Mon. Not. R. Astron. Soc. **345,** 349–364 (2003)

# Virial shocks in galactic haloes?

Yuval Birnboim★ and Avishai Dekel★
*Racah Institute of Physics, The Hebrew University, Jerusalem, Israel*

Accepted 2003 June 24. Received 2003 May 25; in original form 2003 February 1

**ABSTRACT**

We investigate the conditions for the existence of an expanding virial shock in the gas falling within a spherical dark matter halo. The shock relies on pressure support by the shock-heated gas behind it. When the radiative cooling is efficient compared with the infall rate, the post-shock gas becomes unstable; it collapses inwards and cannot support the shock. We find for a monatomic gas that the shock is stable when the post-shock pressure and density obey $\gamma_{\rm eff} \equiv (\mathrm{d}\ln P/\mathrm{d}t)/(\mathrm{d}\ln\rho/\mathrm{d}t) > \frac{10}{7}$. When expressed in terms of the pre-shock gas properties at radius $r$ it reads as $\rho r\Lambda(T)/u^3 < 0.0126$, where $\rho$ is the gas density, $u$ is the infall velocity and $\Lambda(T)$ is the cooling function, with the post-shock temperature $T \propto u^2$. This result is confirmed by hydrodynamical simulations, using an accurate spheri-symmetric Lagrangian code. When the stability analysis is applied in cosmology, we find that a virial shock does not develop in most haloes that form before $z \sim 2$, and it never forms in haloes less massive than a few $10^{11}$ $\mathrm{M}_\odot$. In such haloes, the infalling gas is not heated to the virial temperature until it hits the disc, thus avoiding the cooling-dominated quasi-static contraction phase. The direct collapse of the cold gas into the disc should have non-trivial effects on the star formation rate and on outflows. The soft X-ray produced by the shock-heated gas in the disc is expected to ionize the dense disc environment, and the subsequent recombination would result in a high flux of L$\alpha$ emission. This may explain both the puzzling low flux of soft X-ray background and the L$\alpha$ emitters observed at high redshift.

**Key words:** hydrodynamics – shock waves – cooling flows – galaxies: formation – galaxies: ISM – dark matter.

## 1 INTRODUCTION

The standard lore in the idealized picture of galaxy formation by spherical infall of gas inside dark matter haloes is that the gas is first heated to the halo virial temperature behind an expanding virial shock. It is then supported by pressure in a quasi-static equilibrium while it is cooling radiatively and is slowly contracting to a disc where it can eventually form stars. The cooling process thus determines important galaxy properties such as the star formation rate and the metal enrichment, so it is necessarily an important ingredient in the galaxy formation process.

However, it is not at all clear that a stable shock can persist in the halo gas away from the disc under the conditions valid in many galactic haloes. In the absence of a virial shock, the gas is not heated to the virial temperature until it falls all the way to the disc, where the collapse stops and the gas is heated in a thin layer. This may alter some of the assumed processes of disc formation and, in particular, the star formation rate in it. It may work against blowout by supernova-driven winds in dwarf galaxies. The result of heating near the disc instead of at the virial radius may result in weakening of the soft X-ray emission from such haloes and producing a high flux of L$\alpha$ instead. In this paper we evaluate the conditions for the existence of a virial shock in galactic haloes.

Initial density perturbations are assumed to grow by gravitational instability, reach maximum expansion and collapse into virial equilibrium at roughly half the maximum expansion radius. During the initial phase, and roughly until shells start crossing each other near the virial radius, the gas pressure is negligible compared with the gravitational force, so the shells of gas and dark matter move in a similar manner. Once interior to the virial radius, where shells tend to cross and the gas density becomes high enough, the gas pressure becomes an important factor in the dynamics. Its hydrodynamic properties allow transfer of bulk kinetic energy into internal energy and the pressure prevents the gas element from passing through other gas elements and from being compressed without limit. This makes the infall velocity vanish at the centre. Since in the cold infalling gas the typical velocity is higher than the speed of sound, the information concerning this inner boundary condition cannot propagate outwards in time, and these supersonic conditions create a shock. After the gas crosses the shock, it is heated up, the speed of sound increases and the flow becomes subsonic.

★E-mail: yuval@frodo.fiz.huji.ac.il (YB); dekel@phys.huji.ac.il (AD)

The shock transfers the kinetic energy that has been built up during the collapse into internal gas energy just behind the shock. A stable spherical shock would slowly propagate outwards through the infalling gas, leaving behind it hot, high-entropy gas that is almost at rest. The temperature of the post-shock gas roughly equals the virial temperature. The persistence of the shock depends on sufficient pressure from the post-shock gas, which supports it against being swept inwards due to the gravitational pull together with the infalling matter. Radiative gas cooling makes the gas lose entropy and pressure, which weakens the pressure support behind the shock front. Our approach here is to evaluate the existence of a virial shock by analysing the gravitational stability of the supporting gas behind the shock in the presence of significant cooling.

In Section 2 we first summarize the standard analysis of an adiabatic shock and then generalize the gravitational stability criterion to the case where cooling is important. In Section 3 we describe our spherical hydrodynamic Lagrangian code, which includes gravitating dark matter and gas shells, artificial viscosity, radiative cooling and centrifugal forces. We test the code in this section and in Appendix A. In Section 4 we apply the numerical code to simulations that demonstrate the shock formation and test the validity of the analytical model. In Section 5 we apply the shock stability criterion to realistic haloes forming in cosmological conditions. In Section 6 we summarize our results and discuss potential astrophysical implications.

## 2 SHOCK STABILITY ANALYSIS

Our goal here is to derive a criterion for the existence of a virial shock in terms of the properties of the infalling gas just in front of the shock front. It is based on a gravitational stability analysis of the post-shock gas. We first remind ourselves of the standard stability analysis in the simple adiabatic case, and then derive a more general criterion for stability in the radiative case, under certain assumptions and using a perturbation analysis.

### 2.1 The standard adiabatic case

Throughout this paper, we treat the baryons as an ideal monatomic gas. Their equation of state could therefore be written as

$$P = (\gamma - 1)\, e\rho, \tag{1}$$

where $P$ is the pressure, $e$ is the specific internal energy, $\rho$ is the density of the gas and $\gamma$ is the adiabatic index. Along an isentrope (an adiabatic process of constant entropy) the pressure and density are related via $P \propto \rho^{\gamma}$, so the adiabatic index is defined by

$$\gamma = \left( \frac{\partial \ln P}{\partial \ln \rho} \right)_{\rm s}. \tag{2}$$

For a monatomic gas $\gamma = \frac{5}{3}$.[1]

The virial shock is assumed to be a spherical accretion shock that propagates outwards slowly while infalling gas crosses it inwards.

The kinetic energy of the infalling gas is transformed at the shock front into thermal energy – the post-shock gas is thus heated to a temperature close to the virial temperature of the system of dark matter halo and gas, $V^2_{\rm infall} \approx k_{\rm B} T_{\rm vir}$. Because the original temperature of the infalling gas is negligible compared with the virial temperature, the system obeys the strong-shock limit. When we denote the pre-shock and post-shock quantities by subscripts 0 and 1, respectively, the jump conditions across the shock are in this case (Zel'dovich & Raiser 1966):

$$\rho_0 = \frac{\gamma - 1}{\gamma + 1} \rho_1, \tag{3}$$

$$(u_0 - u_{\rm s}) = \frac{\gamma + 1}{\gamma - 1} (u_1 - u_{\rm s}), \tag{4}$$

$$P_1 = \frac{2\rho_0 u_0^2}{\gamma + 1}, \tag{5}$$

$$T_1 = \frac{\mu}{k_{\rm B} N_{\rm a}} \frac{P_1}{\rho_1} = \frac{\mu}{k_{\rm B} N_{\rm a}} \frac{2\gamma - 1}{(\gamma + 1)^2} u_0^2, \tag{6}$$

where $u$ denotes the radial velocity, $u_{\rm s}$ is the shock velocity, $N_{\rm a}$ is Avogadro's number, $N_{\rm a}/\mu$ is the average number of molecules per unit mass and $k_{\rm B}$ is Boltzmann's constant.

According to standard shock theory, the post-shock gas is always subsonic (in the frame of reference of the moving shock) because of the increasing sound velocity behind the shock. This gas is thus capable of providing the necessary pressure to support the shock against the gravitational pull inwards applied by the self-gravity of the gas and the dark matter halo as well as the pressure applied by the infalling matter at the shock front.

The criterion for gravitational stability of this post-shock gas in the adiabatic case is the standard Jeans stability criterion: $\gamma > \frac{4}{3}$ (e.g. Cox 1980, Chapter 8).

If the post-shock gas is gravitationally unstable, it falls into the galaxy centre on a dynamical time-scale and can no longer support the shock. As a result, the shock weakens and it is swept inwards.

The Jeans criterion can be understood qualitatively in terms of the following heuristic derivation. For a shell of radius $r$, we compare the gravitational pull inwards, $a_{\rm g} = GM/r^2$ (where $M$ is the mass interior to $r$), to the pressure pushing outwards, $a_{\rm p} = \rho^{-1} \nabla P$. We assume an isentrope, $P \propto \rho^{\gamma}$. We also assume homology, such that the local density scales such that the mean density in the sphere interior to $r$, $\rho \propto M/r^3$. Then $\nabla P$ can be replaced by $\sim P/r$ and we obtain

$$\frac{a_{\rm p}}{a_{\rm g}} \propto \rho^{\gamma - 4/3}. \tag{7}$$

If $\gamma < \frac{4}{3}$, we have an unstable configuration. Starting in hydrostatic equilibrium, $a_{\rm p}/a_{\rm g} = 1$, a perturbation involving contraction is associated with a larger $\rho$, and therefore $a_{\rm p}/a_{\rm g} < 1$ by equation (7), implying that the pressure cannot prevent collapse. If $\gamma > \frac{4}{3}$, the pressure force increases until it balances the increased gravitational pull. We note that even this simple derivation of the Jeans criterion had to assume homology – an assumption that we will have to also adopt in our analysis of the radiative case below.

### 2.2 Shock stability under radiative cooling

We wish to replace the adiabatic Jeans criterion by a more general stability condition that will also be valid in the radiative case. This criterion must depend on the cooling rate and should therefore be naturally expressed in terms of time derivatives. We generalize the adiabatic $\gamma$ of equation (2) by an effective $\gamma$ following a comoving volume element along its Lagrangian path:

$$\gamma_{\rm eff} \equiv \frac{{\rm d} \ln P / {\rm d}t}{{\rm d} \ln \rho / {\rm d}t}. \tag{8}$$

[1] As the temperature exceeds the binding energy of the hydrogen and helium atoms, electrons become detached from the nuclei and $\gamma$ becomes smaller. Once the gas becomes fully ionized, the original value of $\frac{5}{3}$ is restored, but with a different effective density. This should have only a marginal effect on our results, and is ignored in this paper.

We expect that the system would be stable when $\gamma_{\rm eff}$ is larger than a certain critical value, the analogue of the requirement $\gamma > \frac{4}{3}$ in the adiabatic case.

In our Lagrangian analysis all the quantities ($r$, $u$, $\rho$, $P$, etc.) refer to comoving shells; they are all functions of the gas mass $m$ interior to radius $r$ and time $t$. Derivatives with respect to time following a comoving volume element will be denoted by an upper dot and derivatives with respect to $m$ will be denoted by a prime.

The effective gamma can be related to its adiabatic analogue given the cooling rate and other post-shock gas quantities. The time derivative of equation (1) yields

$$\dot{P} = (\gamma - 1)(\dot{e}\rho + e\dot{\rho}). \tag{9}$$

Energy conservation in the presence of radiative losses can be expressed by

$$\dot{e} = -P\dot{V} - q = \frac{P\dot{\rho}}{\rho^2} - q, \tag{10}$$

where $q$ is the radiative cooling rate (to be discussed below, e.g. equation 20) and $V = \rho^{-1}$ is the specific volume. Substituting $\dot{e}$ from equation (10) into equation (9), and using it in equation (8), we obtain

$$\gamma_{\rm eff} = \gamma - \frac{\rho}{\dot{\rho}}\frac{q}{e}. \tag{11}$$

Note that in the limit $q/e \ll \dot{\rho}/\rho$ we reproduce the adiabatic case; the process is nearly adiabatic when the cooling time-scale is long compared with the contraction time-scale.

We assume that in the region close behind the shock the pattern of the velocity field is *homologous*. By this we mean that at any given time the (radial) velocity is proportional to the radius (as in a Hubble flow), namely

$$u/r = u_1/r_{\rm s}, \tag{12}$$

thus providing a boundary condition for the post-shock gas. The homology is shown to be a valid approximation in the simulations discussed below, where the post-shock shell trajectories are roughly parallel to each other in the $\log r$–$t$ plane, at any given time close enough to shock crossing. The time evolution of the density can then be evaluated via the *continuity* equation in Lagrangian form for the spheri-symmetric case,

$$\frac{\dot{\rho}}{\rho} = -\nabla\cdot\mathbf{u} = -\frac{1}{r^2}\frac{\partial}{\partial r}(r^2 u) = -\frac{3u_1}{r_{\rm s}}, \tag{13}$$

where the last equality results from the assumed homology, equation (12). The homology thus implies that $\dot{\rho}/\rho$ at a given $t$ is a constant in $m$ throughout the post-shock region. Equation (11) can then be simplified:

$$\gamma_{\rm eff} = \gamma + \frac{r_{\rm s}}{3u_1}\frac{q}{e}. \tag{14}$$

We start with a hypothetical unperturbed state for the post-shock gas, where we assume that the net force *vanishes*, $\ddot{r} = 0$. The system adjusts itself to this state on a time-scale associated with the speed of sound $c_{\rm s}$, provided that it is much higher than the infall velocity $u$. This is expected to be the case in the subsonic post-shock medium, where $c_{\rm s}$ becomes high and $u$ becomes low. The unperturbed equation of *motion* in Lagrangian form is then

$$\ddot{r} = -4\pi r^2 P' - \frac{GM}{r^2} = 0, \tag{15}$$

where $M$ is the total mass interior to radius $r$.

