Mon. Not. R. Astron. Soc. **267,** 401–412 (1994)

Simulations of dissipative galaxy formation in hierarchically clustering universes – II. Dynamics of the baryonic component in galactic haloes

Julio F. Navarro[1] and Simon D. M. White[2]
[1] *Physics Department, University of Durham, Durham DH1 3LE*
[2] *Institute of Astronomy, Madingley Road, Cambridge CB3 0HA*

Accepted 1993 October 21. Received 1993 October 21; in original form 1993 September 20

ABSTRACT

We present self-consistent 3D simulations of the formation of virialized systems containing both gas and dark matter. Using a fully Lagrangian code based on the smooth particle hydrodynamics technique and a tree data structure, we follow the evolution of regions of comoving radius 2–3 Mpc with proper inclusion of the tidal effects of surrounding material. Initial conditions at high redshifts assume an Einstein–de Sitter universe, a biased cold dark matter perturbation spectrum ($b = 2.5$), and a baryonic mass fraction of 10 per cent. The gas is initially cold and radiates in the manner expected for a plasma of primordial composition. We neglect star formation and associated processes. We find that most of the gas settles rapidly into centrifugally supported discs at the centres of small dark matter clumps. These are not disrupted during later evolution, and survive the merging of their dark haloes. They either remain distinct, or merge without reheating at a later time. These results confirm that significant energy input from non-gravitational sources is required to produce a cold baryonic fraction consistent with the observed abundance of gas and stars in galaxies. In our simulations, much of the final disc of gas accumulates through mergers of dense clumps rather than through the smooth infall envisaged in earlier models. This process involves substantial angular momentum loss and large collapse factors. It produces discs that are more concentrated than real spiral discs and have circular velocities up to 70 per cent greater than that of the surrounding halo. Nevertheless, the surface density profile of these discs can be reasonably well fitted by an exponential, and their rotation curves are flat out to ~ 5 disc scalelengths. The high efficiency of cooling also results in very low predicted X-ray luminosities for the residual haloes of hot gas.

Key words: galaxies: formation – cosmology: miscellaneous.

1 INTRODUCTION

The last decade has seen rapid progress in our understanding of how structure formed in the Universe. One of the most successful of current theories, the cold dark matter (CDM) scenario, postulates that most of the mass of the Universe is contributed by some unspecified weakly interacting massive particle with small initial thermal motions, and that structure is the result of quantum processes occurring during a very early inflationary epoch (Peebles 1982, 1984; Blumenthal et al. 1984). In its standard form, the theory also assumes an Einstein–de Sitter universe and that galaxies are biased tracers of the mass distribution (Davis et al. 1985).

Linear theory allows us to compute the statistical properties of density fluctuations at high redshifts, and hence the microwave background fluctuations they should produce. The strongly biased CDM model predicts fluctuations a factor of 2 smaller than those observed by the *COBE* satellite (Smoot et al. 1992). This discrepancy appears to be due to an incorrect *shape* for the power spectrum of density fluctuations, since other data require the amplitude of fluctuations on galactic cluster scales to be quite close to that predicted by standard CDM (Henry & Arnaud 1991; White, Efstathiou & Frenk 1993a). For this paper the relevant fluctuations are on still smaller scales, and indeed the power spectrum in our simulations can hardly be distinguished from a power law, $P(k) \propto k^{-2}$ (see Fig. 1 later). Thus for our purposes standard CDM just provides a convenient way to describe the normalization of the fluctuation spectrum.

In CDM-like model universes, small-scale structures collapse first and then merge hierarchically into more massive systems. This process can be followed in detail with the aid

of numerical simulations. Until recently, however, only the collisionless dark matter component could be treated. As a result, comparison with observation was uncertain and allowed room for considerable speculation. Further progress in the field seems to require the explicit inclusion of the observable component of the Universe (gas and stars). This involves solving the equations that govern the evolution of a collisional fluid and including the many additional physical processes that govern its evolution: star formation and evolution; energy input from stellar winds, stellar radiation and supernova explosions; ionization effects from stellar and quasi-stellar object (QSO) radiation fields; heavy element enrichment; heat conduction; and magnetic fields.

First attacks on these problems have used a variety of approaches, and most notably grid-based Eulerian schemes (see e.g. Cen 1992 and references therein) and Lagrangian schemes based on the smooth particle hydrodynamics technique (Evrard 1988; Hernquist & Katz 1989; Navarro & White 1993, hereafter Paper I). Gravitational interactions and the dark matter component are followed using standard N-body methods, particle-mesh codes, P^3M codes or tree codes (Efstathiou et al. 1985; Barnes & Hut 1986; Benz et al. 1990). This field is still in its infancy, and much work is needed before most of the relevant physical processes can be treated realistically.

One of the processes expected to be of critical importance during galaxy formation is radiative cooling. On galactic scales, cooling times can be significantly shorter than dynamical times (Binney 1977; Rees & Ostriker 1977; Silk 1977). As a result, gas temperatures can drop far below the virial temperature during protogalactic collapse. Without pressure support the gas contracts further, thereby increasing its density and decreasing its cooling time. This runaway process can only be halted if (a) a large fraction of the gas is transformed into stars; (b) most of the gas settles into a rotationally supported disc; or (c) external energy input (e.g. from supernovae) reheats the gas. In combination with the realization that galaxies are embedded in massive dark matter haloes, an elaboration of these ideas led to a first basic understanding of galaxy formation (White & Rees 1978; Fall & Efstathiou 1980; Faber 1982).

Many outstanding questions remain unanswered, however, some of which we consider in this paper. What fraction of the gas in a galactic halo can cool and settle into the central galaxy? How does this fraction depend on mass and formation epoch? Does the gas cool gradually from a hot, low-density atmosphere, as in a cooling flow, or does it contract into dense lumps at high redshifts, thereafter evolving mainly through mergers? Does the gas form discs resembling real spiral galaxies? How does the angular momentum of a gaseous disc relate to that of the surrounding dark halo? How frequent are mergers, and what is their influence on galactic structure?

Some of these questions can be studied using analytic models such as those developed by White & Rees (1978), Cole (1991), White & Frenk (1991), Lacey et al. (1993), Kauffmann, White & Guiderdoni (1993) and Cole et al. (1994). Such models allow the interaction of the many factors influencing galaxy formation to be studied relatively easily, and are now able to account for many of the properties of the observed galaxy population. Even the earliest of them showed quite clearly that the physics included in the simulations of this paper (gravitational evolution of the gaseous and dark matter components together with radiative cooling) leads to the excessive condensation of gas into small dense lumps at early times, and that additional physics is required to produce results consistent with observation. The simulations we present below show explicitly how this problem arises. Such simulations are critical to validate the simple assumptions made in analytic models, and in some cases can give quite precise results. Proper simulation of the full galaxy formation process is not, however, feasible because of the vast range of scales that are important and the critical role played by ill-understood physical processes such as star formation. We have therefore elected to study the effects of additional processes one at a time, rather than to include a rough representation of all relevant processes in a first exploration of the problem.

Galaxy formation experiments similar to the ones reported in this paper have been carried out by Katz & Gunn (1991) and Steinmetz & Müller (1993). They used codes similar to the TREESPH code of Hernquist & Katz (1989) to simulate the collapse of an isolated system containing gas and dark matter. Their initial system was a near-uniform expanding sphere in solid-body rotation and with small-scale density fluctuations. They found that the gas settled into a thin disc soon after the main collapse of the system, and that angular momentum was transferred to the dark halo from the gas during the formation process. Cooling was quite efficient and most of the gas ended up in a cold, rotationally supported disc. The main shortcomings of this work are the artificiality of its initial conditions and the isolated evolution of the system. The relation of the initial conditions to those expected in a CDM-like universe is uncertain, and the isolated boundary condition means that external tides are missing and that infall stops soon after the system first collapses.

The simulations reported follow the evolution of 'typical' objects in the CDM scenario, and treat their interactions with their environments in a fully consistent way. Angular momentum is acquired naturally through torques exerted by neighbouring systems, while infall continues until the present day. Our code is similar in many respects to that of Hernquist & Katz (1989), and is described in detail in Paper I. We will focus our analysis on how cold gas accumulates in the cores of dark haloes, and on the structure of these gaseous systems at $z=0$. We demonstrate the problem of over-efficient cooling referred to above, but find that, despite this, the final discs have a similar structure to real spiral galaxies. The overcooling problem may be resolved by reducing the baryonic fraction, by including the inhibiting effects of a photoionizing background (from QSOs or young stars) on the cooling function (Efstathiou 1992), or by including mechanical heating from stellar winds and supernovae. We intend to study these possibilities in future work. Although clearly not the final answer, we hope that the results of this paper will prove useful as a first step towards understanding some of the complex physics of this subject.

