favoured model being ΛCDM. Coincidentally, these two give similar formation redshifts for the two generations of haloes and have similar baryonic masses. The phenomenologically based critical-density τCDM cosmology, on the other hand, forms its first objects at much lower redshifts, but nevertheless still comfortably above the observed redshifts of the most distant objects yet observed.

The baryonic masses of these first cooled haloes are about 10 times larger in the τCDM cosmology than in the ΛCDM cosmology – in the former they correspond roughly to the mass of globular clusters. Note, however, that the peaks corresponding to the first generation haloes in Figs 2, 3 and 6 are very broad, so that we would expect haloes with a wide range of masses to cool almost simultaneously.

5.2 The H_2 mass fraction required for collapse

We have shown in Section 4.1 that the virial temperature required for collapse agrees with the results from T97 and Abel et al. (1998). These two papers both claim that the H_2 fraction required for collapse of the first haloes is 5×10^{-4}, a result confirmed by Fuller & Couchman (2000). However, the H_2 fractions shown in Figs 4 and 7 are much lower, about 1×10^{-4}. There is no contradiction here, however. The figures show the H_2 fraction after one dynamical time, whereas T97 give the asymptotic H_2 fraction at late times – for which we get a similar value of 4.4×10^{-4} for the SCDM cosmology (similarly, 4.2×10^{-4} for ΛCDM and a slightly lower value of 3.2×10^{-4} for τCDM).

5.3 The relative roles of the two generations of haloes

When talking about the first cosmological objects, we mean all those haloes that can collapse and form stars before being polluted by feedback of metals (and energy) from supernovae. Table 5 lists the differences in collapse times for the two generations of haloes in each cosmology; these are similar to the lifetime of the most massive stars (a 15-M star has a lifetime of 10 Myr).

The time differences in Table 5 are likely to be upper limits, because the enhanced H_2 production will lead to more rapid cooling at low temperatures for the second-generation haloes. The question arises as to whether, even if small objects are the first to collapse, larger objects may collapse around them and form stars before the first generation of supernovae explode.

Table 5. The redshift, $z_{0.75}$ and age, $t_{0.75}$, of the Universe at the time at which the first haloes in each generation have cooled to 75 per cent of the virial temperature. The final column gives the difference in age between the two generations.

Model	$z_{0.75}$	$t_{0.75}$/Myr	$z_{0.75}$	$t_{0.75}$/Myr	Δt/Myr
	Generation 1		Generation 2		
SCDM	19.4	140	18.5	151	10
τCDM	10.8	322	10.7	326	4
ΛCDM	21.8	145	20.6	157	12

The results presented in this paper are for 3σ density fluctuations. In reality, there will be fluctuations with a continuous range of overdensities, leading to an enormous spread in collapse times. Note also that for the masses we are dealing with, the power spectrum is almost flat – in this situation Press-Schechter theory (Press & Schechter 1974) predicts that there is as much collapsed mass in each decade of mass as any other.

We have taken a simple model in which haloes have no substructure. In reality, high-mass object will contain larger fluctuations on smaller scales. For example, the inclusion of density profiles may raise the core density of even small haloes to a point where free-fall collapse could proceed. In this case, we should also consider the effect of ionizing radiation from high-mass stars; however, the amount of ionizing radiation and the degree to which this permeates the halo will be highly dependent upon the initial mass function and the physical conditions in the star-forming region.

The halo-in-halo problem can be addressed by using a statistical model of the collapse of haloes, but a realistic estimate of the star formation time and the structure of star-forming regions requires numerical simulations of the dynamics of the collapse, as discussed in the next section.

5.4 Towards more realistic models

Our model does not attempt to follow the increasing density in objects that do cool on less than a dynamical time. To do so requires the use of some form of hydrodynamics code. Several groups have attempted this. Haiman et al. (1996), Omukai & Nishi (1998) and Nakamura & Umemura (1999) have all simulated spherically symmetric collapses; Fuller & Couchman (2000) go further and simulate a top-hat collapse with and without substructure using a three-dimensional code. In general, these results confirm the simple analytic predictions.

One of the most important simplifications of our model is that it ignores substructure. Any halo that can cool and collapse on a dynamical time will contain smaller fluctuations of even higher overdensity. This will result in fragmentation of the cloud as the collapse proceeds. This has been investigated by Abel et al. (1998) and Abel, Bryan & Norman (2000), who have used a hierarchical grid code to perform very high-resolution simulations of the collapse and fragmentation of the first objects, down to scales of just 1 pc. They find a very filamentary structure develops with the first stars forming from small knots at the intersection of filaments. Only a small fraction of the gas reaches sufficiently high densities to allow star formation. If these first stars were to feed back energy into the surrounding cloud and disrupt it, then that would suggest that the size of the first star clusters may be much smaller than the

size of the cloud from which they formed. However, the masses of these star clusters in the simulations is small, and these clusters may not produce many high-mass stars. Also, the simulations have not yet been run for long enough to see whether the surrounding cloud will collapse before feedback becomes effective, as we have suggested in Section 5.3. It should be noted that very different results were obtained by Bromm, Coppi & Larson (1999) using a particle-based hydrodynamics, but with much poorer resolution. They found that the cloud collapsed to a rotationally supported disc, which then broke up into very massive star clusters. It will be a while, we suspect, before any consensus emerges.

