Mon. Not. R. astr. Soc. (1992) **257**, 11–31

Simulating the formation of a cluster of galaxies

P. A. Thomas[1] and H. M. P. Couchman[2]★

[1] *Astronomy Centre, MAPS, University of Sussex, Falmer, Brighton BN1 9QH*
[2] *Department of Astronomy, University of Toronto, Toronto, Ontario M5S 1A1, Canada*

Accepted 1992 January 14. Received 1991 December 2; in original form 1991 August 2

SUMMARY

We present simulations (P^3M+SPH) containing both dark matter and gas, of the formation of a cluster of galaxies. An efficient potential solver, SPH algorithm and cooling implementation make for an extremely fast scheme. The numerical two-body relaxation time is shown to constrain severely the resolution which can be attained. The cluster in our simulations forms by the merger of three large groups at which time relaxation creates an isothermal mass distribution with no central core. Subsequent infall of small groups does not disturb this structure but gives an outer radius which grows uniformly with time. The density and temperature of the gas fall in the outer regions of the cluster but the combination $T/n^{2/3}$, which is related to the entropy, or shock heating, rises linearly with radius before dropping steeply at the edge of the cluster.

The morphology of the cluster is not determined by the initial shape of the overdensity which gives rise to it, but is strongly affected by the large-scale tidal field. The collapse is well fitted by a simple top-hat model up to the point of maximum expansion but is then slowed either by the presence of substructure or because of the non-spherical shape. Press–Schechter theory underestimates the number density of massive clusters. Pre-virialization does not appear to diminish the degree of collapse.

Fitting an isothermal β-model to the density profiles of the cluster and other large groups gives $\beta_{\rm fit} \approx \frac{2}{3}$. There is incomplete thermalization of the cluster gas but this may well be a numerical artefact arising from inadequate resolution of shocks. After correcting for this, the ratio of dark matter to gas energies is $\beta \approx 1$. We show that the difference between β and $\beta_{\rm fit}$ is consistent with a hydrostatic model. However, this does not explain the long-standing β-discrepancy which, because of the steeper decline of galaxy density with radius, predicts a ratio of galaxy to gas energies of $\beta_{\rm spec} \approx \beta_{\rm fit}$, contrary to observations: possible resolutions of this problem are discussed.

When cooling is added to our simulation the intracluster medium becomes multiphase within the cooling radius and a large quantity of cool gas accumulates at the cluster centre. The gas temperature rises steeply from 10^4 K in low-mass haloes to the virial temperature of 10^7 K or more in high-mass ones, in agreement with simple hierarchical models of galaxy formation.

Subgroups can survive within the virialized portion of the cluster. They tend to be on radial orbits with a velocity dispersion (internal plus external) equal to the cluster mean. Mass estimators based on the virial theorem underestimate the total cluster mass and have large variance. A more stable statistic is the projected mass estimator of Heisler, Tremaine & Bahcall. The relationship between the X-ray luminosity and the mass function of clusters remains obscure but where the surface brightness profile can be measured the mass profile of a cluster is well constrained.

Finally, we discuss several ways in which the simulations presented in this paper can be extended and improved.

★ Present address: Department of Astronomy, University of Western Ontario, London, Ontario N6A 3K7, Canada.

1 INTRODUCTION

1.1 Motivation

This paper reports the results of N-body simulations of the formation of a rich cluster of galaxies. Both dark matter and gas are included so as to model correctly the intracluster medium. This is an increase in complexity from pure gravity codes which are sufficient, in standard models, to describe the large-scale flow of matter.

Clusters of galaxies are interesting because they provide strong discriminants between cosmologies and they are perhaps the simplest condensed systems we can study. Rich clusters are sites of strong X-ray emission, principally via bremsstrahlung radiation from the hot intracluster medium. They can be detected at high redshift and their number density can be used to constrain the amplitude of fluctuations on these scales. To do this, however, we have to understand the structure of the intracluster medium so that we can relate the observed luminosity and temperature functions to the underlying mass function of clusters.