We then introduce a perturbation due to a homologous infall velocity $u$. Over a short time interval $\delta t$, it introduces a small displacement inwards, $\delta r = u\delta t$. In order to distinguish between stability and instability we wish to determine whether the induced acceleration, $\delta\ddot{r}$, is positive or negative, tending to decrease or increase the velocity, respectively. Note that under homology, equation (12), the relative displacement is

$$\delta r/r = u_1\delta t/r_{\rm s}. \tag{16}$$

Writing the equation of motion, equation (15), but for the perturbed quantities $P + \delta P$ and $r + \delta r$, and subtracting the unperturbed equation (15), we obtain to first order

$$\delta\ddot{r} = -4\pi r^2(\delta P)' + \frac{4GM\delta r}{r^3}. \tag{17}$$

Next, we manipulate the right-hand side of equation (17) to obtain a simple expression involving $\gamma_{\rm eff}$.

In the second term we use the homology, equation (16), and then the unperturbed equation of motion, equation (15), to obtain

$$\frac{4GM\delta r}{r^3} = -\frac{16\pi r^2 u_1\delta t}{r_{\rm s}}P'. \tag{18}$$

The manipulation of the first term is somewhat more elaborate. We use the definition of $\gamma_{\rm eff}$, equation (8), to write

$$\delta P = (\dot{\rho}/\rho)P\gamma_{\rm eff}\delta t = -(3u_1/r_{\rm s})P\gamma_{\rm eff}\delta t, \tag{19}$$

where the second equality is due to equation (13). Note that the $m$ dependence in this term is only in the product $P\gamma_{\rm eff}$. We now express $\gamma_{\rm eff}$ in terms of the cooling rate $q$ as in equation (14), and need to take the derivative $(Pq/e)'$. Here we use the standard assumption that the radiative cooling rate is proportional to density,

$$q = \rho\Lambda(T), \tag{20}$$

where $\Lambda\,(T)$ is the macroscopic cooling function and $T$ is the post-shock temperature. The immediate post-shock medium is assumed to be isothermal, reflecting via the jump conditions an assumed approximate uniformity of the pre-shock gas over a short time interval. Using equation (1) we have $P/e = (\gamma - 1)\rho$, and together with equation (20) it becomes

$$Pq/e = (\gamma - 1)\Lambda\rho^2. \tag{21}$$

In the computation of $(Pq/e)'$, we first replace $\rho' = ({\rm d}\rho/{\rm d}P)P'$, then use equation (8) to write ${\rm d}\rho/{\rm d}P = \gamma_{\rm eff}^{-1}\rho/P$, use equation (21) backwards to replace $(\gamma - 1)\Lambda\rho^2/P$ by $q/e$, and finally use equation (14) to obtain $(Pq/e)' = 3u_1/r_{\rm s}[-2\gamma_{\rm eff}^{-1}(\gamma - \gamma_{\rm eff})P']$. We thus have in the first term of the right-hand side of equation (17)

$$-(\delta P)' = \frac{3u_1\delta t}{r_{\rm s}}P'[\gamma - 2\gamma_{\rm eff}^{-1}(\gamma - \gamma_{\rm eff})]. \tag{22}$$

With the right-hand side of equation (17) given by equation (22) and (18), the first-order equation finally becomes,

$$\delta\ddot{r} = \frac{12\pi r^2 u_1\delta t P'}{r_{\rm s}}\left[\gamma - 2\gamma_{\rm eff}^{-1}(\gamma - \gamma_{\rm eff}) - \frac{4}{3}\right]. \tag{23}$$

Since $u_1$ and $P'$ are both always negative, the desired sign of $\delta\ddot{r}$ is determined by the sign of the expression inside the square brackets. Note that in the adiabatic case, $q = 0$, we have $\gamma_{\rm eff} = \gamma$, so we recover the standard stability criterion, $\gamma > \frac{4}{3}$. In the radiative case, $\gamma_{\rm eff} \neq \gamma$, we finally obtain the generalized stability criterion:

$$\gamma_{\rm eff} > \frac{2\gamma}{\gamma + \frac{2}{3}} \equiv \gamma_{\rm crit}. \tag{24}$$

For a monatomic gas, where the adiabatic value is $\gamma = \frac{5}{3}$, the threshold for stability is $\gamma_{\rm crit} = \frac{10}{7} = 1.43$, which is close but not identical to the adiabatic threshold $\frac{4}{3}$.

### 2.3 Stability in terms of pre-shock quantities

Next, we wish to express $\gamma_{\rm eff}$ and the stability criterion in terms of the properties of the pre-shock gas; the infall velocity $u_0$ and the gas density $\rho_0$ at $r_{\rm s}$. We use the jump conditions, equations (3)–(6), in equation (14). In equation (4) we assume $u_{\rm s} = 0$, namely that the shock is temporarily at *rest*, which should be valid when the shock is marginally stable (or unstable). This is because a stable shock is pushed outwards by the post-shock gas, while cooling reduces the pressure, slows the outward motion, and eventually causes it to halt and then be swept inwards by the infalling matter and gravitational pull. The transition from stability to instability can thus be associated with a transition from expansion to contraction of the shocked volume.

According to equation (1) and (5) we have

$$e_1 = \frac{1}{(\gamma-1)}\frac{P_1}{\rho_1} = \frac{2u_0^2}{(\gamma+1)^2}. \tag{25}$$

According to equation (13) and (4) with $u_{\rm s} = 0$ we have

$$\frac{\dot\rho}{\rho} = -\frac{3u_1}{r_{\rm s}} = -3\frac{(\gamma-1)}{(\gamma+1)}\frac{u_0}{r_{\rm s}}. \tag{26}$$

With these and equation (20) we obtain the desired expression for the effective $\gamma$ of the post-shock gas in terms of the pre-shock conditions:

$$\gamma_{\rm eff} = \gamma - \frac{(\gamma+1)^4}{6(\gamma-1)^2}\frac{\rho_0 r_{\rm s}\Lambda(T_1)}{|u_0|^3}. \tag{27}$$

For a monatomic gas, $\gamma = \frac{5}{3}$, we obtain

$$\gamma_{\rm eff} = \frac{5}{3} - 18.96\frac{\rho_0 r_{\rm s}\Lambda(T_1)}{|u_0|^3}. \tag{28}$$

Based on equation (24), the criterion for stability of a $\gamma = \frac{5}{3}$ gas finally becomes

$$\frac{\rho_0 r_{\rm s}\Lambda(T_1)}{|u_0|^3} < 0.0126. \tag{29}$$

The post-shock temperature is related to the pre-shock infall velocity using the jump condition, equation (6), which for $\gamma = \frac{5}{3}$ gives

$$T_1 = \frac{3}{16}\frac{\mu}{k_{\rm B}N_{\rm a}}u_0^2. \tag{30}$$

For a given cooling function $\Lambda(T)$, equation (29) is a simple criterion for determining whether a stable shock can form at some radius $r_{\rm s}$ of the halo. It is in a form that can be directly tested against hydrodynamic simulations (Section 3), and can serve for evaluating shock stability under realistic conditions in cosmological haloes (Section 5).

Under the simplifying assumption that the gas is unclumped, the cooling rate is given by equation (20). The macroscopic cooling function $\Lambda(T)$ is related to the microscopic $\Lambda_{\rm mic}(T)$, the energy-loss rate of a particle, via $\Lambda(T) = (N_{\rm a}^2\chi^2/\mu^2)\Lambda_{\rm mic}(T)$, where $\chi$ is the number of electrons per particle. We assume a helium atomic fraction of 0.1 for $\mu$ and $\chi$. The microscopic cooling function is shown in Fig. 1 for three different values of mean metallicity $Z$. The cooling at temperatures below $10^4$ K is very slow because the main available cooling agent is molecular hydrogen, which is very inefficient. At temperatures slightly above $10^4$ K the cooling function peaks due to L$\alpha$ emission from atomic hydrogen. At very low metallicities, a second peak arises near $10^5$ K due to recombination of atomic helium. Metals give rise to a higher peak at $\sim 10^5$ K and slightly above, due to line emission from the heavier atoms. At $\sim 10^6$ K and above, the cooling is dominated by bremsstrahlung, and the cooling function increases slowly. We use the cooling function as derived by Sutherland & Dopita (1993), and presented in their table, in the manner described in Somerville & Primack (1999).

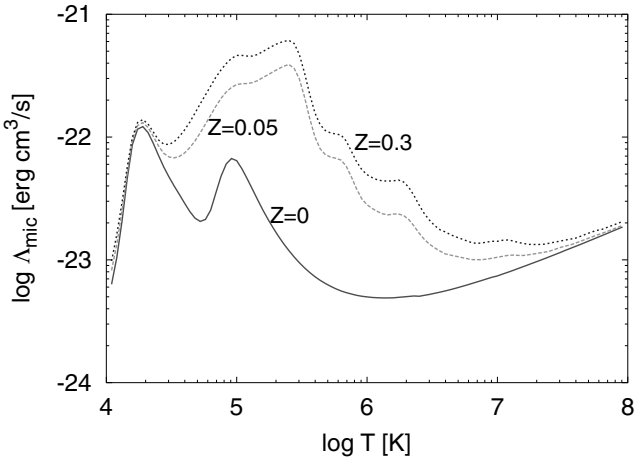


**Figure 1.** The microscopic cooling function of low-density gas for three different mean metallicities as indicated in solar units. The peaks are dominated by atomic H and He (for $Z = 0$) and by line emission from heavier atoms (for the higher $Z$ values). This figure is available in colour in the on-line version of the journal on *Synergy*.

## 3 THE SPHERICAL HYDRODYNAMICS CODE

We test the validity of the shock stability criterion using numerical simulations based on a spherical hydrodynamics code that follows the evolution of shells of dark matter and gas. Since the problem we intend to examine is of global spherical symmetry, and since we need to follow the cooling and the shock with high precision, we use a one-dimensional (1D) code. Most of the simulations presented here were run using 2000 gas shells and 10 000 dark matter shells. A comparable resolution in a three-dimensional (3D) code would require of the order of $10^{10}$ and $10^{12}$ particles, respectively, which is impractical. We use no smoothing in the dark matter shell-crossing scheme, we introduce small-scale smoothing at the halo centre to avoid an artificial singularity there, and we include small artificial viscosity in the hydrodynamics. Tests of the code performance are described in Appendix A.

### 3.1 Dark matter

The dark matter particles are represented by infinitely thin spherical shells of constant mass and of radii $r$ that vary in time. The shell of current radius $r$ obeys the equation of motion

$$\frac{{\rm d}^2 r}{{\rm d}t^2} = -\frac{G(M+m)}{(r+a)^2} + \frac{j^2}{r^3}, \tag{31}$$

where $M$ and $m$ refer, respectively, to the mass of dark matter and gas within the sphere of radius $r$. The last term is a centrifugal acceleration, determined by the specific angular momentum $j$ of the particle represented by the shell. This $j$ is assigned to each shell at the initial conditions and is assumed to be preserved during the simulation. The parameter $a$ is the smoothing length that becomes effective only near the centre; it has been set to be 50 pc throughout this work. The dark matter shells are allowed to cross each other (and the gas shells). The dark matter mass is evaluated by

$$M(r) = \sum_{r_i<r}\Delta M_i + \frac{1}{2}\sum_j \delta(r=r_j)\Delta M_j, \tag{32}$$

where the shell radii and masses are denoted by $r_i$ and $\Delta M_i$, respectively, $i = 1, \ldots, n_{\rm d}$. The second term adds half the mass of a shell when $r$ coincides with one of the shells. Generally, this summation requires $n_{\rm d}^2$ calculations for the dark matter alone. The particles are kept sorted by radius. When two shells cross each other, we resort the array by exchanging pairs that violate the order. This kind of sorting algorithm, termed 'shell's method' in Numerical Recipes (Press et al. 1997), is natural in cases where only a few shells cross each other in each time-step. When two shells cross, they exchange an energy of $G\Delta M_i\,\Delta M_j/r$. In order to conserve energy, the radius at which the shells cross must be known with great precision. We therefore reduce the time-step to a small value, $t_{\rm sc}$, when two shells are about to cross each other (see below).