Fuller & Couchman (2000) simulated a cubical region of side $25\,h^{-1}\,$kpc in the SCDM cosmology with several different realizations of the density fluctuation spectrum, using an *N*-body, hydrodynamics code. This random realization is just what is required to look at the relative importance of haloes of different mass. They found that the most massive objects that collapsed within the region did so over a wide range of redshifts, 15–30, and had a similar large spread in mass. Unfortunately, their simulation volume was not large enough to sample our second-generation haloes.

6 CONCLUSIONS

In this paper we have considered the cooling of gas within spherical, virialized haloes in the high-redshift Universe. Our technique is similar to that used by T97, but with a more up-date cooling function and cosmological model. In addition, we have investigated haloes with a wider range of virial temperatures, in the range of 100 to 100 000 K.

We have followed the abundances of hydrogen and helium species, including molecular hydrogen. We use a more complete destruction term for H_2, taking into account the relative importance of the reverse reaction H7, which other authors have ignored. This makes a difference of a few per cent to the final H_2 abundances for haloes which cool from above 9000 K.

The main coolant for temperatures below 9000 K is molecular hydrogen. Unfortunately, the cooling rate seems to be poorly known, and a recent determination by GP98 gives values at temperatures below 1000 K over an order of magnitude lower than those of LS84. Consequently, the first objects to cool in the former do so later and have much higher masses and virial temperatures.

We follow T97 by defining clouds to have cooled if they lose 25 per cent or more of their energy in the time that the redshift has decreased by 25 per cent. We obtain similar virial temperatures for the smallest haloes that can cool, but note that the *first* objects to cool in each cosmology are more massive and have higher virial temperatures of about 4000 K. In the SCDM and ΛCDM cosmologies the formation redshift is $z_{0.75} \approx 20$; for the τCDM cosmology it is much lower, $z_{0.75} \approx 11$.

We identify a second generation of haloes that cool about 10 Myr after the first one. These are haloes with virial temperatures in excess of 9000 K for which there is a significant fraction of free electrons. The cooling is dominated by electronic transitions at high temperatures and is almost instantaneous, occurring on much less than a dynamical time. Just as significant, however, is the fact that the residual ionization is greater than in low-mass haloes and the production of H_2 is much greater. Consequently, they cool to much lower temperatures in a dynamical time than do the first generation of haloes.

Our model suffers from three deficiencies: we consider only 3σ density fluctuations, we ignore substructure, and we cannot follow the collapse of haloes whose cooling times are shorter than their dynamical times. More sophisticated studies are required to determine whether the first- or second-generation haloes are more important for determining the mass of the first star clusters, or whether they both have a role to play. We hope to report on these in future papers.

ACKNOWLEDGMENTS

PAT is a PPARC Lecturer Fellow. HMPC thanks the Canadian Institute for Advanced Research. We are grateful to Todd Fuller for helpful discussions. FS thanks his parents for their support.

REFERENCES

Abel T., Anninos P., Zhang Y., Norman M. L., 1997, New Astron., 2, 181
Abel T., Anninos P., Norman M. L., Zang Y., 1998, ApJ, 508, 518
Abel T., Bryan G., Norman M., 2000, ApJ, 540, 39
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Bond J. R., Efstathiou G. P., 1984, ApJ, 285, L45
Bromm V., Coppi P. S., Larson R. B., 1999, ApJ, 527, 5
Eke V. R., Navarro J. F., Frenk C. S., 1998, ApJ, 503, 569
Ettori S., Fabian A. C., 1999, MNRAS, 305, 834
Forrey R. C., Balakrishnan N., Dalgarno A., Lepp S., 1997, ApJ, 489, 100
Freedman W. L. et al., 2001, ApJ, 553, 47
Fuller T., Couchman H. M. P., 2000, ApJ, 544, 6
Galli D., Palla F., 1998, A&A, 335, 403 (GP98)
Grachev S. I., Dubrovich V. K., 1991, Astrophysics, 34, 124
Haiman Z., Thoul A. A., Loeb A., 1996, ApJ, 464, 523
Hansen S. H., Villante F. L., 2000, Phys. Lett. B, 486, 1
Lepp S., Shull J. M., 1984, ApJ, 280, 465 (LS84)
MacLow M. M., Shull J. M., 1986, ApJ, 302, 585
Martin P. G., Schwarz D. H., Mandy M. E., 1996, ApJ, 461, 265
Nakamura F., Umemura M., 1999, ApJ, 515, 239
Omukai K., Nishi R., 1998, ApJ, 508, 141
Peebles P. E. J., 1968, ApJ, 153, 1
Peebles P. E. J., 1993, Principles of Physical Cosmology. Princeton Univ. Press, Princeton
Press W. H., Schechter P. G., 1974, ApJ, 187, 425
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in Fortran. 2nd edition Cambridge Univ. Press
Sasaki S., Takahara F., 1993, PASJ, 45, 655
Shapiro P. R., Kang H., 1987, ApJ, 318, 32
Stancil P. C., Lepp S., Dalgarno A., 1998, ApJ, 509, 1
Tegmark M., Silk J., Rees M. J., Blanchard A., Abel T., Palla F., 1997, ApJ, 474, 1 (T97)
Viana P. T. P., Liddle A. R., 1996, MNRAS, 281, 323