On large scales, gravity plus the linear theory of perturbations are sufficient to describe the density distribution and velocity field. Clusters are non-linear but their structure is relatively simple: they are virialized in their centres and the outer regions are well described by a simple infall model. Similarly, the physics of the intracluster medium is relatively straightforward. Here we consider a simple model in which shock heating and radiative cooling are the only important processes.

There have been few simulations of this kind in which the gas has been included as an integral part of the code. The most important previous work is by Evrard (1990) who performed calculations very similar to the ones presented here but with fewer particles. Our results are in broad agreement with his but differ in details because of our higher resolution. We discuss to what extent the finite particle number, particularly its effect in limiting the gas force resolution, still constrains the results: currently the uncertain history of the intracluster medium is just as important. The simulations provide insight into a number of cluster properties, as detailed in the summary, but can only act as guides to theoretical models. In the future, however, it seems that the main uncertainties will lie not in the N-body simulations themselves but in the physical assumptions that go into them.

In Section 2 we first describe the numerical method and then discuss the choice of gravitational softening and its relation to the two-body relaxation time in the cluster core. The results of the main simulation are presented in Section 3 while those of the comparison runs and other topics of interest are discussed in Section 4.

1.2 The simulations

Most of the results presented in this paper come from just one simulation of the formation of a rich, Abell-type cluster. We do not aim to mimic any particular cosmology but rather to investigate general features of any hierarchical model of cluster formation. For definiteness, the initial power spectrum is taken to be that of the standard cold dark matter theory (CDM, e.g. Bond & Efstathiou 1984) with $h_{50} = H_0/50$ km s^{-1} Mpc$^{-1} = 1$ and a density parameter $\Omega = 1$. We have no particular prejudice in favour of this choice but it is a popular cosmology which has moderate power on both small and large scales. The box side is taken to represent 50 Mpc, large enough to contain a reasonably massive cluster whilst giving resolution down to 40 kpc which is much smaller than the expected cluster core size. With 32^3 gas and an equal number of dark matter particles, and with a baryon mass fraction $\Omega_b = 0.1$, this gives particle masses of 2.64×10^{10} and $2.37 \times 10^{11}\ M_\odot$, respectively. The choice of baryon density has minor significance as the dark matter dominates the gravitational potential. Choosing $\Omega_b = 0.05$, in closer agreement with recent limits from primordial nucleosynthesis, would have little effect on our simulations except to make cooling slightly less important.

We normalize to fluctuations in a sphere of radius 16 h_{50}^{-1} Mpc which are deemed to have a root-mean-square (rms) value in linear theory of 0.55 today, where we use a time unit of $t_0 = 2/3H_0^{-1} \approx 1.30 \times 10^{10}\ h_{50}^{-1}$ yr so that $t = 1$ at the endpoint of the simulation; the initial time is 8.0×10^{-3} in these units. This normalization corresponds to a bias of 1.7 ($= 0.935/0.55$) although once again the precise value would make little difference to the results. The time unit must be multiplied by $(b/1.7)^{1.5}$ to convert to a different bias factor, but note that cooling does not scale self-similarly.

In a simulation on this scale there is significant power from fluctuations with wavelengths larger than the box; we allow for this by scaling the box size and effective Hubble constant as in an overdense universe. The initial overdensity $\Omega_{\rm box} - 1 \approx \delta M/M|_{\rm box}$ is chosen to be 0.0312 which is approximately 2.3 times the rms fluctuations on that scale. The maximum fluctuation within the box is likewise 2.3 times the rms on a scale of approximately 16 Mpc, which is larger than any bound halo we will consider. Therefore, although the box overdensity is significant, it never dominates. The box would stop expanding at time 1.58 in the above units.