### 3.2 Gas

The hydrodynamic part of the code is based on Lagrangian finite elements in the form of spherical shells. The basic equations governing the dynamics of each shell are

$$\frac{{\rm d}^2 r}{{\rm d}t^2} = -\frac{1}{\rho}\nabla(P+\sigma) - \frac{G(M+m)}{(r+a)^2} + \frac{j^2}{r^3}, \tag{33}$$

$$\frac{{\rm d}e}{{\rm d}t} = \rho^2(P+\sigma)\frac{{\rm d}\rho}{{\rm d}t} - q, \tag{34}$$

$$\rho = \frac{{\rm d}m}{4\pi r^2\,{\rm d}r}, \tag{35}$$

$$P = (\gamma - 1)e\rho. \tag{36}$$

An artificial viscosity term, $\sigma$, is added to the pressure for numerical purposes, as explained below. The smoothing length effective at the centre, $a$, is the same as for the dark matter, equation (31). As in the model described before, the loss of internal energy due to radiative cooling is represented by the cooling rate $q$.

The gas is divided into discrete shells. The mass enclosed within a shell, $\Delta m$, is assumed to be constant in time, while the inner- and outer-shell boundaries move independently in time. Each boundary is characterized by a temporal position $r$, velocity $v$ and specific angular momentum $j$. The acceleration (equation 33) is evaluated at the boundary position. The variables $\rho$, $P$, $e$, $q$ and $T$ for each shell are evaluated within the shell between the boundaries.

In particular, the pressure term in equation (33) is evaluated at the outer boundary $r_i$ of shell $i$ using equation (35):

$$-\frac{1}{\rho}\nabla(P+\sigma) = \frac{4\pi r_i^2}{\Delta m}[(P+\sigma)_{i+1} - (P+\sigma)_i], \tag{37}$$

where $\Delta m = (\Delta m_i + \Delta m_{i+1})/2$.

The boundary conditions for the outer boundary of the system are $P = \sigma = 0$, and zero mass beyond the outer boundary.

Since gas shells cannot cross each other, the gas mass in the sphere interior to each gas shell is constant throughout the simulation:

$$m(r_i) = \sum_{j=1}^{i} \Delta m_j. \tag{38}$$

For the evolution of the dark matter shell at $r$, we evaluate the gas mass that appears in equation (31) using

$$m(r) = \sum_{j=1}^{i-1} \Delta m_j + \frac{r^3 - r_{i-1}^3}{r_i^3 - r_{i-1}^3}\Delta m_i, \tag{39}$$

where $i$ refers to the gas shell for which $r_{i-1} \leqslant r < r_i$.

### 3.3 Integration and time-step

The discrete integration of $r$ and $v$ is performed using a Runge–Kutta fourth-order scheme (Press 1997). The state of the system at the beginning of each time-step is kept in memory until the time-step is completed, such that it is possible to return to the beginning of the time-step and retry with a smaller time-step if the convergence criteria are not met. The time-steps are set such that the position $r$ and velocity $v$ do not change by too much during a single time-step. For a given accuracy parameter $\epsilon_{\rm rk}$, we demand that the difference between the fourth-order displacement $\Delta r_4$ and the analogous first-order displacement $\Delta r_1$ obeys $|\Delta r_4 - \Delta r_1|/r < \epsilon_{\rm rk}$, both for the dark matter and the gas. A similar requirement is applied to the change in velocity over a time-step. If this condition is not fulfilled, we reduce the time-step by a certain factor and repeat the calculation over this time-step. We use here as our default $\epsilon_{\rm rk} = 0.1$.

In addition, we make sure the time-step for each shell does not violate the Courant condition, for an accuracy parameter $\epsilon_{\rm c}$. This implies $c_{\rm s}\Delta t/\Delta r < \epsilon_{\rm c}$, where ${c^2}_{\rm s} = ({\rm d}P/{\rm d}\rho)_{\rm s} = \gamma P/\rho$ is the speed of sound. We use here as our default $\epsilon_{\rm c} = 0.3$.

A third limitation on the time-step comes from the desire to conserve energy when shells cross. When two shells are about to cross each other within the current time-step ${\rm d}t$, we set the time-step to $\min({\rm d}t, t_{\rm sc})$, and keep it small until they actually cross. We use here as our default $t_{\rm sc} = 10^{-4}$ Gyr.

The values for $\epsilon_{\rm c}$, $\epsilon_{\rm rk}$ and $t_{\rm sc}$ were chosen empirically such that energy is conserved and the dynamics converges to our satisfaction, in the sense that it does not change by much when smaller parameters are used. We demonstrate in Appendix A how well these requirements are met.

Once we have computed the new radii and velocities of the shells at the end of the time-step, we correct the energy of the gas for the $-P\,{\rm d}V$ work term using the states of the system at the beginning and at the end of the time-step. The cooling is explicitly subtracted from the internal energy after the hydrodynamic time-step is completed. Once the final state of the system is ready, it is copied on to the memory array of the initial state, and the simulation is ready to execute a new time-step.

### 3.4 Initial conditions

The simulation starts at high redshift, $z = 100$, with a small spherical density perturbation. The initial density fluctuation profile is set to be proportional to the linear correlation function of the assumed cosmological model, representing the typical perturbation under the assumption that the random fluctuation field is Gaussian (see Dekel 1981 and Appendix C). The amplitude of the density fluctuation at the initial time, averaged over a given mass, determines the time of collapse, as desired. The initial velocity field is assumed to follow a quiet Hubble flow and the radial peculiar velocities build up in time. We assume the standard $\Lambda$ cold dark matter ($\Lambda$CDM) cosmology with $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, $h = 0.7$ and $\sigma_8 = 1$.

### 3.5 Angular momentum

We assume that in a real system the orbits of dark matter particles, and the initial orbits of the gas particles, are quite elongated. Cosmological $N$-body simulations show that the velocity distribution tends to be more radial than tangential (Ghigna et al. 1998) and already for an isotropic distribution the eccentricities are approximately 1:6. The processes we study in this paper occur away from the galactic disc at a radius of the order of the virial radius, namely in a regime

where the centrifugal force can be expected to be negligible compared with the gravitational force and the gas pressure force. The prescribed angular momentum for the shells is thus mainly for numerical purposes, to avoid divergent densities of gas or dark matter shells when they pass through the halo centre. Our results concerning the virial shock are insensitive to the actual way in which we assign angular momentum to each shells.

In the current study we practically assume that the dark matter particles are almost on radial orbits. The angular momentum of the gas is prescribed such that the shells, once they lost their energy by radiation, would settle into an exponential disc with pure circular motions and a characteristic radius of a few kpc, smaller than the inner characteristic radius of the halo. Our spherical 'disc' thus contains gas that is cold and dense compared with the shocked gas.

### 3.6 Artificial viscosity

It is impossible to follow the discontinuous behaviour across the shock using the conventional continuity equation for the density and standard conservation of energy and momentum. The jump conditions can be calculated explicitly, as in equations (3)–(6) (termed 'the characteristic method' or 'Godonov's method'). Alternatively, as proposed by Von Newman, one can slightly smear the discontinuities and then solve them within the framework of the standard hydrodynamic equations. By adding an artificial pressure term in a few shells around the shock, the differential equations become solvable and one can continue the calculation without affecting the energy and the dynamics of the shock (while its internal structure naturally changes).

Artificial viscosity is applied when the inner- and outer-shell boundaries at $r_1$ and $r_2$ approach each other, $\Delta v = v_2 - v_1 < 0$, and when the volume of the shell decreases, $\mathrm{d}V/\mathrm{d}t = 4\pi(r_2^2 v_2 - r_1^2 v_1) < 0$. The artificial viscosity then takes the form

$$\sigma = a_2 \rho (\Delta v)^2 + a_1 \rho c_\mathrm{s} |\Delta v|. \quad (40)$$

The quadratic, common form of artificial viscosity smears discontinuities over approximately three shells. The linear discontinuity affects a slightly larger range, and is usually added with a smaller coefficient $a_1$. The coefficients $a_1$ and $a_2$ are varied for different shells during the course of the simulation in order to overcome a specific numerical problem in the cold 'disc', where the gravitational and centrifugal forces balance each other and the pressure force is negligible. In this case the gas is not a standard hydrodynamic gas because the pressure does not regulate large discontinuities, and information is not transported because of the low speed of sound. When a 'disc' shell vibrates, it is artificially heated by the artificial viscosity in every contraction until its pressure grows and stops the process. If we are not careful to properly tune the artificial viscosity we may end up with one 'disc' shell that has been heated to $10^7$ K, while the rest of the 'disc' is at $10^4$ K. This imposes an undesired drastic decrease in the corresponding time-step. In order to overcome this numerical problem, we gradually turn-off the quadratic term ($a_2$) of the artificial viscosity inside the 'disc'. We define a 'disc' radius $R_\mathrm{disc}$ to be the largest radius for which the difference between the gravitational and centrifugal forces is less than $\frac{1}{4}$ of the gravitational force. Once at $r < 0.6R_\mathrm{disc}$, we continuously decrease the parameter $a_2$ in equation (40) according to $a_2 = (r - 0.3R_\mathrm{disc})/(0.3R_\mathrm{disc})$ and make it completely vanish at $r < 0.3R_\mathrm{disc}$. This prescription was found by trial and error to properly solve the numerical problem in most cases. The linear term of the viscosity, being proportional to the speed of sound, is anyway very small in the cold 'disc', so effectively no artificial viscosity is applied in the inner 'disc'.

Appendix A provides tests and examples of the hydrodynamic simulations in some detail.

## 4 VIRIAL SHOCK IN THE SIMULATIONS

### 4.1 Existence of a virial shock

We now investigate the formation of a virial shock using the spherical hydrodynamical simulations described above. In particular, we wish to test the validity of the analytic stability criterion developed in Section 2.

In order to mimic a typical perturbation in a random Gaussian field (Dekel 1981, Appendix B), the initial density-fluctuation profile was set to be proportional to the correlation function, normalized such that the mean density fluctuation in a sphere enclosing $M_\mathrm{i} = 10^{11}\ \mathrm{M}_\odot$ was $\delta_\mathrm{i} = 0.09$ at $z = 100$. For example, the shell encompassing $M \sim 3 \times 10^{10}\ \mathrm{M}_\odot$ is expected to collapse at $z = 3$, and $M \sim 10^{12}\ \mathrm{M}_\odot$ is expected to collapse at $z = 0$.

Fig. 2 shows the time evolution of the radii of Lagrangian shells in a simulation of the adiabatic case, with the cooling turned off. We find that a shock exists at all times. It appears as a sharp break in the flow lines, associated with a discontinuous decrease in infall velocity (equation 4). Shown in the figure is the shock radius, defined by the outermost shell for which the inner- and outer-shell boundaries approach each other and the volume of the shell decreases (the same conditions that have been used for turning on the artificial viscosity in equation 40). The shock gradually propagates outwards, encompassing more gas mass and dark matter in time. The gas below the shock is pressure supported and at quasi-static equilibrium. Not shown here are the dark matter shells, which collapse, oscillate and tend to increase the gravitational attraction exerted on the gas shells.

Shown in comparison is the evolution of the virial radius, computed from the simulation density as the radius within which the mean overdensity is $\Delta_\mathrm{v}$ times the mean cosmological background

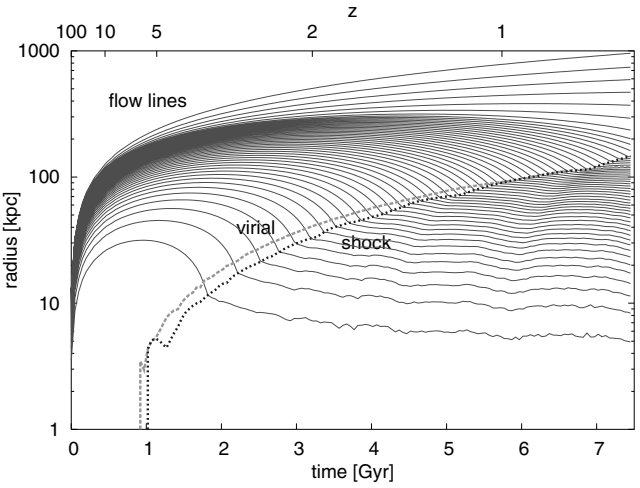


**Figure 2.** Simulation of the *adiabatic* case. The sequence of (solid red) curves describe the log radii of Lagrangian gas shells as a function of time. The simulation was of 2000 gaseous shells (shown here) and 10 000 dark matter shells. The radius of every 20th gaseous shell is plotted. Shown on top are the virial radius and the shock. The shock exists at all times. It gradually propagates outwards, and it practically coincides with the virial radius. This figure is available in colour in the on-line version of the journal on *Synergy*.

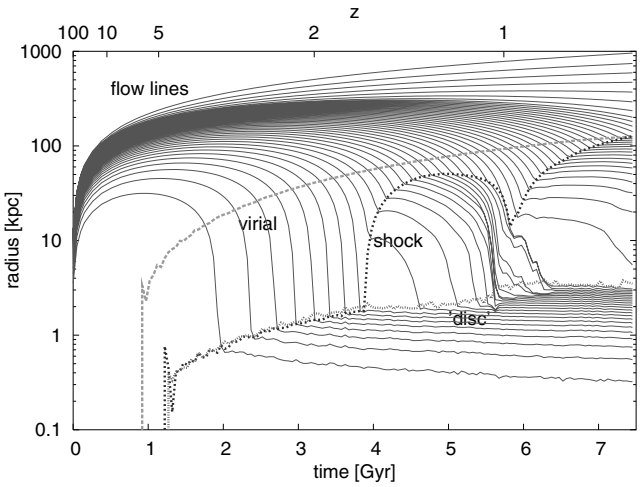


**Figure 3.** Simulation of the *radiative cooling* case, with $Z = 0$. The curves are as in Fig. 2, with the 'disc' radius added. There is no shock outside the 'disc' at early times, when the virial mass is small, because the cooling is too efficient. A shock develops at later times, when the mass is larger, and it quickly propagates outwards. After a couple of oscillations the shock radius approaches the virial radius. This figure is available in colour in the on-line version of the journal on *Synergy*.