The appendices describe two different ways of deriving equation (1) for the rate of change of abundance of H_2. We do this because nowhere in the literature does it seem to be spelt out in detail, and because our equation differs slightly from those used previously, as described in Section 2.1.

APPENDIX A: DERIVATION VIA REMOVAL OF H^- AND H_2^+ FROM THE REACTION NETWORK

The reaction rates for destruction of H^- and H_2^+ are much greater than for formation or destruction of H, H^+, e^- and H_2. Hence we can regard H^- and H_2^+ as short-lived species that are destroyed the instant that they are produced. This has the advantage that we can eliminate them from the reaction network, as described below.

Consider first H^-. This has two destruction channels, reactions

H4 and H5 in Table 1, whose combined rate is

$$R_{\mathrm{H}^-} = n_{\mathrm{H}} R_4 + n_{\mathrm{H}^+} R_5, \tag{A1}$$

where R_4 and R_5 are the reaction rates as listed in the table (at redshifts greater than 110 photodestruction of H^- by cosmic microwave background photons is also important, but we do not consider such high redshifts in this paper). Hence the fractions of H^- that decay via reactions H4 and H5 are $n_{\mathrm{H}} R_4/R_{\mathrm{H}^-}$ and $n_{\mathrm{H}^+} R_5/R_{\mathrm{H}^-}$, respectively. Reaction H4 is always important, whereas reaction H5 is significant only at high ionization levels.

Now reaction H3 for H^- production can instead be rewritten as two different reaction chains:

$$\text{H3H4}\quad 2\mathrm{H} + \mathrm{e}^- \mapsto \mathrm{H}_2 + \mathrm{e}^- + \gamma, \tag{A2}$$

$$\text{H3H5}\quad \mathrm{H} + \mathrm{e}^- + \mathrm{H}^+ \mapsto 2\mathrm{H} + \gamma, \tag{A3}$$

which occur at rates $n_{\mathrm{H}}^2 n_{\mathrm{e}^-} R_3 R_4/R_{\mathrm{H}^-}$ and $n_{\mathrm{H}} n_{\mathrm{e}^-} n_{\mathrm{H}^+} R_3 R_5/R_{\mathrm{H}^-}$, respectively. The first of these reaction chains forms molecular hydrogen using electrons as catalysts, whereas the second leads to a reduction in the ionization level.

We can treat H_2^+ in the same manner. Its decay channels are reactions H7 and H8, which sum to a total destruction rate of

$$R_{\mathrm{H}_2^+} = n_{\mathrm{H}} R_7 + n_{\mathrm{e}^-} R_8. \tag{A4}$$

Once again we have neglected photodestruction, which is important only at very high redshifts.

Then reaction 6 becomes the two reaction chains:

$$\text{H6H7}\quad 2\mathrm{H} + \mathrm{H}^+ \mapsto \mathrm{H}_2 + \mathrm{H}^+ + \gamma, \tag{A5}$$

$$\text{H6H8}\quad \mathrm{H} + \mathrm{H}^+ + \mathrm{e}^- \mapsto 2\mathrm{H} + \gamma, \tag{A6}$$

which are analogous to those for H^-. There is a second couplet starting with H_2:

$$\text{H9H7}\quad \mathrm{H}_2 + \mathrm{H}^+ + \mathrm{H} \mapsto \mathrm{H}_2 + \mathrm{H}^+ + \mathrm{H}, \tag{A7}$$

$$\text{H9H8}\quad \mathrm{H}_2 + \mathrm{H}^+ + \mathrm{e}^- \mapsto 3\mathrm{H}. \tag{A8}$$

Of these, H9H7 is the more important at low ionization levels, and will strongly suppress destruction of H_2 via collisions with protons.