Our paper is organized as follows: Section 2 describes our numerical method and Section 3 describes the main results of the simulations, while Sections 4 and 5 discuss the implications of these results and summarize our conclusions.

2 THE NUMERICAL EXPERIMENTS

2.1 The code

We use a gridless Lagrangian code which combines the speed and accuracy of tree-based N-body techniques with the flexibility of the SPH approach. A detailed description of SPH is beyond the scope of the present paper, but an excellent review can be found in Benz (1990). The particular implementation of SPH used in our code and some relevant tests of the performance of the code are described in Paper I.

In SPH, information about the fluid properties is carried by pseudo-particles which move with the mean fluid velocity and whose locations serve as interpolation centres for the computation of local values and gradients of the thermodynamical variables. If each particle is characterized by its mass m_i, its specific internal energy u_i, and a smoothing length h_i, the hydrodynamical variables at any point in space can then be computed through kernel interpolation and the use of an adequate equation of state. In this work, we use the spline kernel function of Paper I and the equation of state of an ideal gas, $p=(\gamma-1)\rho u$, where p and ρ stand for the pressure and density, respectively, and γ is chosen to be 5/3. The physical processes modelled in the present version of the code include self-gravity, pressure gradients, shock heating and radiative cooling. The radiative cooling function used is a fit to the cooling curve of Dalgarno & McCray (1972), as appropriate for a gas of primordial H and He abundances. This fit is also in very good agreement with other published cooling curves, such as that used by Fall & Rees (1985). Cooling is very inefficient for temperatures below 10^4 K and is neglected in these calculations.

Individual smoothing lengths vary with time so that each gas particle keeps a fixed number of neighbours (~ 30) within its neighbour sphere. Gravitational forces are computed using a nearest neighbour binary tree that is free from geometric restrictions and allows the use of a large number of particles because of the $\sim N\log(N)$ scaling of the required CPU time (Benz et al. 1990). Gravitational interactions between neighbouring particles are softened using the same functional form as the interpolating kernel, but the softening lengths are set independently for gas and dark matter particles and are kept constant in time. Collisionless (dark matter) particles are, of course, subject to gravitational forces only.

2.2 The initial conditions

The generation of initial conditions for a simulation involves several steps. Initially, $\sim 32^3$ particles are arranged on a cubic lattice inside a sphere of radius ~ 20 Mpc. Position and velocity perturbations are computed using the Zel'dovich approximation in order to reproduce a CDM power spectrum normalized so that its variance, $\Delta(r_0)$, equals the inverse of the 'bias' parameter b for $r_0=16$ Mpc [$\Delta(16\ \mathrm{Mpc})=1/b=1/2.5$]. We assume the Hubble constant to be $H_0=50$ km s^{-1} Mpc^{-1} ($h=0.5$), and the density parameter to be $\Omega=1$, throughout this paper.

The perturbations consist of the superposition of N_k plane waves with amplitudes chosen from a Gaussian distribution with variance proportional to the power spectrum. We use equal number of waves per logarithmic interval. The phases of these waves are random, and only wavenumbers between the fundamental and Nyquist frequencies of the lattice are included. Fig. 1 shows the power spectrum of a typical realization with $N_k\sim 10^4$, computed by Fourier-transforming the density field of the perturbed particle distribution, compared to the 'actual' spectrum we are trying to reproduce. A cloud-in-cell scheme is used to assign particles to the grid. As expected, the agreement between the two spectra is very good at low frequencies, but at higher k the simulated $P(k)$ converges to the white noise level corresponding to the number of particles used (Efstathiou et al. 1985).

This system is then evolved (assuming isolated boundary conditions) from a high redshift, $z_i\approx 7$, until the present, $z_0=0$, at which time virialized clumps of different masses are identified. For simplicity, we do not select systems that are undergoing a major merger or are in the process of being accreted on to a larger structure. This may imply that our systems do not represent a completely fair sample of the population of haloes, but they are by no means atypical. All the systems reach their final equilibrium state through a series of mergers, in some cases as recent as $z=0.2$, but they show no indication of severe departures from equilibrium at z_0. A total of four systems with circular velocities ranging between 160 and 250 km s^{-1} are selected, and all the particles inside a sphere of mean overdensity 200 are traced back to z_i. All the particles belonging to each of these systems are fully contained in a sphere of comoving radius 2–3 Mpc at z_i.

Finally, a finer lattice is used to set up particles inside this sphere so that, at z_0, the sphere of overdensity 200 contains a specified number of particles, between 2000 and 3000 in the present simulations. Perturbations are computed as described above, using the *same* waves as before, but including the new ones allowed by the increased Nyquist frequency of the new lattice. The surrounding regions are coarsely sampled in order to provide the appropriate tidal field. In

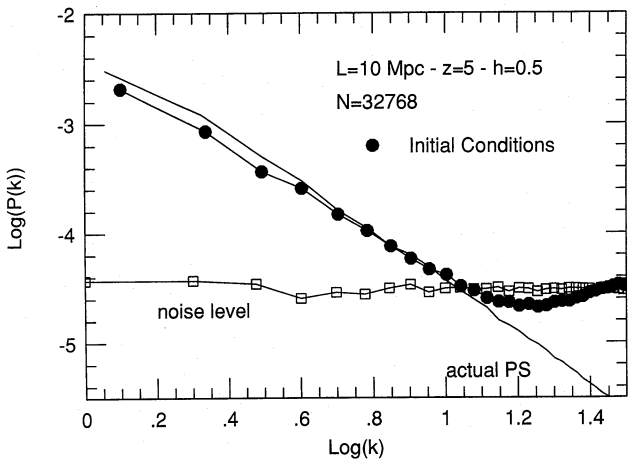

Figure 1. The power spectrum of initial conditions generated with the procedure described in Section 2.2. Filled dots are computed by Fourier-transforming the density field corresponding to the perturbed particle positions on a 32^3 grid. The solid line without symbols is the actual CDM power spectrum that we are trying to reproduce. Empty squares correspond to initial conditions where the particles have been placed at random within the box, and indicate the white noise level corresponding to the number of particles used. Other relevant parameters are indicated.

practice, we use the nodes of the tree structure of the previous simulation so that the accuracy of the tidal forces acting on the particles inside the selected region is very similar to that of the first run.

Tests of this procedure show that the overall properties and the history of the virialized systems formed in the first and second simulations are indistinguishable, except for the increased resolution of the latter. The main advantages of this procedure are that (a) the structural properties of the selected haloes are naturally determined by the self-consistent evolution of the whole system; (b) we can focus our resources on the region we are actually interested in; and (c) we can generate statistically equivalent systems with different numbers of particles. This last feature is most useful when trying to assess the effects of limited resolution (i.e. small N) on the results of the simulations (see e.g. Paper I).

2.3 The simulations

Once the initial dark matter density and velocity fields have been set up as described above, we insert the gaseous component by adding gas particles with an equivalent phase-space distribution. Their masses are determined by our adopted baryonic fraction, $\Omega_b = 0.1$, while the mass of the dark matter component is reduced to maintain $\Omega = 1$ overall. Gas is only included in the high-resolution part of the simulation, and its internal energy is chosen so that the average initial temperature is very low ($\sim 10^2$ K). The gravitational softening used for the simulations is 2 and 5 kpc for the gas and dark matter particles, respectively, in order to allow the structure of the gaseous discs to be adequately resolved.

We discuss four main simulations, C1–C4, whose properties we list in Table 1. The radius of the high-resolution sphere is listed in the column labelled r_0, the total mass inside this radius is M_0, and the total number of particles inside this radius (both gas and dark matter) is given in column N_0. Equal numbers of gas and dark matter particles are used in the high-resolution region. As tests, we repeated C1 with the gaseous fraction reduced by a factor of 5 (C1a) and repeated C3 with half as many particles in the high-density region (C3a). A typical run took about 200 h of CPU time on a SUN IPC workstation.