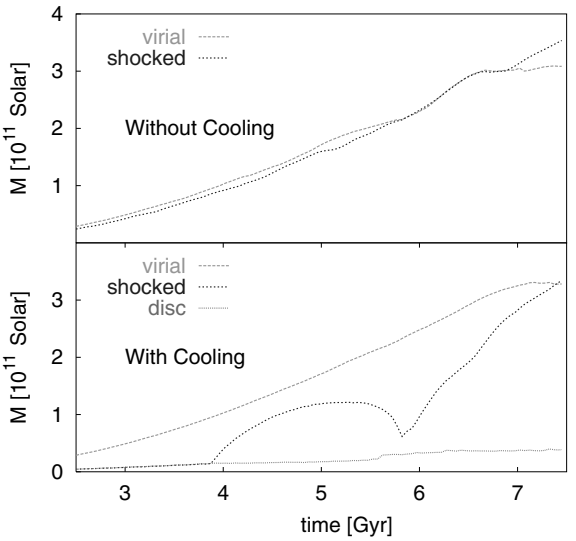


**Figure 4.** Evolution of the total mass interior to the virial radius, shock radius and 'disc' radius, in the adiabatic simulation (top) and the radiative simulation (bottom). This figure is available in colour in the on-line version of the journal on *Synergy*.

density. The virial overdensity $\Delta_v$ is provided by the dissipationless spherical top-hat collapse model; it is a function of the cosmological model, and it may vary with time. For the Einstein–de Sitter cosmology, the familiar value is $\Delta_v \simeq 176$ at all times.[2] For the family of flat cosmologies ($\Omega_m + \Omega_\Lambda = 1$), the value of $\Delta_v$ can be approximated by Bryan & Norman (1998)

$$\Delta_v \simeq (18\pi^2 + 82x - 39x^2)/(1 + x), \tag{41}$$

where $x \equiv \Omega_m(z) - 1$, and $\Omega_m(z)$ is the ratio of mean matter density to critical density at redshift $z$. For example, in the ΛCDM cosmological model that serves as the basis for our analysis in this paper ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$), the value at $z = 0$ is $\Delta_v \simeq 340$. We see in Fig. 2 that the shock radius almost coincides with the virial radius at all times. This is hardly surprising, as the shock is likely to appear at the outermost radius at which shell crossing first occurs, which is near the virial radius (to be demonstrated in Fig. 8 below).

Fig. 3 is the result of a similar simulation, but now with realistic radiative cooling for $Z = 0$. We see that a stable shock does not exist in this case before $t = 3.9$ Gyr. During this period, the cooling makes the gas lose its pressure support and lets it collapse freely under gravity into the halo centre. The collapse stops by the assumed angular momentum, in a 'disc' where the marked radius can be identified at the bottom of the plot by the abrupt change of the infalling flow lines into horizontal lines. The matter in the 'disc' is angular-momentum supported. As is visible in the figure, a shock, in the sense of a discontinuity in velocity and density, is present at the edge of the disc. Once the stability criterion is met, a shock forms and propagates outward abruptly. The propagation of the shock causes it to re-enter a regime for which $\gamma_{\rm eff}$ is below the critical value. Consequently, the post-shock gas becomes non-supportive again and falls. The oscillatory behaviour of the shock continues with an increasing period until it stabilizes at the largest radius for which the stability criterion is met. The shock never expands beyond the virial radius because shells do not tend to cross there (see Fig. 8 below).

Fig. 4 shows the evolution of the total mass interior to the characteristic radii in the adiabatic and radiative simulations of Figs 2 and 3. In the adiabatic case, the shock mass practically coincides with the virial mass, and the gas never forms a disc. In the radiative case, the shock is initially at the disc radius, and only after 3.9 Gyr does it start propagating outwards.

### 4.2 Testing the stability criterion

We first test the assumption made in Section 2.2 that the post-shock gas has a homologous velocity profile in the vicinity of the shock. In Figs 2 and 3 we see that the flow lines beneath the shock tend to be nearly parallel in the $\log r$–$t$ plane. This means that $\mathrm{d}\log r/\mathrm{d}t = u/r \simeq$ constant, namely homology, equation (12). A similar behaviour has been found in all the simulations that we have performed.

Next, we use the simulations to test the validity of the stability criterion derived in Section 2, equation (29), where $\gamma_{\rm eff}$ is given by equation (27). Recall that $\gamma_{\rm eff}$ is expressed in terms of the pre-shock quantities. In order to map the value of $\gamma_{\rm eff}$ in the different regions of the free-falling gas in the forming halo, we ran the same simulation as in the previous section except that the cooling rate was set to be very high, such that the virial shock never develops. Fig. 5, top panel shows the flow lines in this case, on which four contours of equal $\gamma_{\rm eff}$ values are overlaid, evaluated via equation (27) with equation (30) and the cooling rate from Sutherland & Dopita (1993). As shells are falling into the halo, their $\gamma_{\rm eff}$ is gradually increasing. Also, as time progresses, the value of $\gamma_{\rm eff}$ at the same radius is increasing. By following the value of $\gamma_{\rm eff}$ just above the 'disc' radius (shown as the break at the bottom of the plot), in comparison with the critical value for stability $\gamma_{\rm crit}$, we can therefore use our model to predict when we expect the virial shock to form. This is shown in the middle panel of Fig. 5. We see that at early times (and smaller masses) we have $\gamma_{\rm eff} < \gamma_{\rm crit}$, predicting no stable shock. The system is predicted to

[2] This can be derived from the top-hat formalism of Appendix B, once the final radius is assumed to be fixed at half the maximum expansion radius but the overdensity is evaluated at the time when the top-hat sphere would have collapsed to a singularity.

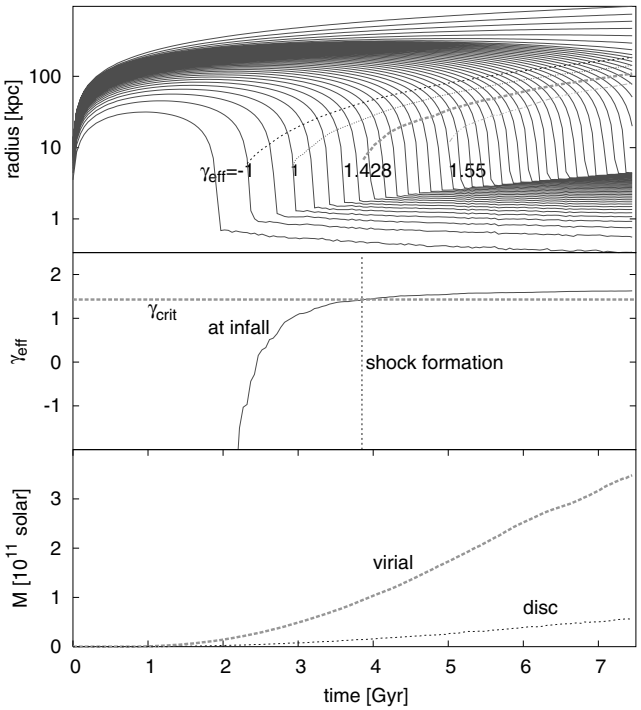


**Figure 5.** Model predictions for shock stability. (This figure is available in colour in the on-line version of the journal on *Synergy*.) Top: the flow lines describe a simulation similar to that shown in Fig. 3, except that the cooling rate has been set to be unrealistically high such that the shock cannot develop. The 'disc' radius is marked by the break in the flow lines at the bottom. Shown on top are four contours of equal $\gamma_{\rm eff}$ values (dashed lines, for $\gamma_{\rm eff} = -1.0$, 1.0, 1.428, 1.55). They are computed by the model, (equation 27), from the local ('pre-shock') gas quantities corresponding to the flow lines at the background, but assuming realistic cooling in the computation of $\gamma_{\rm eff}$. The long-dashed (green in the on-line version) curve corresponds to $\gamma_{\rm eff} = \gamma_{\rm crit} \simeq 1.428$, above which a shock would have formed under realistic cooling. The negative value, $\gamma_{\rm eff} = -1$, is not physical. Middle: time evolution of $\gamma_{\rm eff}$ at infall just above the 'disc' radius (solid) in comparison with the model critical value for stability $\gamma_{\rm crit} = 1.428$. Since $\gamma_{\rm eff}$ is monotonically increasing with decreasing radius, the shock is expected to form first at the 'disc' radius. Shock formation is predicted by the model at $t \simeq 3.9$ Gyr. Bottom: evolution of the characteristic masses in the simulation with the unrealistically high cooling rate. The virial mass at the time of shock formation is predicted to be approximately $10^{11}$ $M_\odot$.

enter the stable-shock regime at approximately $t = 3.9$ Gyr, where $\gamma_{\rm eff}$ becomes larger than $\gamma_{\rm crit}$. A comparison with the realistic radiative simulation described in Fig. 3 yields that this model prediction is very accurate: the shock indeed starts forming at $t = 3.9$ Gyr, and is globally stable thereafter.

Fig. 6 shows the actual evolution of $\gamma_{\rm eff}$ at either the 'disc' radius or the shock radius, whichever is larger, as computed directly from the pre-shock quantities in the simulation with realistic cooling. Shown at the same times are the characteristic masses, already shown in Fig. 4. The virial shock is first generated at time $t = 3.9$ Gyr, when the virial mass is $10^{11}$ $M_\odot$. Starting at this time, the shock is propagating outwards very rapidly. As a result of this fast expansion, the $\gamma_{\rm eff}$ of the pre-shock infalling matter at the shock, which is decreasing with $r$, drops below the threshold. This makes the shock lose its pressure support, it becomes temporarily unstable and its expansion slows down until it is eventually swept back on a dynamic time-scale. The associated drop in total mass behind the shock, seen around $t = 5.8$ Gyr, is due to the fact that the dark

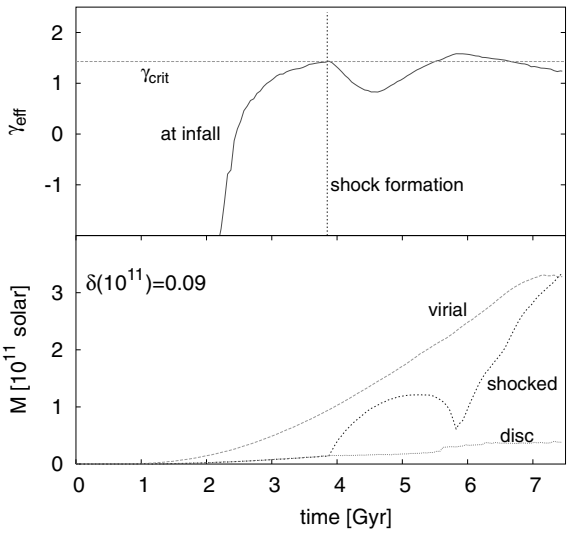


**Figure 6.** Time evolution of $\gamma_{\rm eff}$ at the 'disc' radius or the shock radius, whichever is larger (top), and the associated characteristic masses (bottom), in the simulation with realistic cooling of Fig. 3. The shock forms for the first time when $\gamma_{\rm eff}$ becomes larger than $\gamma_{\rm crit} = \frac{10}{7}$. The shock then oscillates while it approaches a steady state near the virial radius. This figure is available in colour in the on-line version of the journal on *Synergy*.

matter is not swept back with the gas. Once the shock is shrunk to a low enough radius, $\gamma_{\rm eff}$ rises again to above $\gamma_{\rm crit}$; the shock becomes stable again and it resumes its associated expansion towards the virial radius. After the conditions for the shock stability are first met, the shock is visible most of the time. In the rest of this paper, we treat haloes at this state as ones containing a virial shock.

Fig. 7 presents results similar to Fig. 6 for two other simulations with different initial overdensities, and therefore different masses collapsing at different times. The small difference seen in one case between the $\gamma_{\rm eff}$ at which the shock actually forms and the predicted $\gamma_{\rm crit}$ may be totally due to numerical inaccuracies in the simulation. Such inaccuracies may occur when a dark matter shell crosses a gaseous shell, which, near the threshold, may lead to a slightly premature shock formation. This is seen in the convergence test of our code described in Appendix A, when we compare the shock formation times in Table A2. Thus, the model stability criterion is found to be valid within the accuracy of the simulations in all the cases studied, indicating that the model is not limited to a special range of masses and collapse times.