Initial conditions were generated by populating a regular 32^3 cubic grid with independent modes and then Fourier transforming to obtain a potential which was used to calculate the Zel'dovich particle displacements (Efstathiou *et al.* 1985). Individual Fourier modes had amplitudes set to the square root of the power spectrum with phases drawn randomly from a distribution uniform in $[0, 2\pi]$. The Fourier modes were corrected for the interpolation and differencing errors that arise due to the finite grid resolution as described in Carlberg & Couchman (1989). The initial, unperturbed particle distribution was uniform to minimize the effects of shot noise on small scales. The resulting power spectrum of the particle distribution is an accurate representation of the CDM spectrum up to the Nyquist frequency, here $\frac{1}{16}k_0$, where $k_0 = 2\pi/50$ Mpc^{-1} is the fundamental frequency. We checked the effectiveness of this procedure by smoothing with spherical top-hat filters containing 500, 1000 and 2000 particles of each kind and comparing the distribution of measured overdensities with a Gaussian: there is good agreement. On the scale of 2000 particles the cluster shows up as a rare 3.1σ fluctuation which we positioned at the box centre, for convenience.

We have also performed two other simulations, with slightly different parameters, in order to investigate the effect of various physical changes on our results. Of these, one is identical except that $\Omega_{\rm box} = 1$, and in the other the gas is allowed to cool radiatively with the emission spectrum of a 0.3-solar metallicity gas. The three simulations are labelled

Table 1. A summary of the differences between our three simulations.

	initial overdensity	cooling
S1	**0.0312**	**no**
S2	**0**	**no**
S3	**0.0312**	**yes**

S1, S2 and S3 and their differences are summarized in Table 1.

2 THE NUMERICAL METHOD

Our simulations evolve $32^3 = 32\,768$ 'gas', or baryonic, particles under the action of gravity and pressure forces, and an equal number of 'dark' matter particles under gravity only. The geometry is that of a cubical box with periodic boundary conditions on opposite faces.

2.1 Calculating the forces

Gravitational forces are evaluated using a particle-mesh (PM) potential solver with a softening length of 3 grid units which is supplemented by interparticle (PP) forces at smaller separations. The combined P^{3}M pairwise force is accurate in magnitude to within 10 per cent (rms error <4 per cent for all separations) and in direction to within 4° (rms error $<2°$). The force is implemented so as to conserve linear momentum. A 64^3 grid, finer than the particle mesh, is used in order to speed up the calculation of forces from neighbours in highly clustered regions. Further details can be found in Couchman (1992, in preparation).

Gas forces are calculated using smooth-particle hydrodynamics (SPH) (e.g. Gingold & Monaghan 1977; Lucy 1977). The standard technique is admirably described by Hernquist & Katz (1989) and is used here except as detailed below.

We use a smoothing kernel

$$W(\boldsymbol{r}_1, \boldsymbol{r}_2; h) \equiv \frac{W(|\boldsymbol{r}_1 - \boldsymbol{r}_2|/h)}{h^3} = \frac{21}{256\pi h^3}(2-x)^2(4-x^2),$$

$$x = \frac{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}{h} < 2, \tag{1}$$

where we continually adjust the smoothing length, h to average over approximately 32 neighbouring particles subject to a minimum value set equal to the gravitational softening (see below) and a maximum equal to 2 grid units (for numerical convenience: the SPH force at this separation is negligible).

The adopted smoothing kernel is similar to the standard one:

$$W_s(x) = \frac{1}{4\pi}\begin{cases} 4 - 6x^2 + 3x^3, & x \le 1, \\ (2-x)^3, & 1 < x \le 2, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

except that the absolute value of its gradient is monotonic decreasing with increasing x. We choose to satisfy this condition because we noticed in trial runs, using W_s to mimic a uniform, non-gravitating gas, that it was possible for particles to clump together artificially because the pairwise pressure force goes to zero linearly at small separations. This may be a problem in certain kinds of simulations, notably those which identify small groups of three or four particles with galaxies! An alternative and perhaps better way to prevent this happening is to add a conservative repulsive force at small separations to force the gradient to be monotonic,

$$\frac{dW_s'(x)}{dx} = -\frac{1}{4\pi}\begin{cases} 4, & x \le \frac{2}{3}, \\ 3x(4-3x), & \frac{2}{3} < x \le 1, \\ 3(2-x)^2, & 1 < x \le 2, \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

whilst leaving W_s unaltered. The most efficient way to calculate the value of the kernel and its gradient is to interpolate from a table and so their formal complexity does not alter the running time.