3 RESULTS

3.1 The evolution

Different stages of the evolution of one particular model, C3, are shown in Fig. 2. These snapshots are centred on the particles that at redshift zero form one of the disc/halo systems selected for analysis. The snapshots are 500 kpc across and contain about 3000 particles each. Most of the gas is seen to gather at the centre of collapsing dark matter clumps, and to form tightly bound cores of cold gas. Interactions between neighbouring systems endow each system with a small amount of angular momentum, which is what ultimately stops the collapse of the gas and enables thin structures resembling galactic discs to form. These discs often merge as their surrounding haloes coalesce. Due to the efficient conversion of kinetic energy into heat by shocks and to the short cooling times typical of such dense material, the gas settles into a new disc very soon after a merger. Blown-up face-on and edge-on views of the gaseous disc at $z = 0$ are shown in Fig. 3.

Table 1. Parameters of the runs.

Run	r_0 [Mpc][a]	M_0 [$10^{10} M_\odot$]	N_0	Ω_b
C1	3.5	1234	14000	0.10
C2	3.0	777	16000	0.10
C3	3.0	777	16000	0.10
C4	3.2	943	12000	0.10
C1a	3.5	1234	14000	0.02
C3a	3.0	777	8000	0.10

[a]Comoving.

The velocity field in the disc is also shown in Fig. 3; nearly all the kinetic energy is invested in coherent rotation. Most of the gas in the system (~ 90 per cent in this case) is actually in the central disc, which is surrounded by a hot, pressure-supported halo (see Fig. 8, later) and sometimes by smaller satellites like the one seen in Fig. 3. These satellites will eventually merge with the main disc as a result of dynamical friction. This is, in fact, the way that most of the mass is accreted on to the disc: by the addition of individual baryonic cores. The dark halo shows no indication of such substructure, suggesting that gaseous cores evolve differently from their surrounding dark matter and may survive the merging of their parent haloes. When proper account is taken of the orbital parameters of the satellites and of their own dark haloes, simple dynamical friction estimates give on average a good indication of when these satellites will merge with the central disc. Full details concerning this issue are deferred to a later paper, where we analyse a larger number of simulations (Navarro, Frenk & White, in preparation).

Fig. 4 illustrates how the mass of the major halo in each simulation grows with time. The symbols connected by dashed lines give this mass in units of the total mass of the system at $z = 0$, $M_{\rm vir}(z_0)$. A halo mass is defined to be the mass within the largest sphere centred on the gaseous core with a mean overdensity exceeding 200. We consider this sphere to be the boundary of the virialized halo, and denote its radius as $r_{\rm vir}$. Major mergers can be identified in Fig. 4 as sudden jumps in the evolution of the mass curves. If we define the redshift of formation ($z_{\rm f}$) of each halo as the time at which half its dark mass has been assembled, the average $z_{\rm f}$ for these four simulations is ≈ 0.6. This late formation time is in reasonable agreement with the analytic estimates of Lacey & Cole (1993), who derive an average formation redshift $z_{\rm f} \sim 0.63$ for a halo of 10^{12} $M_\odot$ and a CDM spectrum with a bias parameter $b = 2.5$.

Do the gaseous cores merge as their surrounding dark haloes coalesce? This is an interesting question because thin stellar discs may be destroyed by the accretion of satellites, a fact that has been used to argue against hierarchical clustering theories of galaxy formation in an $\Omega = 1$ universe (Toth & Ostriker 1992). The symbols connected with solid lines in Fig. 4 illustrate how the mass of the main gaseous disc ($M_{\rm disc}$) grows with time. We define this to be the gas mass within 20 (physical) kpc of the centre of the most massive core, and we plot it in units of $\Omega_b M_{\rm vir}(z_0)$. The overall evolution of the core mass is in general not very different from a constant accretion rate, and indicates that the gaseous cores have increased their mass by about a factor of ~ 4 since $z = 1$. Half

the mass of the disc was already in place by $z = 0.4$, and at $z = 0$ the gas mass in the inner 20 kpc is a large fraction of the total baryonic content of the halo.

If all the gas were immediately collected at the centre of a dark halo, the dark matter and gas curves would overlap in Fig. 4. This is apparently not the case. Mergers that show up clearly in the dark matter curves are not seen in the disc evolution until some dynamical times later, indicating that gaseous cores survive for some time after their surrounding haloes merge. Consider, for example, run C1 (open triangles), where the halo undergoes a major merger at $z = 0.67$ ($t \sim 700$). The cores of these haloes do not merge until $z = 0.3$, implying a survival time of almost 3×10^9 yr. The survival time can be as short as $\sim 10^9$ yr (run C2), or can exceed $\sim 5 \times 10^9$ yr. The latter is the case in run C4, where the main gaseous cores have yet to merge at $z = 0$.

Mergers appear to proceed more rapidly when systems of similar mass are on nearly radial orbits (C2), and to take longer when they involve a number of haloes of differing mass merging almost simultaneously (as is the case in C4). If we define a 'major merger' to be the sudden accretion of more than 20 per cent of the existing disc mass, the last

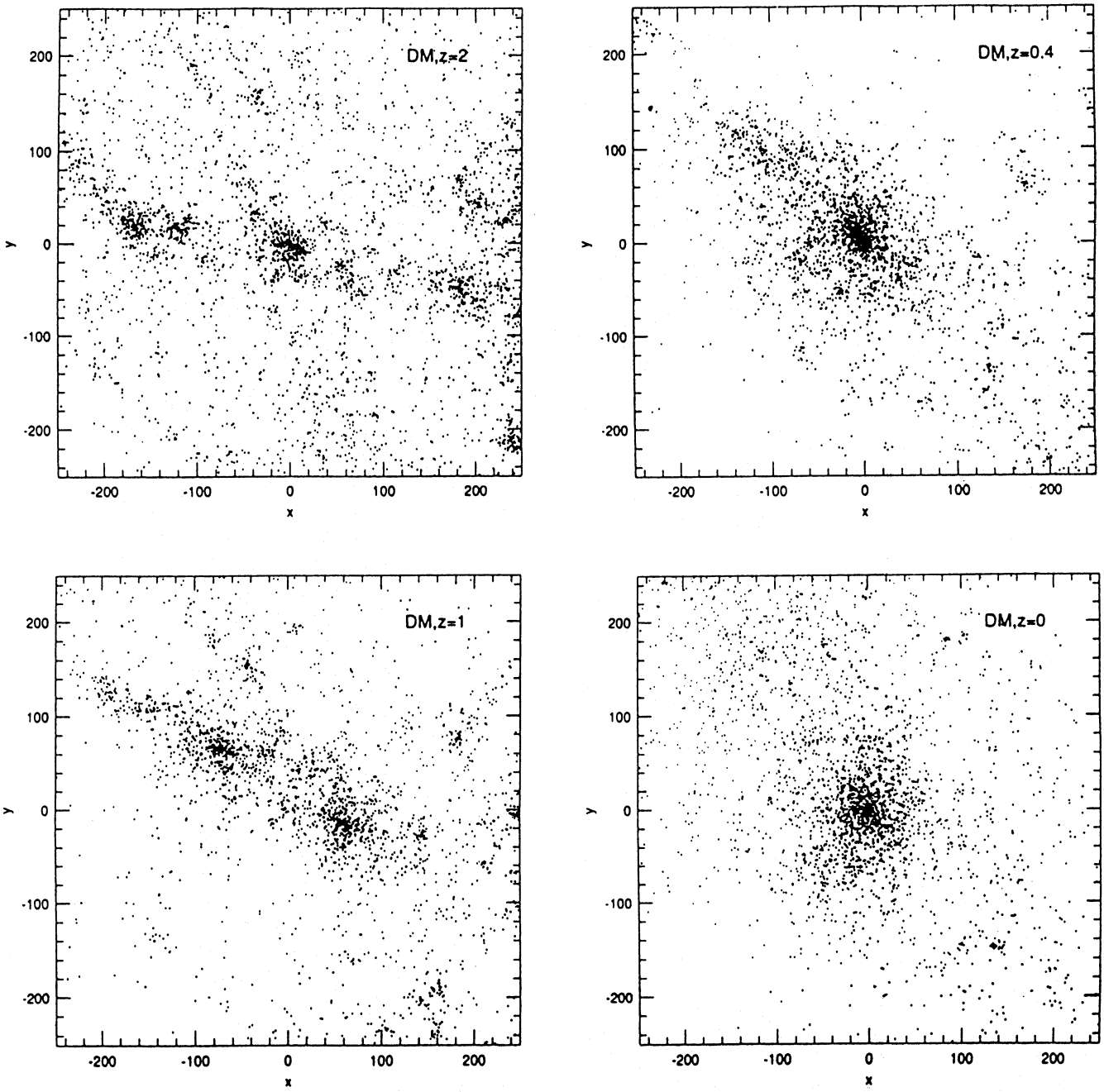

Figure 2. Projected positions of gas and dark matter particles at different times in one of the simulations (C3). Redshifts of 2, 1, 0.4 and 0 are shown. Each box is 600 (physical) kpc across, and contains about 3000 particles. Note how the gas forms tightly bound cores at the centres of non-linear clumps. These gaseous knots are hardly disturbed when the dark haloes collide, and they take longer than the haloes to complete a merger. For example, note that at $z = 0.4$ two large gaseous clumps survive, while there is little substructure in the dark matter.