## 5 SHOCK STABILITY IN COSMOLOGY

The analysis of Section 2 thus provides a successful criterion for shock stability, equation (29), as a function of the pre-shock properties of the infalling gas at radius $r$: the density, velocity and metallicity. In order to apply this criterion to a given protogalaxy in a cosmological background, we wish to evaluate the gas density and velocity just before it hits the disc, for a gas shell initially encompassing a total mass $M$ that virializes at redshift $z_{\rm v}$. In this calculation we assume an Einstein–de Sitter cosmology, as a sensible approximation at $z > 2$ (where $\Omega_{\rm m} > 0.9$).

We assume a given universal baryonic fraction $f_{\rm b}$, and a global spin parameter $\lambda$, which determines the ratio of disc to virial radius. The initial mean density perturbation profile, $\bar{\delta}_{\rm i}(M)$, is given at some fiducial time in the linear regime; it is the average profile

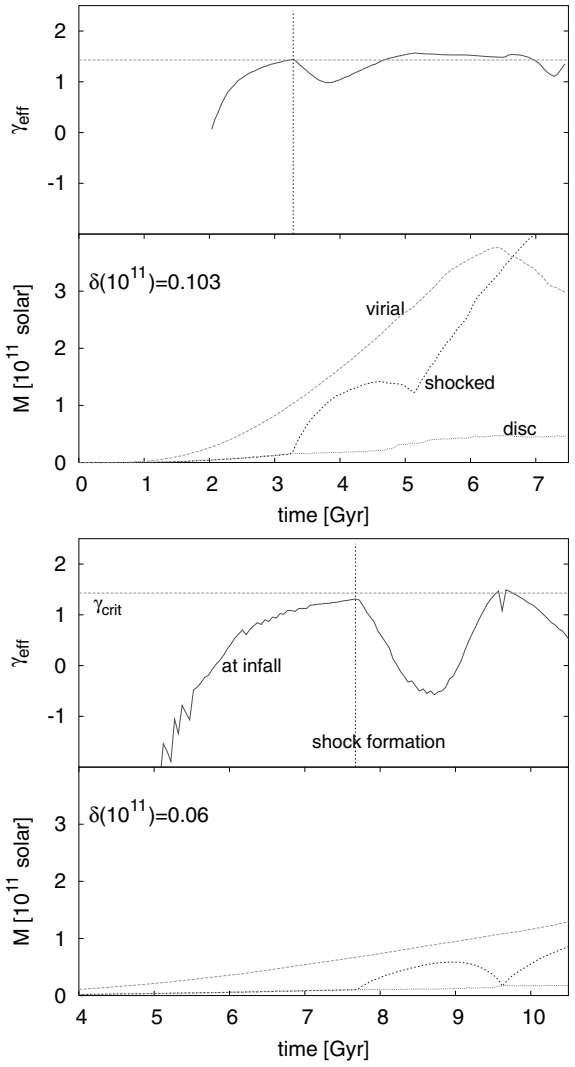


**Figure 7.** Same as in Fig. 6 for two other initial conditions. In all three cases the cosmology is ΛCDM and the initial fluctuation density profile is proportional to the correlation function, but with different amplitudes, as indicated at $z = 100$ within the sphere enclosing $10^{11}$ $M_\odot$. Note the different time-scale in the bottom panel. The qualitative behaviour of the shock is similar in the three cases, and so is the success of the model in predicting shock stability. This figure is available in colour in the on-line version of the journal on *Synergy*.

derived from the power spectrum of initial density fluctuations, as described in Appendix C. In the cosmological toy model used here we approximate the power spectrum as a power law, $P_k \propto k^n$, where $n \simeq -2.4$ to mimic the ΛCDM power spectrum on galactic scales.

We follow gas shells from the initial perturbation until they approach the disc using a two-stage model. During the expansion, turnaround, and until an assumed *virialization* at half the maximum expansion radius, we assume no shell crossing, the total mass interior to the shell remains constant in time, and we follow the radii, density and velocity of the shell via the spherical top-hat model (see Appendix B). From the virial radius inwards we assume that the gas shells, which do not cross each other, contract inside the fixed potential well of an isothermal dark matter halo. This idealized model involves several crude approximations, such as the instantaneous transition at the virial radius, and neglecting the effect of the angular momenta of the individual gas particles at small radii, but we show using spherical simulations that this model predicts the minimum halo mass for which a stable shock first appears to an accuracy of better than 25 per cent. This allows us to use the model for exploring the critical mass as a function of cosmological parameters such as galaxy formation time, metallicity, spin parameter, fluctuation power spectrum and baryonic fraction.

### 5.1 Toy model until virialization

For a given shell $M$ and initial mean perturbation profile $\delta(M)$ [standing for the $\bar{\delta}_{\rm i}(M)$ of Appendix B], the top-hat model (equations B8 and B2) yields the implicit solution

$$r(M, \eta) = r_{\rm v}(M)(1 - \cos\eta), \tag{42}$$

$$t(M, \eta) = t_{\rm v}(M)(\eta - \sin\eta), \tag{43}$$

where the mass dependence enters via the virial quantities

$$r_{\rm v} = C_{\rm i} M^{1/3} \delta(M)^{-1}, \tag{44}$$

$$t_{\rm v} = r_{\rm v}/v_{\rm v} = C_{\rm i}^{3/2} G^{-1/2} \delta(M)^{-3/2}, \tag{45}$$

with $v_{\rm v}^2 = GM/r_{\rm v}$. The coefficient $C_{\rm i} = (6/\pi)^{1/3}(0.15/\rho_{\rm ui})$ is determined by $\rho_{\rm ui}$, the cosmological density at the initial time when $\delta(M)$ is given, independent of $M$.

The velocity of the shell $M$ is

$$u = \frac{\partial r/\partial\eta}{\partial t/\partial\eta} = v_{\rm v}\frac{\sin\eta}{(1-\cos\eta)}. \tag{46}$$

At virialization, $\eta = 3\pi/2$, it is simply $u = -v_{\rm v}$.

In order to evaluate the local density, we follow the radii of two adjacent shells, encompassing masses $M$ and $M + {\rm d}M$, respectively, at a given time $t$, e.g. the time when shell $M$ virializes (at half its maximum expansion radius). Let $\eta$ correspond to shell $M$ at that time, and $\eta + {\rm d}\eta$ to shell $M + {\rm d}M$. In order to express ${\rm d}\eta$ in terms of ${\rm d}M$ we use the fact that the time $t$ is the same for the two shells: $0 = {\rm d}t = (\partial t/\partial M)\,{\rm d}M + (\partial t/\partial\eta)\,{\rm d}\eta$. Using equation (43) this gives

$${\rm d}\eta = \frac{3}{2}\frac{(\eta - \sin\eta)}{(1-\cos\eta)}\frac{\delta'}{\delta}\,{\rm d}M, \tag{47}$$

where we denote $\delta' \equiv {\rm d}\delta/{\rm d}M$. Expressing ${\rm d}r$ in terms of ${\rm d}M$ and ${\rm d}\eta$ based on equation (42), we obtain using equation (47) and after some algebra

$$\frac{{\rm d}r}{r} = \frac{1}{3}\frac{{\rm d}M}{M}\left\{1 - \frac{3M\delta'}{\delta}\left[1 - \frac{3}{2}\frac{\sin\eta(\eta-\sin\eta)}{(1-\cos\eta)^2}\right]\right\}. \tag{48}$$

At virialization of shell $M$, $\eta = 3\pi/2$, the quantity in square brackets equals $(10 + 9\pi)/4$. Not surprisingly, if the initial perturbation is of uniform density, $\delta' = 0$, we are left with ${\rm d}r/r = \frac{1}{3}({\rm d}M/M)$, the straightforward result of $M \propto r^3$. Recall that the virial radius $r_{\rm v}$ of shell $M$ can be obtained either from the universal density at the time of virialization using equation (B4), or from the initial perturbation using equation (B8). The desired local density $\rho$ can be obtained from equation (48) via ${\rm d}M = 4\pi r^2 \rho\,{\rm d}r$.

If the initial perturbation profile is a power law, $\delta(r) \propto r^{-(n+3)}$, using $M \propto r^3$ we have $3M\delta'/\delta = -(n+3)$. So finally

$$\frac{{\rm d}r}{r_{\rm v}} = \frac{1}{3}\frac{{\rm d}M}{M}\left[1 + (n+3)\frac{(10+9\pi)}{4}\right]. \tag{49}$$

Equation (48) (or equation 49 in the power-law case) allows us to compute the desired radii of the two adjacent shells at the time of virialization of shell $M$.

### 5.2 Toy model after virialization

Given the radii of the two adjacent shells at $r_v$, we enter the shell-crossing regime and continue to follow the shells down to the disc radius by numerical integration. The shell radius $r$ and velocity $u$ are related via energy conservation. Assuming that the gas shells contract without crossing each other inside a dark matter halo that is a fixed isothermal sphere, the total mass interior to the shell that originally encompassed a total mass $M$ is

$$M(r) = f_b M + (1 - f_b)\frac{M}{r_v} r, \tag{50}$$

and the gravitational potential at $r$ is

$$\phi(r) = -\frac{GM(r)}{r} - (1 - f_b)\frac{GM}{r_v} \ln\left(\frac{r_v}{r}\right). \tag{51}$$

The integration is performed by advancing $r$ according to the velocity $u$ and then recalculating $u$ according to energy conservation:

$$\frac{1}{2}u^2 + \phi(r) = \frac{1}{2}v_v^2 + \phi(r_v) = \text{constant}. \tag{52}$$

We follow shell $M$ for the time it falls from $r = r_v$ to the disc radius $r = \lambda r_v$, and shell $M + dM$ for the same time interval. Denoting the separation between the shells at the end of this time interval by $dr$, we compute the desired gas density by

$$\rho = \frac{f_b}{4\pi r^2}\frac{dM}{dr}. \tag{53}$$

The resultant values of $r$, $u$ and $\rho$ are inserted into equation (28) in order to obtain an approximation for $\gamma_{\rm eff}$ and then to evaluate stability by equation (24). This allows us to check stability for the case where mass $M$ virializes at redshift $z_v$, with metallicity $Z$, spin parameter $\lambda$, baryonic fraction $f_b$ and a given power spectrum.

### 5.3 Model versus simulations

In Figs 8 and 9 we compare the evolution of the quantities of a given gas shell according to the toy model described in the previous subsections and according to the spherical hydrodynamics simulation described in the earlier sections. We follow a specific shell that hits the disc at approximately $t \simeq 3.8$ Gyr, just before the shock starts propagating into the halo (see Fig. 3). The quantities shown as a function of radius $r$ are total mass $M$ interior to $r$, radial velocity $u$, gas density $\rho$ and the corresponding value of $\gamma_{\rm eff}$. For $M$ and $u$ the evolution starts at the top left-hand corner and ends at the bottom left, while for $\rho$ the upper part of the curve corresponds to the expansion phase and the lower part to the contraction phase. The evolution of $\gamma$ is followed only during part of the contraction phase.

In Fig. 8 we calibrate the toy model to match the simulation at the maximum expansion radius. We see that while the mass interior to the shell is reproduced by the model only to a limited accuracy in the final stages of the collapse, the velocity, density and the resulting value of $\gamma_{\rm eff}$ are recovered very well by the model. This allows us to predict quite accurately the point where $\gamma_{\rm eff}$ exceeds $\gamma_{\rm crit}$.

Since we wish to use the toy model without an exact knowledge of the conditions at maximum expansion, we normalize the model evolution in Fig. 9 based on $M$ and $z_v$.

The slight deviations in $u$ and in $\rho$ now translate into a larger error in $\gamma_{\rm eff}$. The error in the toy model originates mostly from the slight ambiguity in the definition of the virial radius. On one hand we assume it to equal half the maximum expansion radius, and on the other hand we assume it to represent an overdensity of $\Delta_v$ as in equation (41). These two assumptions are not fully consistent

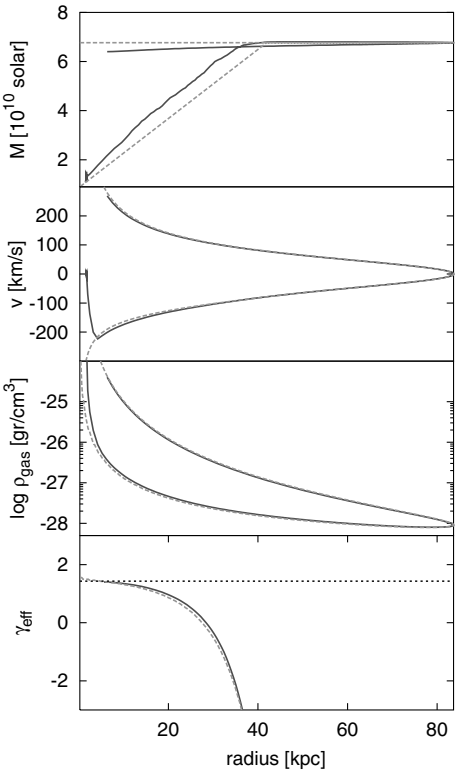


**Figure 8.** Model versus simulation. Evolution of the pre-shock gas properties along a trajectory of one Lagrangian shell as a function of the radius. The solid (red) curve corresponds to the smallest shell from the simulation of Fig. 3 for which the shock formed. The dashed (green) curve is the prediction of the toy model, calibrated here in an ideal way to match the simulation at maximum expansion. The bottom panel shows the $\gamma_{\rm eff}$ derived from the above quantities for the model and the simulation, in comparison with the critical value of $\gamma_{\rm eff}$ marked by the horizontal line. We note that indeed $\gamma_{\rm eff} = \gamma_{\rm crit}$ as $r \to 0$. The fit between model and simulation is remarkable. This figure is available in colour in the on-line version of the journal on *Synergy*.

with the actual behaviour of the virializing system in the simulation. Nevertheless, we see below that our approximate model allows us to estimate the critical halo mass below which the shock does not form to an accuracy of better than 25 per cent, which is quite satisfactory for our purpose here.