In order to conserve momentum we make use of the following identity,

$$\frac{\nabla P}{\rho} = \nabla\left(\frac{P}{\rho}\right) + \frac{P}{\rho^2}\nabla\rho, \tag{4}$$

to write the pressure force on particle i as

$$\boldsymbol{f}_i = \sum_j \boldsymbol{f}_{ij} = -\sum_j m_i m_j \frac{P_i}{\rho_i^2}\nabla_i W(\boldsymbol{r}_i, \boldsymbol{r}_j; h_i) + \sum_j m_i m_j \frac{P_j}{\rho_j^2}\nabla_j W(\boldsymbol{r}_i, \boldsymbol{r}_j; h_j), \tag{5}$$

where m_i is the mass, P_i the pressure and ρ_i the density of particle i. Hence it is only necessary to loop over the particles twice, once to evaluate the short-range gravitational force, density, pressure gradient and velocity divergence at a given particle position, and a second time to apply equal and opposite SPH forces to the particle and each of its neighbours.

We use a perfect-gas equation of state, $P_i = \frac{2}{3}\rho_i\varepsilon_i$ where ε_i is the specific energy, related to the gas temperature T_i by $\varepsilon = 3kT/2\mu m_H$ where k is Boltzmann's constant, m_H is the mass of the hydrogen atom and the relative molecular mass $\mu = 0.6$. Energy conservation is ensured by writing the rate of change of specific energy ε in the form

$$\frac{d\varepsilon_i}{dt} = -\sum_j \frac{\varepsilon_i}{m_i\varepsilon_i + m_j\varepsilon_j}(\boldsymbol{v}_i - \boldsymbol{v}_j)\cdot\boldsymbol{f}_{ij} - \text{cooling terms}. \tag{6}$$

Radiative cooling is implemented in integrated form assuming constant density. The change in specific energy, $\Delta\varepsilon$, is given by

$$\int_{\varepsilon}^{\varepsilon-\Delta\varepsilon} \frac{d\varepsilon}{\Lambda} = -\frac{n_i^2}{\rho_i}\Delta t, \tag{7}$$

where Δt is the time-step, n_i is the number density and $n_i^2\Lambda$ is the power radiated per unit volume. This ensures that cooling never limits the time-step.

SPH also requires an artificial viscosity which is used to dissipate convergent motions. We simply add an extra component to the pressure used in the force calculation:

$$P \leftarrow P + \rho[-1\cdot c_s h\nabla\cdot\boldsymbol{u} + 2\cdot(h\nabla\cdot\boldsymbol{u})^2], \qquad \nabla\cdot\boldsymbol{u} < 0, \tag{8}$$

where $c_i = \sqrt{5P_i/3\rho_i}$ is the sound speed and the constants are empirically chosen so as to give good jump conditions in a compressive shock with the minimum of interpenetration. We tested this formula using one-dimensional shocks (adiabatic and isothermal) in a three-dimensional simulation (this differs from a one-dimensional simulation in that the particles are not aligned so as to collide but may slip around one another). Some interpenetration does occur but the correct jump conditions are attained with a shock width of about 6 particles. *This means that in a spherically symmetric collapse approximately 1000 particles are required for efficient conversion of kinetic energy into heat*; also many particles may avoid passing through a complete shock if they have a finite impact parameter.