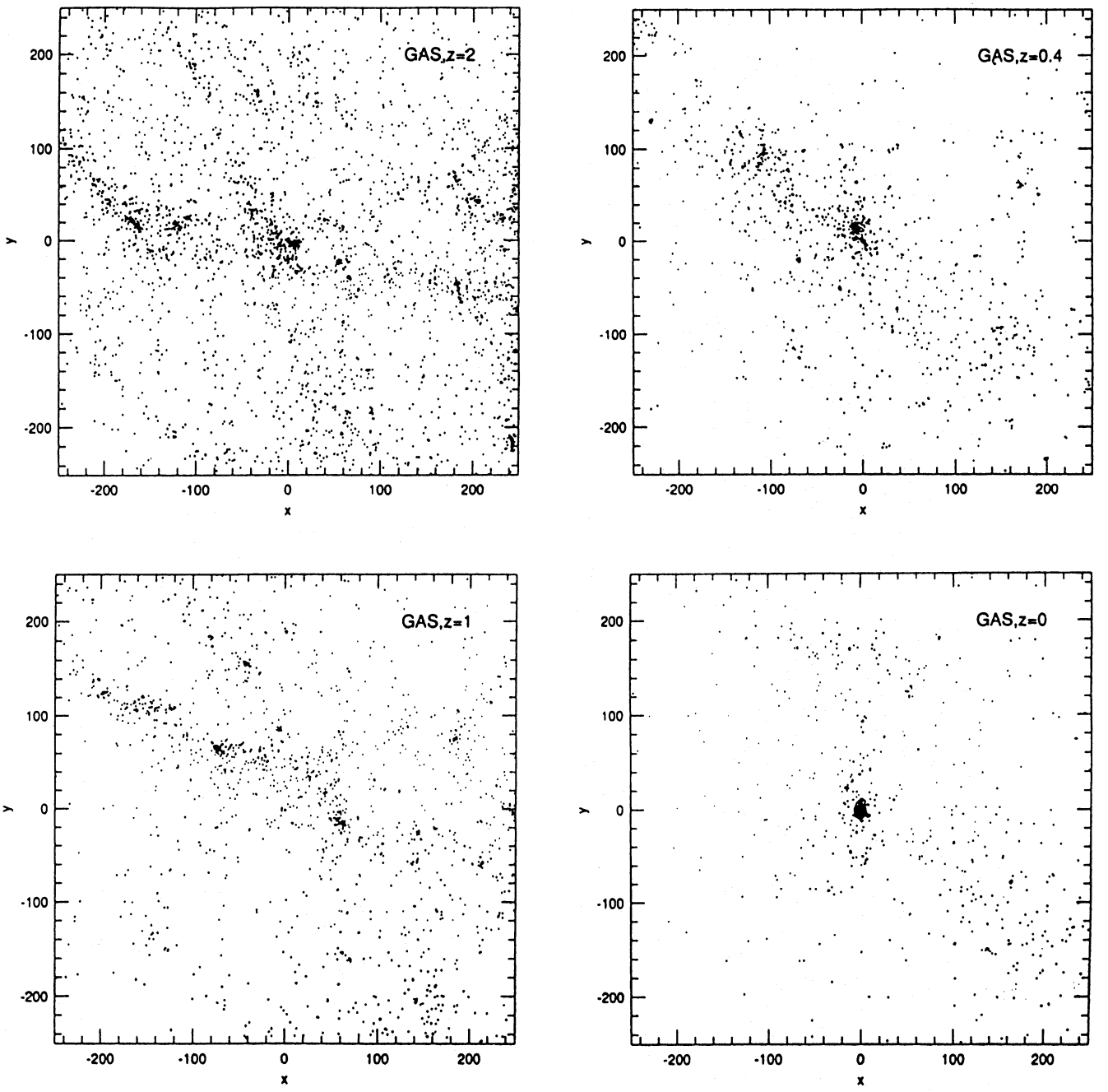

Figure 2 – *continued*

major mergers occurred at $z \sim 0.76$, 0.50, 0.25 and 0.20 in our four simulations.

Fig. 4 also gives some indication of how our results may be affected by the parameters chosen for these simulations. If Ω_b is reduced below our adopted value of 0.1, cooling should become less efficient. This might produce a significantly lower concentration of gas in the disc at $z = 0$. This effect is shown in the C1 panel by the curves without symbols, which correspond to lowering the baryonic abundance by a factor of 5 (to $\Omega_b = 0.02$). Although the dark matter evolution is barely affected, only ~70 per cent of the available gas reaches the disc by $z = 0$, compared to ~90 per cent in the case where $\Omega_b = 0.1$. The remainder is in the hot, pressure-supported halo which surrounds the disc. Although significant, this reduction is still not enough to overcome the 'overcooling' problem mentioned in Section 1, although it does lessen some of the problems associated with disc structure which we discuss in the next section. A test for artefacts due to our limited resolution is shown by the curves without symbols in panel C4. The evolution is quite similar for the two simulations shown, even though the additional model uses only half as many particles to represent the mass distributions.

The dense, gaseous cores that form in these simulations are presumably the regions of the universe in which most star formation activity will occur. It is thus important to investi-