### 5.4 Critical mass for shock formation

Fig. 10 shows for several different cases the critical halo mass, below which a shock does not propagate into the halo, versus the redshift at which this critical mass virializes. For each case we compare the model prediction to the shock formation as actually seen in the simulation. The cases differ by the mean metallicity, $Z = 0$ and 0.05 for the lower and upper sets of points, respectively, and by the amplitude of the initial perturbation, corresponding to a range of shock-formation redshifts at every given $Z$. The assumed baryonic fraction is always $f_b = 0.13$, but the assumed spin parameter may be different for the different shells in a given simulation because we set it for each shell such that the final disc has an exponential surface density profile. However, the $\lambda$ values vary in the range 0.02–0.05, compatible with the distribution of spin parameter in cosmological simulations (Bullock et al. 2001). We see that the model predicts the critical mass with an accuracy of better than 25 per cent, such that we can use it for mapping the parameter space in more detail.

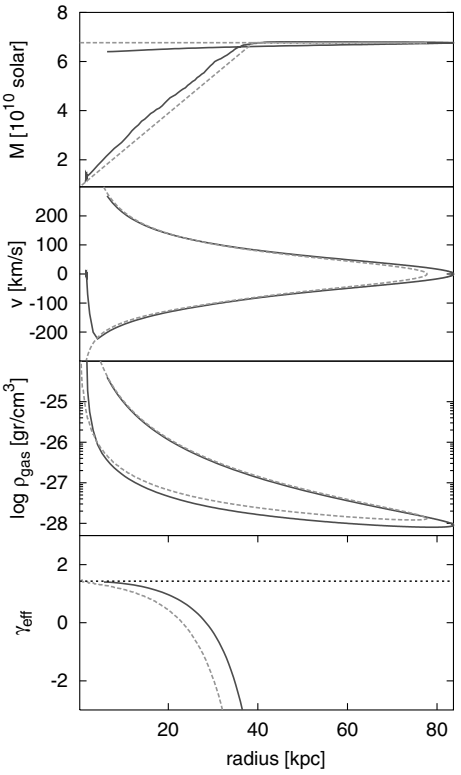


**Figure 9.** Model versus simulation. Same as in Fig. 8, except that the model was calibrated in a practical way based on $M$ and $z_{\rm v}$ without reference to the simulation results. The model predictions now deviate somewhat from the simulation results, but the deviation amounts to less than 25 per cent in the estimate of the critical mass. This figure is available in colour in the on-line version of the journal on *Synergy*.

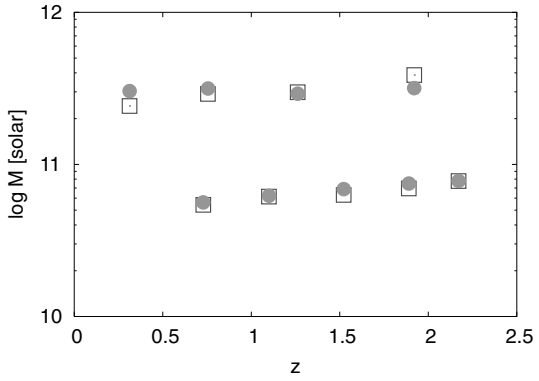


**Figure 10.** Model (green circles) versus simulations (red squares). Displayed is the critical halo mass for shock formation versus the virialization redshift. In all cases $f_{\rm b} = 0.13$, $n = -2.4$, and $0.02 < \lambda < 0.05$. The high set of points is for $Z = 0.05$ and the low set is for $Z = 0$. This figure is available in colour in the on-line version of the journal on *Synergy*.

Fig. 11 shows for several different choices of parameters the critical halo mass for shock formation versus the halo virialization redshift as predicted by the model. A virial shock does not form in haloes of masses below the line. The lines are not always monotonic due to the non-monotonic features in the cooling curves (Fig. 1). Shown in comparison is $M_\star$, the characteristic mass for haloes forming at $z$ according to the Press–Schechter approximation (Lacey & Cole

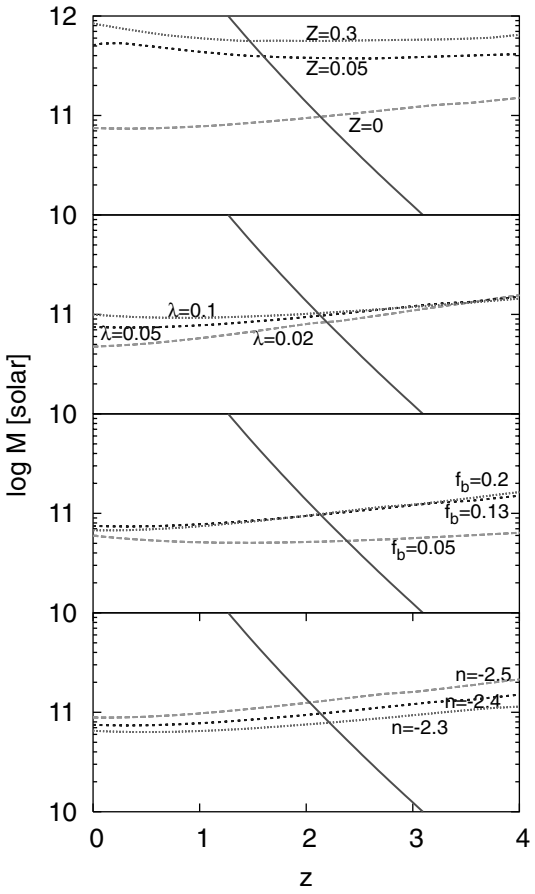


**Figure 11.** Critical mass for shock formation as a function of virialization redshift, according to the model tested against simulations in Fig. 10 (dashed curves). Shown for comparison is the characteristic Press–Schechter $M_\star$ as a function of $z$ for ΛCDM (solid curve). The default values of the parameters are: $f_{\rm b} = 0.13$, $\lambda = 0.05$, $Z = 0$ and $n = -2.4$. One parameter is varied in each panel as indicated. This figure is available in colour in the on-line version of the journal on *Synergy*.

1993). The default values of the parameters, used unless specified otherwise, are $Z = 0$, $\lambda = 0.05$, $f_{\rm b} = 0.13$ and $n = -2.4$.

The upper panel has the metallicity varying from $Z = 0$ to 0.3. The critical mass tends to be higher at higher redshifts (especially for $Z = 0$) because the higher density implies more efficient cooling. It is striking that even for the case of zero metallicity, for which the cooling is not at its maximum efficiency, an $M_\star$ halo cannot produce a shock until a relatively late redshift, $z \simeq 2.1$. The addition of a small amount of metal, $Z = 0.05$, increases the cooling rate significantly (see Fig. 1) such that $M_\star$ haloes start producing virial shocks only after $z \simeq 1.6$.

The second panel has $\lambda$ varying as marked. The shock forms slightly earlier if the disc is smaller (lower $\lambda$), because the conditions become more favourable for shock formation closer to the centre. At high redshifts the increase in infall velocity happens to balance out the increase in density, temperature and cooling rate. The post-shock temperature there is a few $10^6$ K. The bottom line is that the critical mass is not too sensitive to $\lambda$.

The third panel has $f_{\rm b}$ varying as marked. The critical mass is monotonic with the baryonic fraction because the cooling rate is monotonic with gas density. The parameter $f_{\rm b}$ can be interpreted as the fraction of the baryons that actually take part in the shock formation. This can be smaller than the universal baryonic fraction

if some of the gas falls into the halo in the form of dense clumps. Even with $f_b$ as low as 0.05, meaning that most of the gas is not participating in the cooling, an $M_\star$ halo would not produce a shock until $z \simeq 2.4$. The conclusion is that the critical mass is not too sensitive to $f_b$ either.

The bottom panel explores three values for the initial power index $n$ approximating the power spectrum of ΛCDM on galactic scales. The dependence of the critical mass on $n$ in this regime is weak.

## 6 DISCUSSION

The heating of the gas behind a virial shock in haloes has been a basic component in galaxy formation theory (Rees & Ostriker 1977). We studied the conditions for the existence of such a virial shock in spherical haloes. We first pursued an analytic stability analysis in the presence of cooling, and then demonstrated its validity using high-resolution spherical hydrodynamical simulations. The obtained criterion for shock stability in terms of the post-shock quantities is

$$\gamma_{\rm eff} \equiv \frac{{\rm d}\ln P/{\rm d}t}{{\rm d}\ln \rho/{\rm d}t} > \frac{10}{7}. \tag{54}$$

In terms of the pre-shock gas properties, this condition reads

$$\frac{\rho_0 r_s \Lambda(T_1)}{|u_0|^3} < 0.0126, \tag{55}$$

where $\rho_0$ and $u_0$ are the gas density and infall velocity in front of the shock, $r_s$ is the shock radius, $\Lambda(T)$ is the cooling function, which depends on the metallicity $Z$, and $T_1 \propto u_0^2$ is the post-shock temperature as a function of the pre-shock infall velocity.

Based on this criterion, we find that a virial shock forms only in big haloes forming at late redshifts. A virial shock does not form in smaller haloes forming early where the cooling behind the shock efficiently removes its pressure support. For example, we find that most galactic haloes that have collapsed and virialized by $z \sim 2$ did not produce a virial shock. Haloes less massive than $\sim 10^{11}$ M$_\odot$ never produce a shock even if the gas is of zero metallicity. If the metallicity is non-negligible (e.g. $Z \sim 0.05$), this lower bound to shock formation rises to $\sim 7 \times 10^{11}$ M$_\odot$. When a shock does not exist, the gas is not heated to the halo virial temperature until it falls all the way to the disc at the inner halo.

Forcada-Miro & White (1997), in an unpublished work, have pursued independently a numerical analysis along similar lines, involving a more detailed treatment of the cooling processes involved. They also find that the virial shock radius is significantly reduced due to the cooling in haloes of small masses, $M < 10^{11}$ M$_\odot$. In their case the shock never completely disappears because of a different feature in their numerical scheme; they put all the cooled post-shock gas in one central 'shell' to avoid numerical difficulties at the centre. This makes the inner boundary of the system follow the shock quite closely in cases where there is efficient cooling behind the shock, and allows the presence of a small-radius shock even in such cases. Overall, our numerical results are in encouraging agreement, and our analytic model provides a natural explanation for their numerical results as well.

The most severe uncertainty when attempting to apply our results to real galaxies arises from the assumed spherical symmetry in both the model and the numerical simulations. The validity of this approximation for the asymmetric halo configurations in the hierarchical clustering scenario is an open question to be addressed in future work. Nevertheless, we note that Katz et al. (2001) and Fardal et al. (2001) observe in their cosmological simulations that a large fraction of the mass accreted on to haloes indeed remains cold and is never heated to the virial temperature. Toft et al. (2002) find in their simulations, using a similar TREESPH code as that used by Katz et al. (2001), that the soft X-ray radiation is mainly emitted from within the innermost 20 kpc of their haloes, well inside the virial radius, in encouraging agreement with our results. On the other hand, it is not obvious that the resolution in these simulations is adequate for studying the shock physics involved; our estimates indicate that three-dimensional simulations with proper resolution are not practical at present (Appendix A).

Another complication may arise from radiative effects. Even when there is no virial shock, the kinetic energy of the gas eventually turns into radiation when the gas infall motion is brought to a halt at the disc. At such densities, the width of the shock front is much smaller than the width of the cooling front behind the shock, $\sim 10^{-2}$ pc versus $\sim 10^2$ pc. Thus, the gas in a thin shell behind the shock is heated to a temperature corresponding to its kinetic energy, and it cools by radiating soft X-rays. The X-ray radiation is expected to generate an ionized H II bubble, in which the ionization rate balances the recombination rate. The Strömgren radius of this bubble is relatively small, of the order of a few kiloparsec, because the high gas density implies a high recombination rate. The recombination process then generates a flux of L$\alpha$ radiation, emitted at the inner few kpc of the halo.

A naive inspection of cross-sections might indicate that the L$\alpha$ radiation would be trapped inside the halo. This could, in principle, affect the shock stability in three different ways: by increasing the radiation pressure, by heating up the infalling matter and by slowing down the radiative cooling responsible for the shock instability. It has been argued by Rees & Ostriker (1977) that the radiation pressure at these low temperatures must be insignificant compared with the gas pressure even if all the internal energy was drained from the baryons into the radiation field. One might add that since the radiation pressure behaves like a $\gamma = \frac{4}{3}$ gas, it could at most make the system marginally stable. When work is performed on the radiation field, any leakage of radiation out would turn the energy into cooling rather than $P\,{\rm d}V$, and will thus reduce the effective gamma, making the system unstable.