2.2 The integration method

Because shocks can lead to discontinuous velocities, densities and temperatures, we use a low-order integration method which does not rely on stored information: this also helps to reduce memory requirements. To step forward by time Δt we make use of one intermediate time-step $t_\theta = t_0 + \theta\Delta t$. First the acceleration, $\boldsymbol{a}_0(\boldsymbol{r}, \boldsymbol{u}, m, \varepsilon)$, and cooling rate, $\dot{\varepsilon}_0$, are evaluated and used to update to t_0:

$$\boldsymbol{v} \leftarrow \boldsymbol{v} + \boldsymbol{a}_0 \theta \Delta t, \tag{9a}$$

$$\varepsilon \leftarrow \varepsilon + \dot{\varepsilon}_0 \theta \Delta t, \tag{9b}$$

$$\boldsymbol{r} \leftarrow \boldsymbol{r} + \boldsymbol{v}\theta\Delta t - \boldsymbol{a}_0 \tfrac{1}{2}(\theta\Delta t)^2. \tag{9c}$$

The acceleration, $\boldsymbol{a}_0$, and cooling rate, $\dot{\varepsilon}_\theta$, are then re-evaluated and used to calculate the solution after the full time-step:

$$\boldsymbol{v} \leftarrow \boldsymbol{v} + \left\{\boldsymbol{a}_0\left[(1-\theta) - \frac{1}{2\theta}\right] + \boldsymbol{a}_\theta \frac{1}{2\theta}\right\}\Delta t, \tag{9d}$$

$$\varepsilon \leftarrow \varepsilon + \dot{\varepsilon}_\theta (1-\theta)\,\Delta t, \tag{9e}$$

$$\boldsymbol{r} \leftarrow \boldsymbol{r} + \boldsymbol{v}(1-\theta)\,\Delta t - \frac{1}{6\theta}\{\boldsymbol{a}_0[2(1-\theta)^3 - \theta^3] + \boldsymbol{a}_\theta[3\theta - 2]\}(\Delta t)^2. \tag{9f}$$

We set $\theta = \frac{1}{2}$ which when $\boldsymbol{a}$ is a function of $\boldsymbol{r}$ and m only (i.e. no shocks or strong cooling) gives r correct to terms of order $(\Delta t)^5$. The ε-equation is necessarily of lower order to allow for the integrated form of the cooling, described above. We have tested the ability of our scheme to integrate simple systems such as a single particle in a simple harmonic oscillator potential or two particles in circular orbits about one another, and these are followed with high accuracy.

2.3 The softening length and two-body relaxation times

The softening length in our simulations is chosen to give a balance between small-scale resolution and a reasonable time-step, while at the same time ensuring that two-body interactions are negligible. We discuss the various constraints in some detail here because we feel that they are often neglected and yet are crucial for interpreting the simulations.

The characteristic minimum time-scale for interactions under gravity with a particle of mass m and with a minimum effective impact parameter equal to the softening s is

$$t_2 = \left(\frac{s^3}{Gm}\right)^{1/2} \approx 0.018 \left(\frac{s}{40\ \mathrm{kpc}}\frac{50\ \mathrm{Mpc}}{box}\right)^{3/2} \left(\frac{N_\mathrm{p}}{32^3\Omega_\mathrm{p}}\right)^{1/2} t_0, \tag{10}$$

where $t_0 \approx 1.30 \times 10^{10}\ h_{50}^{-1}$ yr is the current age of the Universe, *box* is the box size, N_p is the total number of relevant particles and Ω_p is their density parameter. Associated with this is a two-body temperature

$$T_2 = \frac{\mu m_\mathrm{H}}{k}\frac{Gm}{s} \approx 2.1 \times 10^6\ \mathrm{K}\ h_{50}^2 \left(\frac{40\ \mathrm{kpc}}{s}\frac{box}{50\ \mathrm{Mpc}}\right)^3 \left(\frac{32^3\Omega_\mathrm{p}}{N_\mathrm{p}}\right). \tag{11}$$

The softening length must be chosen so that T_2 is lower than the virial temperature of any halo of interest.