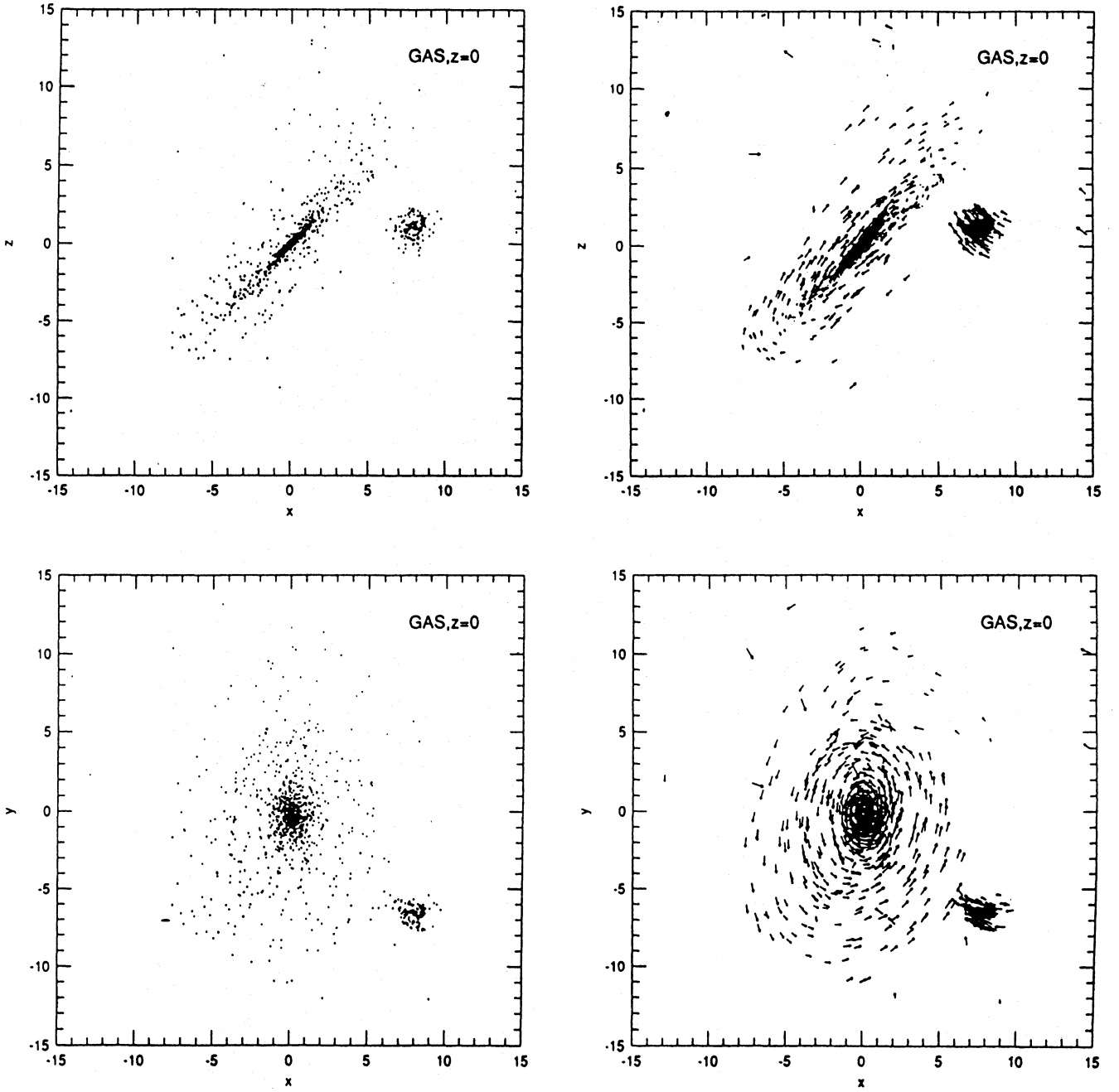

Figure 3. A blow-up of the gaseous disc in Fig. 2 at $z = 0$. The box is now 30 kpc across. The left-hand panels show two orthogonal projections of the system. The right-hand panels show the velocity field of the gas (see also Fig. 7). Note the satellite that orbits very close to the main disc.

gate how much mass is actually locked up in them. Fig. 5 shows how much of the final baryonic content of our main galactic haloes is in cold, dense clumps at each earlier time. Gas particles with densities higher than $\sim 200\Omega_b$ times the cosmic mean are considered part of dense clumps. Due to the extremely short cooling times corresponding to this high density, the temperatures of such clumps are always $\sim 10^4$ K, the cut-off temperature of our cooling curve. By a redshift of 1, ~60–70 per cent of the gas of each final galactic halo is already in high-density cores, even though Fig. 4 shows that only a small fraction of this is actually in the central disc. Fig. 5 also shows that by $z = 0$ baryons are slightly overabundant relative to the dark matter inside a region of overdensity 200. This overabundance is always smaller than about 20 per cent, and is expected for rapidly cooling gas collapsing with a collisionless dark matter halo (White et al. 1993b).

Previous semi-analytic models estimated the mass of gas accumulating in the core of a dark halo from simple energetic or timing arguments. A rough upper limit to the mass of a disc forming at the centre of a dark halo of circular velocity

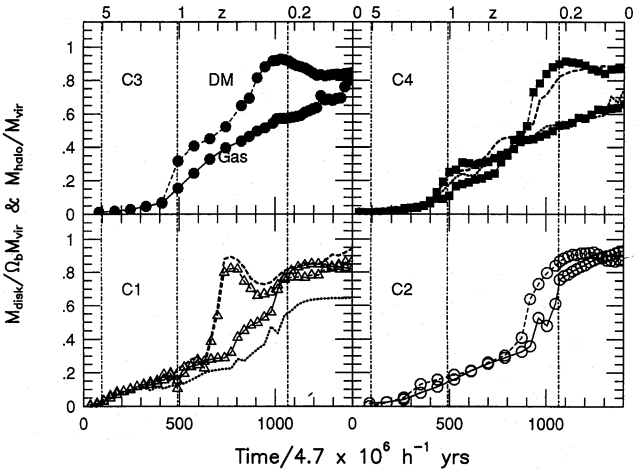

Figure 4. The mass of the main disc (dashed lines) and that of its surrounding halo (solid lines) in runs C1, C2, C3 and C4 as a function of time. Disc masses are plotted in units of the total baryonic mass inside a sphere of overdensity 200. Dark halo masses are in units of the total mass inside a sphere of overdensity 200. Vertical dot–dashed lines indicate times corresponding to redshifts of 5, 1 and 0.2. Note that the disc is constantly accreting mass, and that about half its mass has been accreted in the last 6–7 Gyr. Merger events can be identified as sudden increases in mass. Dark masses jump suddenly during a merger, and then decrease somewhat as the system relaxes to equilibrium. Gaseous cores take longer to merge than the dark haloes, as demonstrated by the time offset between halo and gas mergers. The curves without symbols in panel C1 correspond to a run with $\Omega_b = 0.02$, and those in panel C4 correspond to a run in which the number of particles is halved.

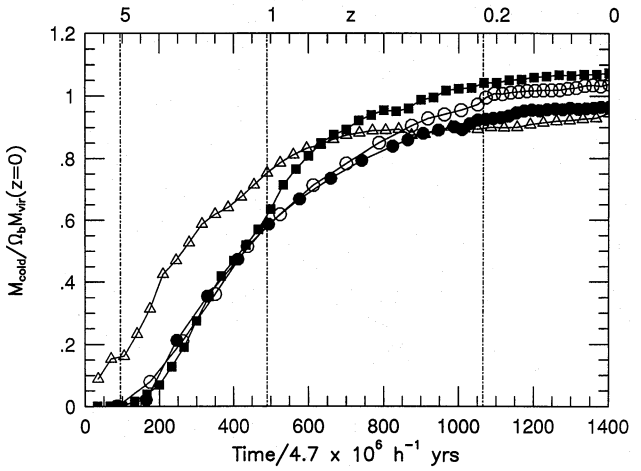

Figure 5. The growth of the amount of the gas in systems identified at $z = 0$ that is already condensed into cold, dense clumps at each earlier time. Symbols are as in Fig. 4. Masses are normalized to the total baryonic mass expected for a halo of the same circular velocity. Most of the mass forms tightly bound cores by $z = 1$. Values of the ordinate larger than 1 indicate a small overabundance of baryons relative to the initial value of 10 per cent ($\Omega_b = 0.1$).

V_c at redshift z is given by the total mass of the virialized region of the halo (i.e. within a density contrast of 200; see White & Frenk 1991; White et al. 1993b), $\Omega_b M_{vir}$, where

$$M_{vir}(V_c, z) = 0.1 G^{-1} H_0^{-1} (1+z)^{-3/2} V_c^3. \quad (1)$$

An alternative estimate can be obtained by assuming that, when a halo forms, the gas is shock-heated to the virial temperature and is distributed similarly to the dark matter. Since the equilibrium structure of virialized collisionless systems is fairly well approximated by an isothermal sphere, it is straightforward to define a 'cooling radius', r_{cool}, the distance from the centre of the system at which the local cooling timescale equals the age of the universe. The gas mass encompassed by r_{cool} (M_{cool}) provides another estimate of the mass that may be able to assemble at the centre of a dark halo (Cole 1991; White & Frenk 1991). In massive systems and at late times, M_{cool} is smaller than $\Omega_b M_{vir}$ so that cooling may limit the accumulation of gas, whereas in small systems and at early times the reverse is true, suggesting that all the gas within each halo will be in the cold central core.

Fig. 6 shows core mass as a function of halo circular velocity for systems identified at $z = 1$ and 0. We include all systems with more than 200 gas particles in our four main simulations, and we overplot the predicted values of M_{cool} and $\Omega_b M_{vir}$. It is clear that core masses are generally very close to $\Omega_b M_{vir}$, and that at $z = 0$ they are significantly larger than M_{cool} for the more massive systems. This occurs because most of the gas that ends up in non-linear clumps is locked up in the dense cores of smaller clumps at early times. Subsequent mergers are unable to disrupt these tightly bound cores, which spiral into the centre of the newly formed haloes and merge to form thin discs like the one shown in Fig. 3. The gas is never reheated to the virial temperatures of the larger haloes.

A related result is that the residual fraction of hot gas is quite small, in fact not larger than ~ 10 per cent in any of the systems plotted in Fig. 6. As a consequence, these systems are weak X-ray emitters, the bulk of the gas being too cold to produce X-rays. Estimates of the X-ray luminosity emitted in the (0.5, 10)-keV passband are given in Table 1.