Partial heating of the infalling gas should not affect our analysis as long as the temperature of the infalling matter is significantly below the virial (post-shock) temperature such that the strong shock approximation remains valid. The effect of the reduced cooling rate is yet to be investigated. In practice, we do not expect the radiation trapping to be very efficient, because the effective opacity is reduced by thermal broadening and by the systematic blueshift due to the gas infall motion. When the opacity is high, the radiation heats up the gas, which enhances the thermal broadening and the collisional ionization rate. This reduces the opacity and allows for radiation escape. The system is likely to reach a steady state in which it gradually cools. This process is under current investigation.

Feedback effects may further complicate the picture and affect shock stability. The energy fed back to the gas from stars, supernovae and active galactic nuclei (AGN) may heat the halo gas and expel part of it. Merging substructures may have additional complicated effects. These effects cannot be captured by our idealized spherical analysis, and a proper study would probably require high-resolution three-dimensional hydrodynamical simulations. While observations and certain theoretical considerations indicate that feedback effects are likely to be important in galaxies as large as $\sim 10^{11}$ M$_\odot$ and may thus affect the shock stability (e.g. Dekel & Silk 1986; Dekel & Woo 2003 and references therein), it has proven difficult for numerical simulations to reproduce such effects in any but small

dwarf galaxies, indicating that feedback effects may not be so crucial for the understanding of shock stability. Until the dust settles on the role of feedback effects, our preliminary conclusions based on the spherical analysis should be taken with a grain of salt.

The general absence of a virial shock might have three direct implications, which we study in associated papers.

First, as explained above, when the gas is heated at the disc rather than near the halo virial radius, the generated X-ray radiation serves to ionize the gas and is not emitted outwards. The result would be a suppression of the X-ray emission in the range $5 \times 10^5$ to $2 \times 10^6$ K. This may help to explain the missing X-ray problem pointed out by Pen (1999) and Benson et al. (2000). Pen (1999) argue that there is an order-of-magnitude discrepancy between the soft X-ray flux as observed by Cui et al. (1996), after subtracting the contribution of quasars, and the predicted flux from haloes constructed by a Press–Schechter hierarchical model under the assumption of shock heating to the virial temperature.

Secondly, the infall energy, via the ionizing X-ray, is efficiently transformed into L$\alpha$ radiation at the inner few kpc of the halo. A related increase in the L$\alpha$ flux has indeed been seen in the cosmological simulations of Fardal et al. (2001). This may explain the observed high flux of L$\alpha$ emitters at high redshift (e.g. Breuck et al. 2000; Pentericci et al. 2000, 2001). Based on the high observed flux and the assumption that the L$\alpha$ is emitted from stars, Pentericci et al. (2000) estimate large masses for the L$\alpha$ emitters, but the much higher flux per unit mass predicted by our model may lead to significantly lower mass estimates. Based on our analysis, most of the L$\alpha$ flux is expected to be emitted from the inner few kpc of the halo, where the gas is at $\sim 10^4$ K. Neglecting line shifts and broadening, the halo might be opaque to L$\alpha$, thus eventually emitting its energy from an outer photosphere where the halo becomes transparent. However, a careful study of the thermal broadening and the systematic redshifts within the halo is required in order to determine whether the system is opaque or transparent to the L$\alpha$ photons. This is a subject of an ongoing investigation.

Finally, the direct collapse of cold gas into the disc may have interesting theoretical consequences that need to be worked out. It may induce an efficient starburst in analogy to the burst originating in the shock between two colliding gas clouds. In turn, the strong inwards flow of gas may prevent an efficient gas removal by supernova-driven winds. In particular, current cosmological semi-analytic models (SAM) of galaxy formation (Kauffmann, White & Guiderdoni 1993; Cole et al. 1994; Kauffmann et al. 1999; Somerville & Primack 1999; Maller et al. 2001 and related works) use the standard picture of heating behind a virial shock in their modelling. This has strong effects on the disc formation rate, star formation rate, feedback, etc. Other semi-analytic models (White & Frenk 1991; Efstathiou 2000) also appeal to the slow gas infall rate as a mechanism that regulates the gas input into the disc. Since the cooling time for a $10^{11}$-$M_\odot$ halo is relatively short, the SAM predictions for such haloes may be only slightly affected by the inhibition of heating. However, given some metal enrichment, no heating is expected for haloes as massive as $\sim 7 \times 10^{11}$ $M_\odot$, for which the cooling time is longer, and the effect on the SAM predictions may be more severe. Shocks, when present, are also expected to alter the properties of the gas; for example, extinct dust particles. These effects can change SAMs that incorporate dust extinction.

## ACKNOWLEDGMENTS

We acknowledge advice from Z. Barkat and E. Livne, J. Ostriker and stimulating discussions with S. Balberg, E. Bertschinger, T. Broadhurst, D. Gazit, Y. Hoffman, W. Mathews, A. Nusser, N. Shaviv and S. D. M. White. This research has been supported by the Israel Science Foundation grant 213/02, by the German–Israel Science Foundation grant I-629-62.14/1999. and by NASA ATP grant NAG5-8218.

## REFERENCES

Bardeen J.M., Bond J.R., Kaiser N., Szalay A.S., 1986, ApJ, 304, 15
Benson A.J., Bower R.G., Frenk C.S., White S.D.M., 2000, MNRAS, 314, 557
Bertschinger E., 1985a, ApJS, 58, 1
Bertschinger E., 1985b, ApJS, 58, 39
Bryan G.L., Norman M.L., 1998, ApJ, 495, 80
Bullock J.S., Dekel A., Kolatt T.S., Kravtsov A.V., Klypin A.A., Porciani C., Primack J.R., 2001, ApJ, 555, 240
Cole S., Aragón-Salamanca A., Frenk C., Navarro J., Zepf S., 1994, MNRAS, 271, 781
Chi W., Sanders W.T., McCammon D., Snowden S.L., Womble D.S., 1996, ApJ, 486, 117
Cox J.P., 1980, Theory of Stellar Pulsation. Princeton Univ. Press, Princeton
De Breuck C., Rottgering H., Miley G., van Breugel W., Best P., 2000, A&A, 362, 519
Dekel A., 1981, A&A, 101, 79
Dekel A., Silk J., 1986, ApJ, 303, 39
Dekel A., Woo J., 2003, MNRAS, in press (astro-ph/0210454)
Efstathiou G., 2000, MNRAS, 317, 3, 697
Efstathiou G., Bond J.R., White S.D.M., 1992, MNRAS, 258, 1P
M.A., Katz N., Gardner J.P., Hernquist L., Weinberg D.H., Davé R., 2001, ApJ, 562, 605
Forcada-Miro M.I., White S.D.M., 1997, astro-ph/9712204 (unpublished)
Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 1998, MNRAS, 300, 146
Katz N., Keres D., Davé R., Weinberg D.H., 2001, preprint (astro-ph/209279)
Kauffmann G., White S.D.M., Guiderdoni B., 1993, MNRAS, 264, 201
Kauffmann G., Colberg J.M., Diaferio A., White S.D.M., 1999, MNRAS, 303, 1, 188
Lacey C., Cole S., 1993, MNRAS, 262, 627
Maller A.H., Prochaska J.X., Somerville R.S., Primack J.E., 2001, MNRAS, 326, 4, 1475
Peebles P.J.E., 1993, Principles of Physical Cosmology. Princeton Univ. Press, Princeton
Pen U.L., 1999, ApJ, 510, 2, L1
Pentericci L. et al., 2000, A&A, 361, 2, L25
Pentericci L., Kurk J.D., Rottgering H.J.A., Miley G.K., Venemans B.P., 2001, ASP Conf. Ser. Astron. Soc. Pac., San Francisco (astro-ph-0110223)
Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., 1997, Numerical Recipes is Fortran 77. Cambridge Univ. Press, Cambridge
Rees M.J., Ostriker J.P., 1977, MNRAS, 179, 541
Somerville R.S., Primack J.R., 1999, MNRAS, 310, 1087
Sutherland R., Dopita M., 1993, ApJS, 88, 253
Toft S., Rasmussen J., Sommer-Larsen J., Pedersen K., 2002, MNRAS, 335, 799
White S.D.M., Frenk C., 1991, ApJ, 379, 52
Zel'dovich Ya. B., Raiser Yu. P., 1966, Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena. Academic, New York

## APPENDIX A: TESTING THE HYDRODYNAMICS CODE

The numerical code HYDRA has been developed specifically for simulating the evolution of a single spherical halo through collapse and feedback processes. A proper computation of the cooling and shock formation requires high precision. In this appendix we describe a few of the tests performed in order to verify that the code works properly. In the following three subsections we test for energy

conservation, spatial convergence and the performance of the code in a self-similar case.

### A1 Energy conservation

Our numerical scheme does not use the total energy equation in the integration of the partial differential equations. Furthermore, the total energy of the system is not a straightforward sum of other variables that are involved in the calculation. The requirement of energy conservation is therefore an independent test for the accuracy of the numerical scheme. Energy conservation is harder to achieve than spatial convergence for several reasons. First, the error in total energy is systematic, in the sense that when dark matter shells cross each other the energy tends to increase. Secondly, since our system is only marginally bound, the total energy is a small difference between two large quantities. We note that energy conservation is simpler to achieve when there is no cooling or when dark matter is absent (and thus there is no shell crossing).

The total energy of the system at time $t$ is the sum of the terms

$$E = K_{\rm d} + T_{\rm d} + K_{\rm g} + T_{\rm g} + U + Q, \tag{A1}$$

where the subscripts d and g refer to dark matter and gas, respectively, $K$ denotes kinetic energy, $T$ denotes potential energy, $U$ is the gas internal energy and $Q$ is the thermal energy lost to radiation by time $t$.

For the dark matter, these are straightforward sums over the discrete dark matter shells:

$$K_{\rm d} = \Delta M_i \frac{1}{2} \sum_{i=1}^{n_{\rm d}} \left( v_i^2 + \frac{j_i^2}{r_i^2} \right), \tag{A2}$$

$$T_{\rm d} = -\sum_{i=1}^{n_{\rm d}} \frac{G \Delta M \mathcal{M}_i}{r_i + a}, \tag{A3}$$

where $\mathcal{M}_i$ is the total mass interior to dark matter shell $i$, as defined in Section 3.

For the gas shells, recall that the quantities $r$, $v$ and $j$ are given at the inner- and outer-shell boundaries, $i - 1$ and $i$, respectively, so we compute the shell energies by averaging over the two boundary values:

$$K_{\rm g} = \frac{1}{2} \sum_{i=1}^{n_{\rm g}} \Delta m_i \left( \frac{v_i r_i^3 + v_{i-1} r_{i-1}^3}{r_i^3 + r_{i-1}^3} \right)^2 + \frac{1}{2} \sum_{i=1}^{n_{\rm g}} \Delta m_i \left( \frac{j_i r_i^2 + j_{i-1} r_{i-1}^2}{r_i^3 + r_{i-1}^3} \right)^2, \tag{A4}$$

$$T_{\rm g} = -\sum_{i=1}^{n_{\rm g}} \frac{G \Delta m_i (\mathcal{M}_i + \mathcal{M}_{i-1})/2}{\left[ \left( r_i^3 + r_{i-1}^3 \right)/2 \right]^{1/3} + a}. \tag{A5}$$

The internal energy is a straightforward sum

$$U = \sum e_i \Delta m_i. \tag{A6}$$

The energy radiated away, $Q = \int {\rm d}t \int q \, {\rm d}m$, is computed by

$$Q_{=} \sum_{j=1}^{n_{\rm t}} \Delta t^j \sum_{i=1}^{n_{\rm g}} q_i^j \Delta m_i, \tag{A7}$$

where $\Delta t^j$ is the length of time-step $j$ and $q_i^j$ is the cooling rate in shell $i$ at time-step $j$ (in units of erg g$^{-1}$ s$^{-1}$).

In a run with 10 000 dark matter shells and 2000 gas shells, we require and obtain energy conservation at the level of 1 per cent in a Hubble time, using a typical Runge–Kutta time-step of approximately $5 \times 10^{-6}$ Gyr. (Such a run takes approximately 10 h on an Alpha-6 DEC processor.)

**Table A1.** Energy conservation.

| Description | $n_{\rm g}$ | $n_{\rm d}$ | $\epsilon_{\rm c}$ | $\epsilon_{\rm rk}$ | $t_{\rm sc}$/ Gyr | $E_{\rm fin}/E_{\rm init}$ |
|---|---|---|---|---|---|---|
| Nominal | 2k | 10k | 0.3 | 0.1 | $10^{-4}$ | 0.992 |
| Small $\epsilon$s | 2k | 10k | 0.1 | 0.01 | $10^{-5}$ | 0.994 |
| More shells | 3k | 15k | 0.3 | 0.1 | $10^{-4}$ | 0.997 |

**Table A2.** Shock formation times and energy conservation.

| $n_{\rm gas}$ | $n_{\rm dark}$ | $E_{\rm final}/E_{\rm initial}$ | Formation of shock (Gyr) |
|---|---|---|---|
| 3000 | 15 000 | 0.997 | 3.9 |
| 2000 | 10 000 | 0.994 | 3.9 |
| 1000 | 5000 | 0.976 | 4.01 |
| 500 | 2500 | 0.923 | 3.76 |
| 250 | 1250 | 0.711 | 3.27 |
| 125 | 625 | 0.252 | 2.82 |

We check the conservation first by varying the accuracy parameters presented in Section 3.3, and then by varying the number of shells. The three cases presented in Table A1 demonstrate that the results converge when the accuracy is increased. The simulations shown in this table are of the standard case with realistic cooling shown in Fig. 3. When cooling is shut off, energy conservation is much better. With the nominal choice of accuracy parameters the final energy is 0.9999 of the initial energy.