Using the reactions listed in Table 1, but replacing reactions H3 to H9 with the reaction chains derived above, we can write down an equation for the rate of change of H_2 abundance:

$$\frac{\mathrm{d}n_{\mathrm{H}_2}}{\mathrm{d}t} = n_{\mathrm{H}}^2 \left(n_{\mathrm{e}^-} \frac{R_3 R_4}{R_{\mathrm{H}^-}} + n_{\mathrm{H}^+} \frac{R_6 R_7}{R_{\mathrm{H}_2^+}} \right) - n_{\mathrm{H}_2} \left(n_{\mathrm{H}^+} n_{\mathrm{e}^-} \frac{R_9 R_8}{R_{\mathrm{H}_2^+}} + n_{\mathrm{H}} R_{10} + n_{\mathrm{e}^-} R_{11} \right). \tag{A9}$$

APPENDIX B: DERIVATION USING EQUILIBRIUM VALUES OF H^- AND H_2^+

The second derivation uses the fact that the destruction rates for H^- and H_2^+ are high to derive equilibrium values for n_{H^-} and $n_{\mathrm{H}_2^+}$ which can then be eliminated from the equations.

Consider first the equation for the rate of change of n_{H^-}:

$$\mathrm{d}\frac{n_{\mathrm{H}^-}}{\mathrm{d}t} = n_{\mathrm{H}} n_{\mathrm{e}^-} R_3 - n_{\mathrm{H}^-} (n_{\mathrm{H}} R_4 + n_{\mathrm{H}^+} R_5). \tag{B1}$$

The destruction rates within the brackets are very large, which means that n_{H^-} will rapidly evolve to the equilibrium value in which creation and destruction of n_{H^-} balance and $\mathrm{d}n_{\mathrm{H}^-}/\mathrm{d}t \approx 0$. Then we have

$$n_{\mathrm{H}^-} \approx \frac{n_{\mathrm{H}} n_{\mathrm{e}^-} R_3}{n_{\mathrm{H}} R_4 + n_{\mathrm{H}^+} R_5} \equiv \frac{n_{\mathrm{H}} n_{\mathrm{e}^-} R_3}{R_{\mathrm{H}^-}}. \tag{B2}$$

Similarly, for $n_{\mathrm{H}_2^+}$ we have

$$n_{\mathrm{H}_2^+} \approx \frac{n_{\mathrm{H}} n_{\mathrm{H}^+} R_6 + n_{\mathrm{H}_2} n_{\mathrm{H}^+} R_9}{n_{\mathrm{H}} R_7 + n_{\mathrm{e}^-} R_8} \equiv \frac{n_{\mathrm{H}} n_{\mathrm{H}^+} R_6 + n_{\mathrm{H}_2} n_{\mathrm{H}^+} R_9}{R_{\mathrm{H}_2^+}}. \tag{B3}$$

We can now write down the equation for the rate of change of H_2, then substitute for n_{H^-} and $n_{\mathrm{H}_2^+}$.

$$\begin{aligned} \frac{\mathrm{d}n_{\mathrm{H}_2}}{\mathrm{d}t} &= n_{\mathrm{H}} n_{\mathrm{H}^-} R_4 + n_{\mathrm{H}} n_{\mathrm{H}_2^+} R_7 \\ &\quad - n_{\mathrm{H}_2} (n_{\mathrm{H}^+} R_9 + n_{\mathrm{H}} R_{10} + n_{\mathrm{e}^-} R_{11}) \\ &= n_{\mathrm{H}}^2 n_{\mathrm{e}^-} \frac{R_3 R_4}{R_{\mathrm{H}^-}} + n_{\mathrm{H}}^2 n_{\mathrm{H}^+} \frac{R_6 R_7}{R_{\mathrm{H}_2^+}} + n_{\mathrm{H}} n_{\mathrm{H}_2} n_{\mathrm{H}^+} \frac{R_9 R_7}{R_{\mathrm{H}_2^+}} \\ &\quad - n_{\mathrm{H}_2} (n_{\mathrm{H}^+} R_9 + n_{\mathrm{H}} R_{10} + n_{\mathrm{e}^-} R_{11}) \end{aligned} \tag{B4}$$

$$\begin{aligned} &= n_{\mathrm{H}}^2 \left(n_{\mathrm{e}^-} \frac{R_3 R_4}{R_{\mathrm{H}^-}} + n_{\mathrm{H}^+} \frac{R_6 R_7}{R_{\mathrm{H}_2^+}} \right) \\ &\quad - n_{\mathrm{H}_2} \left(n_{\mathrm{H}^+} n_{\mathrm{e}^-} \frac{R_9 R_8}{R_{\mathrm{H}_2^+}} + n_{\mathrm{H}} R_{10} + n_{\mathrm{e}^-} R_{11} \right). \end{aligned} \tag{B5}$$

Equation (B5) is identical to equation (A9) derived above.

This paper has been typeset from a TEX/LATEX file prepared by the author.