A third related result, that the gaseous discs in our major haloes appear too massive (or too concentrated), is discussed in the next section. All these results are closely related to the 'overcooling' problem. The fraction of the mass of our simulations that ends up in dense baryonic clumps of galactic density (most of the gas, and so almost 10 per cent of the mass) is much greater than the fraction of the closure density in observed galaxies (roughly 0.5 per cent). Furthermore, if cold dense gas were always able to form stars with high efficiency, most of the baryons in our models would be turned into stars at early times, and there would be little gas left to make discs in our final haloes. We would thus end up with a central 'elliptical galaxy' formed as a result of the merger of stellar cores, surrounded by stellar satellites some of which might still retain a disc structure.

Avoidance of these problems clearly requires physical processes beyond those included here. Previous semi-analytic models have all recognized this, and have shown that UV background-suppressed cooling, inefficient star formation, reheating of gas by supernovae and stellar winds, and interaction-induced star formation are all plausibly important, and may all offer ways both to reduce the overall amount of material that cools, and to bias cooling and star formation towards large objects and late times (White & Rees 1978; Cole 1991; White & Frenk 1991; Cole et al. 1994; Kauffmann et al. 1993; Lacey et al. 1993). Inclusions of such effects in our models would substantially reduce the masses

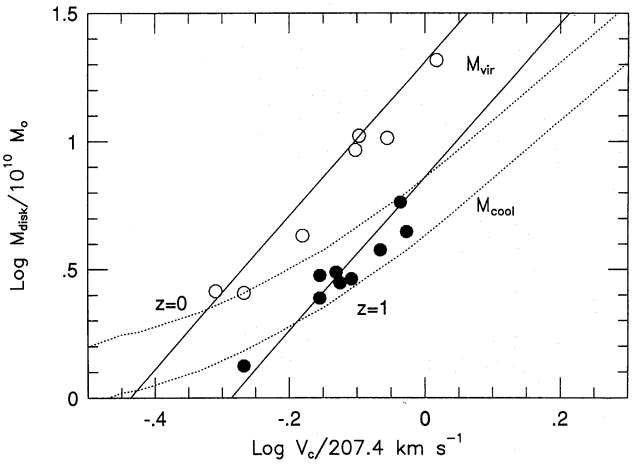

Figure 6. The mass of cold, gaseous cores (defined to be mass within the inner 20 kpc) as a function of the circular velocity of their surrounding haloes. All clumps with $N_{gas} > 200$ particles are included. Open symbols correspond to $z = 0$ and filled symbols to $z = 1$. Estimates based on the total gas mass within the virialized region ($\Omega_b M_{vir}/h$) are given by the solid lines, while the dotted lines are estimates based on the 'cooling mass' (M_{cool}) computed from a no-metal cooling function as described in the text. In each case, upper and lower lines correspond to $z = 0$ and $z = 1$, respectively.

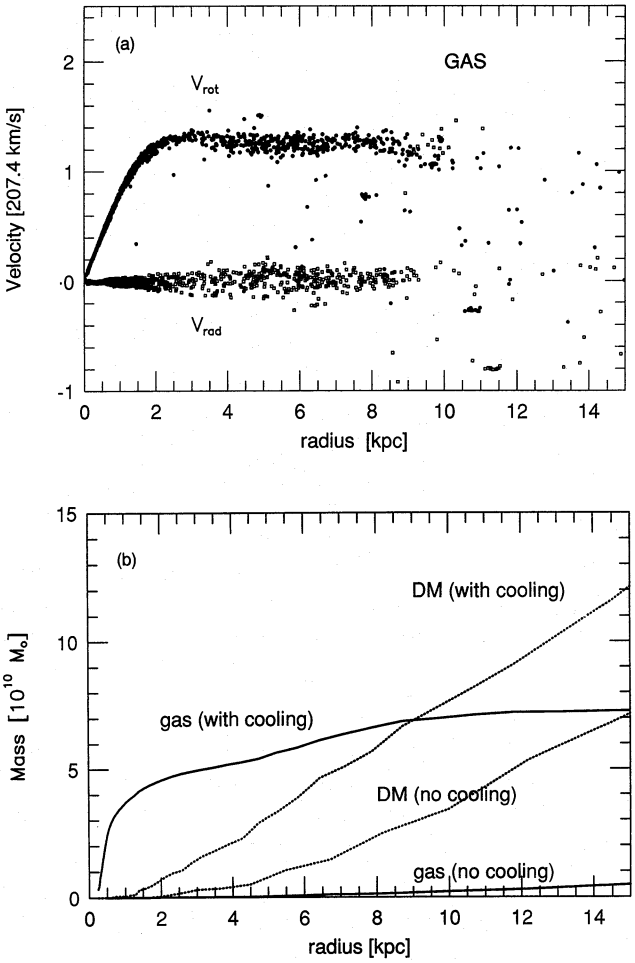

Figure 7. (a) The radial and tangential velocities of all the gas particles in the disc shown in Fig. 3. The rotation curve is rather flat outside 2 kpc, and no major radial motions are seen, indicating that the disc is in equilibrium. (b) The cumulative distributions of mass for gas and dark matter in this same system, together with the corresponding distributions for an equivalent run without cooling. The gaseous disc is completely self-gravitating close to the centre, and at ~ 10 kpc the gas and dark matter contribute similar amounts to the total mass. Note that dissipation draws not only gas into the inner regions but also a substantial amount of dark matter.

of cores of cold gas, particularly at early times, and would greatly enhance the amount of gas in a hot, diffuse, X-ray-emitting component. This is likely to produce major changes in the behaviour discussed in this and the next section. We do not, however, believe that the interaction of such complex processes can be reliably modelled without a clear understanding of the evolution of the simpler system treated in this paper.

3.2 The equilibrium structure of the discs

Fig. 3 shows the velocity field of the gas in one of the discs identified at $z = 0$ (the primary disc in C3). It is clear from this figure that the gas is spinning rapidly about an axis perpendicular to the plane of the disc, and that radial motions are very small. This visual impression is confirmed by Fig. 7(a), which plots radial and tangential velocities for the individual gas particles. The rotational velocity curve of the disc is similar to that of a typical spiral galaxy. Near the centre the rotational velocity rises linearly out to about 2 kpc, where it flattens and stays nearly constant as far as the edge of the disc, at about 10 kpc. The radial velocity dispersion of the gas plays a negligible role in the dynamics of the disc. Moreover, the disc seems to be in equilibrium, without any significant inwards or outwards bulk motions. The central linear portion of the rotation curve reflects the softening of gravitational forces in our code and is therefore non-physical.

Most of the disc mass is concentrated near the centre; in fact, 61 per cent is packed into the inner 1 softening length (2 kpc). The disc is almost completely self-gravitating, and only at about 10 kpc does the cumulative dark matter mass equal that of the gas (Fig. 7b). This transition radius is in rough agreement with observational estimates for a typical disc galaxy. However, the mass fraction in an unresolved central clump, which ranges from 33 to 90 per cent in our four models (column labelled M_{inn} in Table 2), is much too large to be compatible with real galaxies. For example, in a spiral with a rotational velocity of 300 km s^{-1} the half-mass radius of the observed stellar material is typically 10 kpc, whereas in our models it is in the range 1–3 kpc. It appears that the gas in the models has contracted too much relative to the dark matter. Values of the collapse factors R_{DM}/R_{gas}, defined as the ratio of the half-mass radius of a dark halo to that of its cold baryonic core, are in the range 15 to 70, much larger than the value $\sim 0.4/\lambda$ expected for dissipative contraction of the gas at fixed angular momentum (Fall & Efstathiou 1980; λ here is the spin parameter of the halo; see Table 2 for values of these quantities). Note, however, that precise values of the R_{DM}/R_{gas} ratio may depend on the numerical procedure, as in many cores R_{gas} is smaller than the gravitational softening.

Table 2. Parameters of the systems at $z = 0$.