The relaxation time, t_r, is the characteristic time in which particles undergo a significant change in velocity due to two-body interactions. For a uniform distribution of N particles contained within a volume $\frac{4}{3}\pi R^3$ and with velocity dispersion $v^2 \sim \gamma GmN/R$ (see, e.g. Binney & Tremaine 1987, p. 187),

$$t_\mathrm{r} \sim \frac{v^3 R^3}{6(Gm)^2 N \ln(R/s)} \sim \frac{\gamma^{3/2} N^{1/2}}{6\ln(R/s)}\left(\frac{R^3}{Gm}\right)^{1/2} \sim \frac{\gamma^{3/2}N^{1/2}}{6\ln(R/s)}\left(\frac{R}{s}\right)^{3/2} t_2. \tag{12}$$

Note that this time-scale is mainly determined by the total particle number and for $R \gg s$ depends only logarithmically on the smoothing length. The strongest constraint (minimum time-scale) comes from particles at the core of the halo where the density is highest. We can treat the core as isolated because gravitational perturbations from particles at larger radii will cause bulk motion rather than alter the peculiar velocities; the maximum possible effect is anyway small compared to that from the core itself. For an isothermal sphere we have $v^2 = 6\pi G\rho_0 R^2 \approx \frac{9}{2} GmN/R$ where ρ_0 is the central mass density and we have assumed a uniform mass distribution within the core radius R. Then

$$\frac{t_\mathrm{r}}{t_0} \sim 0.03 \frac{(R/s)^{3/2}}{\ln(R/s)} N^{1/2} \left(\frac{s}{40\ \mathrm{kpc}}\frac{50\ \mathrm{Mpc}}{box}\right)^{3/2} \left(\frac{N_\mathrm{p}}{32^3\Omega_\mathrm{p}}\right)^{1/2}. \tag{13}$$

Observed clusters have core radii of order 400 h_{50}^{-1} kpc (Sarazin 1986), i.e. $R \approx 10s$, which gives $t_\mathrm{r}/t_0 \sim 0.4N^{1/2} \gg 1$ as required. The clusters in our simulations, however, have core radii which are equal to the softening length. The above formula is minimized (for fixed N/R) when $R = \sqrt{e}\,s$, giving an initial coefficient in front of the scaling factors in equation (13) of $0.12N^{1/2}$: the relaxation time is therefore short in the core. We would like to ensure that this ratio is never less than unity in galaxy-sized haloes, however this would require an unreasonably large softening, $s > 160$ kpc, which would make it difficult to resolve cluster cores anyway, or an order of magnitude increase in the number of particles. At the time of writing no published cluster simulations have a sufficient number of particles to resolve the core, and some are dominated by numerical relaxation throughout.

2.4 Time-steps

It is crucial in any N-body simulation that the required forces should be accurately represented, that transport processes should be kept to a minimum and that gas particles should not be allowed to free-stream through one another. These effects are controlled by the number of particles, the form of the interparticle forces, discussed above, and the time-step. We identify several different time-scales in our simulation, the most important of which is the time over which significant displacements or changes in velocity due to an acceleration can occur:

$$t_a = \begin{cases} \min_i \left(\dfrac{s^2 + \min_j \Delta r_{ij}^2}{a_i^2} \right)^{1/4}, & \text{gravity,} \\ \min_i \left(\dfrac{h_i^2}{a_i^2} \right)^{1/4}, & \text{SPH,} \end{cases} \tag{14}$$

where Δr_{ij} is the particle separation. This is the major condition which helps to conserve energy. Even in the absence of interparticle forces, the relative positions of particles will change on a time-scale

$$t_v = \begin{cases} \min_{ij} \left(\dfrac{s^2 + \Delta r_{ij}^2}{\Delta v_{ij}^2} \right)^{1/2}, & \text{gravity,} \\ \min_{ij} \left(\dfrac{h^2}{\Delta v_{ij}^2} \right)^{1/2}, & \text{SPH.} \end{cases} \tag{15}$$

This is similar to the Courant condition in grid-based codes but is not required here for mass or energy conservation. We do not restrict the time-step to be less than t_v, as this gives very short time-steps in the cluster core where the changes in geometry are anyway random in nature. This may lead to particle diffusion, although none is obvious in our simulations.