### A2 Spatial convergence: 3D versus 1D

A proper treatment of the competition between the pressure increase due to contraction and the pressure decrease due to cooling requires high temporal and spatial resolution. In particular, when the spatial resolution is increased, the shock appears earlier. Table A2 shows results from simulations of the case with realistic cooling (Fig. 3), all with the same accuracy criteria ($\epsilon_{\rm c}$, $\epsilon_{\rm rk}$ and $t_{\rm sc}$), but with different spatial resolutions. The average distance between gas shells near the centre ranges from approximately 80 pc to 2 kpc. With the poorest resolution of 125 gas shells the virial shock appears almost immediately after the virialization of the first shells of the simulation. The energy changed by approximately 75 per cent during this simulation. Even if we assume that the precision of a 3D calculation is as good as that of an analogous 1D calculation (actually smooth particle hydrodynamics codes converge slower than finite-element schemes for problems involving shocks), we still need to cube the number of particles or grid points in order to achieve the same resolution. A three-dimensional simulation with $2 \times 10^6$ gas particles and $2.5 \times 10^8$ dark matter particles, which is close to the limit of what is computationally feasible today, would correspond to the unsatisfactory case with the lowest spatial resolution in Table A2.

### A3 A self-similar case

When the initial conditions are scale-free (unlike the initial conditions assumed in the body of this paper, motivated by ΛCDM), and when the cooling function is also scale-free (unlike the realistic cooling function used above), the results should be self-similar. This can provide a test for the accuracy of our numerical code. We

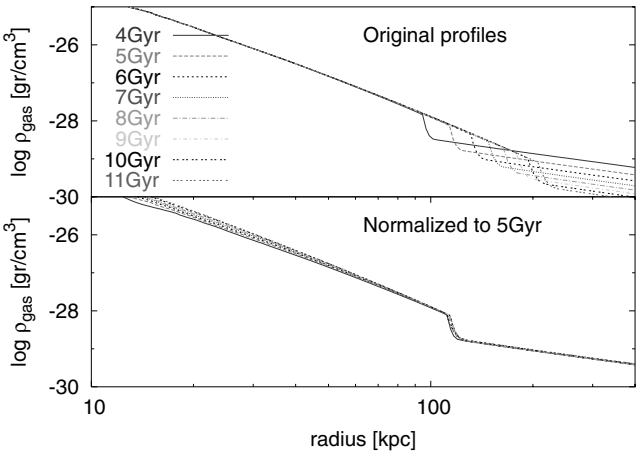


**Figure A1.** Self-similar profiles by our hydrodynamics code with gas only. Top: density profiles at times 4–11 Gyr. Bottom: the same profiles normalized according to the self-similarity law $r \propto t^{8/9}$ to $t = 5$ Gyr.

follow Bertschinger (1985a,b) in using an initial perturbation consisting of a point-mass embedded in a uniform-density background. Far from the point mass, the system should be self-similar. We ran a simulation of such a case using our code with gas only ($f_b = 1$) and no cooling, starting at $z = 200$ with an overdensity of 10 inside the innermost 2 kpc.

The upper panel of Fig. A1 shows the density profile at different times. As expected, a shock appears at every time as a density jump by a factor of 4 [$=(\gamma + 1)/(\gamma - 1)$ for $\gamma = \frac{5}{3}$], and the post-shock gas settles to a complete rest after it is shocked. [The slope of the post-shock density profile is somewhat different from Bertschinger (1985b), because our calculation assumes a ΛCDM cosmology rather than the Einstein–de Sitter assumed by Bertschinger.] The lower panel shows the same profiles after they were scaled to the same time (5 Gyr) according to the scaling relation of Bertschinger (1985b): $r \propto t^{8/9}$. We see that our simulations recover the expected scaling relation almost perfectly.

Fig. A2 shows an analogous test for the case where both gas and dark matter are present, with $f_b = 0.13$. The results are similar except for the somewhat higher noise level caused by the dark matter component.

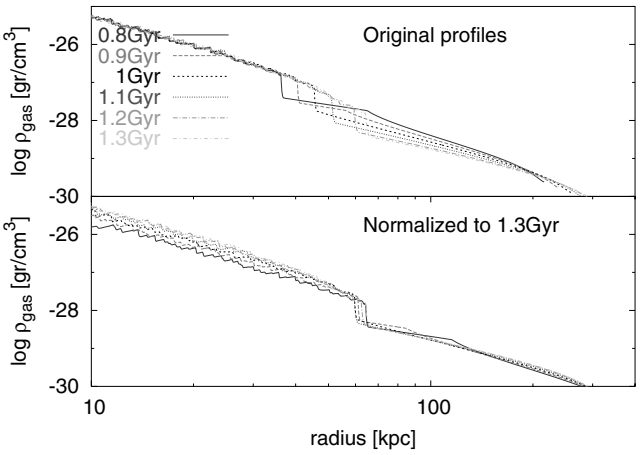


**Figure A2.** Similar to Fig. A1 but with dark matter included, $f_b = 0.13$. The profiles are scales to $t = 1.3$ Gyr.

## APPENDIX B: TOP-HAT MODEL

Consider a bound spherical perturbation encompassing mass $M$, the mean density fluctuation profile of which at some fiducial initial time in the linear regime is $\bar{\delta}_i(M)$, embedded in an Einstein–de Sitter (EdS) cosmological background when the universal expansion factor is $a_i$. We wish to express the shell radius $r$ as a function of time in terms of $M$.

The implicit solution for a closed 'mini-universe', via a conformal time parameter $\eta$ specific to this perturbation, is

$$r = r_v(1 - \cos\eta), \tag{B1}$$

$$t = \frac{r_v}{v_v}(\eta - \sin\eta). \tag{B2}$$

Maximum expansion occurs when $\eta = \pi$, and then a collapse to half this radius, which we identify with *virialization*, is obtained at $\eta = 3\pi/2$, with virial radius $r_v$ and corresponding virial velocity $v_v^2 = GM/r_v$. We normalize the universal expansion factor $a$ by identifying it at the initial time with the shell radius $r$. Assuming $\eta_i \ll 1$, we have $r_i \simeq \frac{1}{2} r_v \eta_i^2$ and $t_i \simeq \frac{1}{6}(r_v/v_v)\eta_i^3$, so $a_i = r_i$ yields $a_i = (9 r_v v_v^2/2)^{1/3} t^{2/3}{}_i$. The EdS expansion factor, $a \propto t^{2/3}$, can now be related using equation (B2) to the $\eta$ of the perturbation at any time:

$$a = (9 r_v v_v^2/2)^{1/3} t^{2/3} = \frac{9}{2}^{1/3} r_v(\eta - \sin\eta)^{2/3}. \tag{B3}$$

The mean density within the perturbation relative to the universal density at the same time becomes a straightforward function of $\eta$:

$$\frac{\bar{\rho}}{\rho_u} = \frac{a^3}{r^3} = \frac{9}{2}\frac{(\eta - \sin\eta)^2}{(1 - \cos\eta)^3}. \tag{B4}$$

This is a standard result of the top-hat model.

In order to relate the density to the small initial perturbation at $\eta_i \ll 1$, we obtain from equation (B4) by a proper Taylor expansion to the first non-vanishing order:

$$\bar{\delta}_i \simeq 0.15\eta_i, \tag{B5}$$

where $\bar{\delta} \equiv \bar{\rho}/\rho_u - 1$ is the mean fluctuation. Using this in the linear term of equation (B1) we obtain

$$a_i = \frac{1}{2} r_v \eta_i^2 = (1/0.3) r_v \bar{\delta}_i. \tag{B6}$$

This allows us to write the mean density at any $\eta$, using equation (B1), as

$$\frac{\bar{\rho}}{\rho_{ui}} = \frac{a_i^3}{r^3} = \frac{\bar{\delta}_i^3}{0.3(1 - \cos\eta)^3}. \tag{B7}$$

Recalling that $\bar{\rho} = M/[(4\pi/3)r^3]$ we finally obtain at any $\eta$

$$r = C_i \frac{M^{1/3}}{\bar{\delta}_i(M)}(1 - \cos\eta), \tag{B8}$$

where

$$C_i \equiv \left(\frac{6}{\pi}\right)^{1/3} \frac{0.15}{\rho_{ui}}. \tag{B9}$$

In particular, at $\eta = 3\pi/2$, we obtain for the virial radius $r_v = C_i M^{1/3}/\bar{\delta}_i(M)$. The constant $C_i$ is independent of $M$; the universal density $\rho_{ui}$ [$=(1 + z_i)^3 \rho_{u0}$] is determined by the choice of the fiducial redshift $z_i$ at which $\bar{\delta}_i(M)$ is given.

## APPENDIX C: INITIAL PROFILE

We adopt in the linear regime the typical density fluctuation profile for the assumed power spectrum of fluctuations. For a Gaussian random field, this profile is proportional to the two-point correlation function (Dekel 1981):

$$\delta_{\rm i}(r) = \delta_{0\rm i}\frac{\xi(r)}{\xi(0)}, \tag{C1}$$

where $\delta_{0\rm i}$ specifies the amplitude normalization. For a given power spectrum $P(k)$, the correlation function is given by (Peebles 1993, equation 21.40)

$$\xi(r) = 4\pi\int_0^\infty k^2\,{\rm d}k\,P(k)\frac{\sin kr}{kr}, \tag{C2}$$

and the local variance is

$$\xi(0) = 4\pi\int_0^\infty k^2\,{\rm d}k\,P(k). \tag{C3}$$

The mean density fluctuation interior to radius $r$, containing mass $M = (4\pi/3)\rho_{\rm ui}r^3$ when the fluctuation is small, is

$$\bar{\delta}_{\rm i}(r) = \frac{3}{r^3}\int_0^r r^2\,{\rm d}r\delta_{\rm i}(r). \tag{C4}$$

This involves the integral (Peebles 1993, equation 21.62)

$$J_3(r) \equiv \int_0^r r^2\,{\rm d}r\xi(r) = \frac{4\pi r^3}{3}\int_0^\infty k^2\,{\rm d}k\,P(k)\tilde{W}_{\rm s}(kr), \tag{C5}$$

where $\tilde{W}_{\rm s}(kr)$ is the Fourier transform of the top-hat window,

$$\tilde{W}_{\rm s}(kr) = 3\left[\frac{\sin kr}{(kr)^3} - \frac{\cos kr}{(kr)^2}\right] = \frac{3}{kr}j_1(kr), \tag{C6}$$

with $j_1$ being the spherical Bessel function.

In the simulations of this paper we use the ΛCDM power spectrum from the fitting formula of Bardeen et al. (1986)

$$P(k) = AkT^2(k), \tag{C7}$$

$$T(k) = \left\{1 + \left[ak/\Gamma + (bk/\Gamma)^{3/2} + (ck/\Gamma)^2\right]^\nu\right\}^{1/\nu} \tag{C8}$$

with $a = 6.4\,h^{-1}$, $b = 3.0\,h^{-1}$, $c = 1.7\,h^{-1}$ Mpc, $\nu = 1.13$ and $\Gamma = 0.21$ (the $\tau$CDM model of Efstathiou, Bond & White 1992). The normalization is such that $\sigma_8 = 1$.

In the cosmological toy model we approximate the ΛCDM by a power-law power spectrum $P_k \propto k^n$, for which the two-point correlation function is also a power law, $\xi(r) \propto r^{-(n+3)}$, and then

$$\bar{\delta}_{\rm i}(r) = \left(\frac{r}{r_1}\right)^{-(n+3)}, \tag{C9}$$

where $r_1$ provides the normalization. In terms of mass we obtain

$$\bar{\delta}_{\rm i}(M) = \left[\frac{M}{(4\pi/3)\rho_{\rm ui}r_1^3}\right]^{-(n+3)/3}. \tag{C10}$$

This serves as the input to equation (B8), or equation (48).

We normalize the initial perturbation such that a specific mass $M_1$ reaches virialization at some cosmological epoch $a_{\rm v} = 1/(1 + z_{\rm v})$. Using equation (B3) and (B6) we obtain the linear analogue to the non-linear fluctuation growth rate:

$$\frac{\bar{\delta}(\eta)}{\bar{\delta}_{\rm i}} \equiv \frac{a(\eta)}{a_{\rm i}} = 0.3\left(\frac{9}{2}\right)^{1/3}(\eta - \sin\eta)^{2/3}. \tag{C11}$$

At virialization, this gives $\delta_{\rm v} \equiv \bar{\delta}(\eta = 3\pi/2) \simeq 1.58$. Then,

$$\bar{\delta}_{\rm i}(M_1) = \delta_{\rm v}\left(\frac{a_{\rm i}}{a_{\rm v}}\right). \tag{C12}$$

The normalization parameter $\delta_{0\rm i}$ (or $r_1$) at $a_{\rm i}$ is obtained by equating this with equation (C4) (or equation C10) at $M = M_1$.

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