Run	λ	V_c [km/s]	V_{rot} [km/s]	M_{disk} [$\Omega_b M_{vir}$]	M_{hot} [$\Omega_b M_{vir}$]	M_{inn} [M_{disk}]	R_{DM} [R_{gas}]	$\log L_X$ [erg/s]
C1	0.053	216	380	0.88	0.12	0.33	15.4	39.0
C2	0.075	166	280	0.94	0.07	0.90	66.1	38.5
C3	0.037	164	260	0.85	0.07	0.61	19.3	38.6
C4	0.039	180	310	0.68	0.05	0.48	41.1	38.0
C1a	0.053	216	241	0.65	0.20	0.55	31.2	38.1
C3a	0.037	164	250	0.73	0.11	0.42	17.0	38.3

A related problem is that the amplitude of the disc rotation curve (V_{rot}) greatly exceeds the circular velocity of the surrounding dark halo (V_c) measured at substantially larger radius (~300 kpc). Table 2 quotes both values for the main discs in our four simulations; V_c is on average 40 per cent smaller than V_{rot}. However, data on satellites orbiting bright spiral galaxies show that the real drop over this radial range is much smaller (~20 per cent; Zaritsky & White 1994). The difference between V_{rot} and V_c results from the large collapse factors of the gas together with the particular baryonic fraction adopted in these simulations ($\Omega_b = 0.1$). A lower Ω_b would lead to lower disc rotational velocities, but would not help to produce discs as large as those observed (see, for example, model C1a in Table 2).

As anticipated in the previous subsection, the discs have a temperature close to the cut-off temperature adopted in our cooling algorithm, $T_{min} = 10^4$ K. This is well below the typical virial temperature of the surrounding haloes, and implies that pressure plays no role in supporting the disc structure. The surrounding halo, however, is filled with a tenuous atmosphere of hot, pressure-supported gas with a cooling time longer than a Hubble time. At large radii, this hot halo dominates the temperature profile shown in Fig. 8. Note, however, that its mass is well below that of the disc (Table 2). The gaseous halo is not made up of this hot atmosphere alone. In fact, it is a two-phase system that also includes cold, dense satellites like the one shown in Fig. 3.

The surface density profiles of our gaseous discs are shown in Fig. 9. Outside the central unresolved lump, three of the four profiles are approximately exponential with scalelengths between 3 and 5 kpc. The exception is the disc formed in run C2 (open circles), where almost all the disc mass lies in the inner 2 kpc. The extreme concentration of this case can be traced to the last major merger it experienced. This happened at $z = 0.2$ and involved two discs of near-equal mass on an almost radial orbit. Since the internal spins of these discs were not parallel, the remnant ended up with very little angular momentum. Without rotational support, the gas contracts until it lies within the gravitational softening radius, where the collapse is artificially halted.

The gaseous core in run C2 is an extreme example of an effect that is present in all our simulations. In general, as dark haloes merge their cores transfer much of their orbital angular momentum and orbital energy to the surrounding dark matter (Frenk et al. 1985; Barnes 1988; Quinn & Zurek 1988; Navarro & Benz 1991). The result is a substantial reduction in the angular momentum of the gaseous component. This is illustrated in Fig. 10, which plots the specific angular momenta of the gas and the dark matter contained in cylindrical shells aligned with the total angular momentum

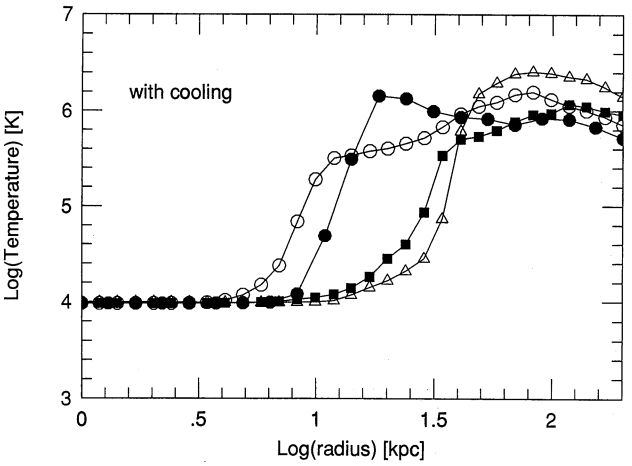

Figure 8. Temperature profiles of the gaseous component (runs with cooling). Symbols correspond to those in Fig. 4. The discs are nearly isothermal and the gas stays at $T \sim 10^4$ K, because its high density implies very short cooling times. Each disc is surrounded by a hot, low-density atmosphere of pressure-supported gas at approximately the virial temperature of the system. The transition from the cold phase to the hot phase is rather abrupt, and occurs at 10–30 kpc. The average density of gas in the hot phase is very low, which implies that these systems would be very weak emitters of X-rays.

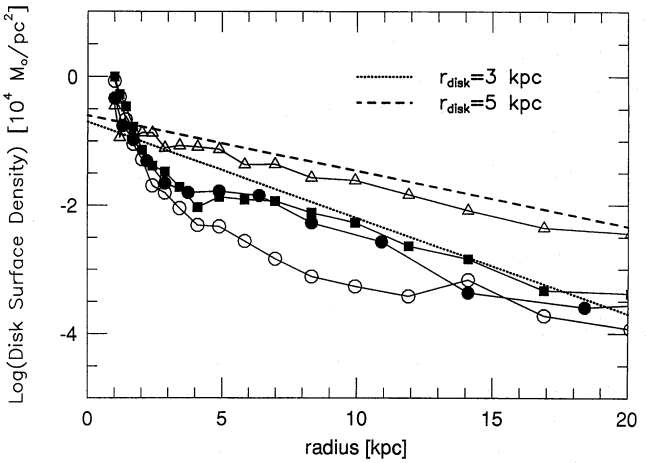

Figure 9. The surface density profiles of the gaseous discs. The symbols correspond to different simulations as in Fig. 4. Exponential profiles with scalelengths of 3 and 5 kpc are also shown.

vector, each of which contain 20 per cent of the mass of the relevant component. Dotted curves correspond to the gas, and solid curves to the dark matter. If there had been no exchange of angular momentum between the two components these two sets of curves would be expected to coincide. This is far from the case, and it is clear that the formation and merging of dense, gaseous cores at the centres of dark haloes has been accompanied by large losses of angular momentum. These losses are in turn responsible for the large contraction factors needed to reach centrifugal equilibrium.

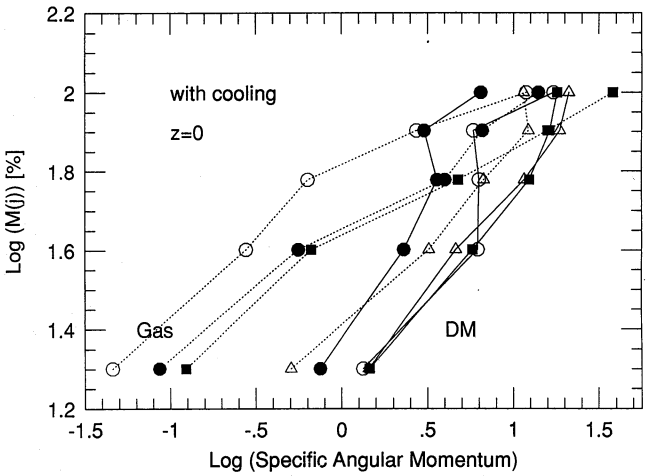

Figure 10. The specific angular momentum distribution of the gaseous and dark matter components in cylindrical, equal-mass bins aligned parallel to the total angular momentum vector. The ordinate is the cumulative mass of each component in these bins. Dotted lines are used for the gaseous component and solid lines for the dark matter. Symbols are as in Fig. 4.

4 DISCUSSION

An important result of our simulations is that, in the absence of processes other than radiative cooling, gas collects at the centre of a system very soon after it collapses and forms a thin, rotationally supported disc (Fig. 3). As clumps merge to form new systems these cores also merge and make new discs which contain most of the gas. In hierarchically clustering scenarios like CDM it is thought that *all* large structures are formed by the amalgamation of smaller ones. If gas collects so efficiently at the centres of dark haloes and is only slightly perturbed by subsequent mergers, it is hard to understand why the largest virialized structures in the present Universe (rich galaxy clusters) have most of their observed baryons in the form of a hot, smoothly distributed gas (see, for example, Jones & Forman 1992 and references therein). How did the gaseous component avoid being locked up in low-mass clumps at high redshifts?

Another way of presenting the same problem is to consider that, although the primordial abundance of baryons appears, from nucleosynthesis arguments, to be $\Omega_b \approx 0.05$ (for $h = 0.5$; Walker et al. 1991), the directly observed abundance of baryons (mainly stars in galaxies) is only about a tenth of this (Persic & Salucci 1992). If star formation were to proceed unhindered in the baryonic discs in our simulations, it would certainly produce an overabundance of stars at the present time. This is, of course, the 'overcooling' problem discussed above. Our simulations seem to confirm the earlier theoretical work which asserted that additional effects, for example radiative and hydrodynamic feedback from star formation, must have a critical effect on the formation history of a galaxy.

Fig. 4 indicates that the mass of a disc is continuously growing, and that most of the mass is added through mergers. Can this mechanism really produce the thin stellar discs characteristic of spiral galaxies? The answer is still far from certain. Mergers have occurred as recently as $z = 0.2$ in our models, and the mean redshift of the last major merger is ~ 0.4. This means that a typical galaxy would have been seriously perturbed as recently as ~ 5 Gyr ago. It is dangerous, however, to draw conclusions about merging statistics from a set of only four simulations. In addition, the cure for the overcooling problem may reduce both the number of merging lumps and their mass, while increasing the fraction of the disc that is accreted from a relatively smooth component. We are presently carrying out a large number of simulations to address the first problem, and we are gradually incorporating additional physical processes into our modelling in order to investigate the second.

Given the excessive contraction of the gas to form our simulated discs, it is surprising that the shape of the rotation curves of the discs agrees quite well with that observed in bright spirals (see one example in Fig. 7a). Similar results were found by Katz & Gunn (1991) who claimed that the flatness of these curves is due to the intrinsic structure of the gaseous disc together with the increased amount of surrounding dark matter (as compared with adiabatic simulations) which has been drawn inwards by the central concentration of gas (Blumenthal et al. 1986). Fig. 7(b) shows that this process also operates in our simulations. We compare the inner mass profiles of the gas and dark matter with those obtained in an equivalent simulation where cooling effects are switched off. It appears that the dissipative contraction of the gas does indeed bring in just enough dark matter to maintain a flat rotation curve.

In our simulations the cold gas contracts in radius by a large factor, and much of it ends up in an unresolved central knot. This shows that angular momentum is transferred very efficiently from gas to dark matter as the discs form and merge. Clearly, models that assume detailed conservation of angular momentum are a very poor description of such systems. However, as noted by Navarro & Benz (1991), although this process could conceivably give rise to concentrated stellar systems such as elliptical galaxies and the bulges of spirals, the loss of angular momentum seems too great for it to be a viable mechanism for making real spiral discs. They argued that feedback processes related to star formation can result in less concentration of the baryonic component at early times, and so less transfer of angular momentum to the dark matter. The present study seems to reinforce their conclusions, which were based on much lower resolution simulations.

While these problems mean that our models cannot yet be considered acceptable descriptions of the formation of real disc galaxies, it is encouraging that with very simple physics –cooling gas in a dynamically evolving population of dark haloes – it is indeed possible to produce condensed centrifugally supported systems with approximately the sizes and masses of galaxies.

5 CONCLUSIONS

We have presented self-consistent simulations of the formation of virialized clumps of gas and dark matter in an $\Omega = 1$ universe dominated by cold dark matter. The clumps are chosen to have potential-well depths corresponding to galactic haloes ($V_c \approx 200$ km s^{-1}). The simulations include the effects of gravity, of pressure gradients, of shocks, and of radiative cooling, but they do not include star formation and the attendant gas-heating processes. Our main conclusions are as follows.

(i) Practically all the gas in a virialized system ends up in a dense, cold core at its centre or in dense cold satellites which are the remnant cores of earlier clumps. These cores can merge during subsequent evolution, but they are never reheated. This confirms earlier suggestions that some physical mechanism (e.g. supernovae) must prevent this 'overcooling problem' if we are to reconcile the CDM model with the low baryonic density (in the form of stars) at the present time, and with the large amount of X-ray-emitting gas in galaxy clusters.

(ii) The gaseous cores at the centres of dark haloes are thin, rotationally supported structures. These discs are quickly regenerated after a merger. Their outer regions have rotation curves and surface density profiles that have the same shape as those of disc galaxies.

(iii) Hierarchical clustering with dense gaseous cores involves the transfer of large amounts of angular momentum from the gas to the dark matter. As a result, the gaseous cores are substantially smaller than predicted by simple analytic arguments that assume conservation of angular momentum during contraction within any given halo. Indeed, the cores formed in our models are substantially more compact than real disc galaxies. This problem may well be solved by the same processes that solve the overcooling problem.

(iv) A simulated galactic disc has on average accreted more than half its present mass since $z = 0.5$, and has undergone its last major merging encounter at $z \sim 0.4$. The average formation time of its surrounding dark halo is $z_f \sim 0.6$. It is unlikely, then, that thin stellar discs much older than ~ 5 Gyr can form in most galaxies in this scenario. More simulations are needed to give a proper statistical basis for this conclusion, which may also be affected by the final resolution of the overcooling problem.

ACKNOWLEDGMENTS

JFN acknowledges discussions with Shaun Cole and Carlos Frenk, and the support of an SERC postdoctoral fellowship.

REFERENCES

Barnes J. E., Hut P., 1986, Nat, 324, 446
Barnes J. E., 1988, ApJ, 331, 699
Benz W., 1990, in Buchler J. R., ed., The Numerical Modelling of Nonlinear Stellar Pulsations: Problems and Prospects. Kluwer, Dordrecht, p. 269
Benz W., Bowers R. L., Cameron A. G. W., Press W. H., 1990, ApJ, 348, 647
Binney J., 1977, ApJ, 215, 483
Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nat, 311, 527
Blumenthal G. R., Faber S. M., Flores R., Primack J. R., 1986, ApJ, 301, 27
Cen R., 1992, ApJS, 78, 341
Cole S., 1991, ApJ, 367, 45
Cole S., Aragon A., Frenk C. S., Navarro J. F., Zepf S., 1994, MNRAS, in press
Dalgarno A., McCray R. A., 1972, ARA&A, 10, 375
Davis M., Efstathiou G. P. E., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Efstathiou G., 1992, MNRAS, 256, 43P
Efstathiou G., Davis M., Frenk C., White S. D. M., 1985, ApJS, 57, 241
Evrard A. E., 1988, MNRAS, 235, 911
Faber S. M., 1982, in Brück H. A., Coyne G. V., Longair M. S., eds, Astrophysical Cosmology. Pontificia Academia Scientiarium, Vatican City, p. 219
Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189
Fall S. M., Rees M., 1985, ApJ, 298, 18
Frenk C., White S. D. M., Davis M., Efstathiou G., 1985, Nat, 317, 595
Henry J. P., Arnaud M., 1991, ApJ, 372, 410
Hernquist L., Katz N., 1989, ApJS, 70, 419
Jones C., Forman W., 1992, in Fabian A. C., ed., Clusters and Superclusters of Galaxies. Kluwer, Dordrecht, p. 49
Katz N., Gunn J., 1991, ApJ, 377, 365
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lacey C. G., Guiderdoni B., Rocca-Volmerange B., Silk J., 1993, ApJ, 402, 15
Navarro J. F., Benz W., 1991, ApJ, 380, 320
Navarro J. F., White S. D. M., 1993, MNRAS, 265, 271 (Paper I)
Peebles P. J. E., 1982, ApJ, 258, 415
Peebles P. J. E., 1984, ApJ, 277, 470
Persic M., Salucci P., 1992, MNRAS, 258, 14
Quinn P., Zurek W., 1988, ApJ, 331, 1
Rees M., Ostriker J., 1977, MNRAS, 179, 541
Silk J., 1977, ApJ, 211, 638
Smoot G. et al., 1992, ApJ, 396, 1
Spitzer L., 1968, Diffuse Matter in Space. Wiley, New York
Steinmetz M., Müller E., 1993, A&A, 268, 391
Toth G., Ostriker J. P., 1992, ApJ, 389, 5
Walker P. N., Steigman G., Schramm D. N., Olive K. A., Kang H. S., 1991, ApJ, 376, 51
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
White S. D. M., Efstathiou G. P. E., Frenk C. S., 1993a, MNRAS, 262, 1023
White S. D. M., Navarro J. F., Evrard A. E., Frenk C. S., 1993b, Nat, 366, 429
Zaritsky D., White S., 1994, ApJ, in press