The shortest time-scale is the cooling time in the cluster core,

$$t_\varepsilon = \min_i \left(\frac{\varepsilon}{|\dot{\varepsilon}|} \right). \tag{16}$$

We have avoided the severe limitation this would impose by implementing cooling in an integrated form, as described in equation (7).

Finally, in the initial stages of the expansion it is the instantaneous Hubble time which is the limiting factor,

$$t_H = \frac{a}{|\dot{a}|}. \tag{17}$$

In our simulations we use a variable time-step

$$\Delta t = 0.5 \min(t_a, 0.25 t_H). \tag{18}$$

Test simulations of the collapse of a uniform spherical ball of particles conserve energy to 1 part in 10^4 and follow a simple top-hat model down to particle separations of order the softening. With gas shocks and cooling included, the energy change during collapse and virialization of a cold gas sphere is a few per cent, sufficient for our purposes.

3 RESULTS OF THE SIMULATION

In this section we describe the results of the basic simulation, S1, where the cluster forms in an overdense box with no radiative cooling of the gas.

3.1 The growth of structure

We first discuss the baryonic particles which, because of the gas pressure, are distributed much more smoothly than the dark matter and more closely trace the underlying potential.

Fig. 1 shows the growth of the density, n, temperature, T, and $T/n^{2/3}$, averaged over all gas particles, as a function of time. This latter quantity is an increasing function of the specific entropy

$$s = (3k/2\,\mu m_H) \ln(T/n^{2/3}\,\mathrm{K\,cm^2}) + \text{constant},$$

which is related to the dissipational heating of particles. The density decreases at first with the cosmic expansion but then rises above it and by $t \approx 0.21$ has saturated at a constant value which suggests that a large fraction of particles are beginning to fall into overdense regions. Two-fifths of the particles have been heated to above 10^4 K at this stage and the fraction with measured density $n > 5.6\bar{n}$, corresponding to regions which have stopped expanding, is 0.55. These groups gradually merge with time, until at the end of the simulation only 15 per cent of the gas particles have temperatures below 10^4 K, and almost all lie in bound groups. Large groups and clusters grow by addition of subgroups with a wide range of masses, ranging from individual particles up to equal-sized partners. This can be seen in Fig. 2 which shows a projection through the box at four different times. At late times the mean density is dominated by the particles in the largest groups. The jump in density at $t \approx 0.55$ is due to the merger of three subgroups to form the central cluster (Fig. 3), while that at $t \approx 0.9$ results from the collapse of the other major groups.

The evolution of the central cluster is shown in more detail in Fig. 3 for both the gas and the dark matter. The variation of mean gas density, shown in Fig. 4, resembles that

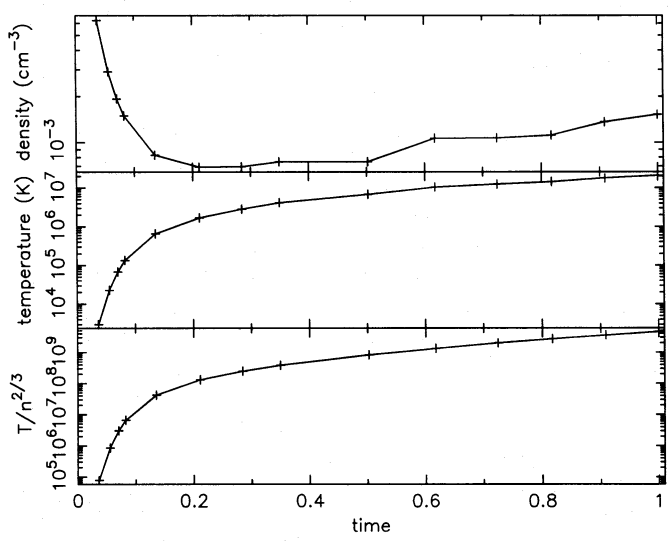

Figure 1. Mean properties of the gas distribution in simulation S1 as a function of time. The three panels show density temperature and an entropy-related quantity.

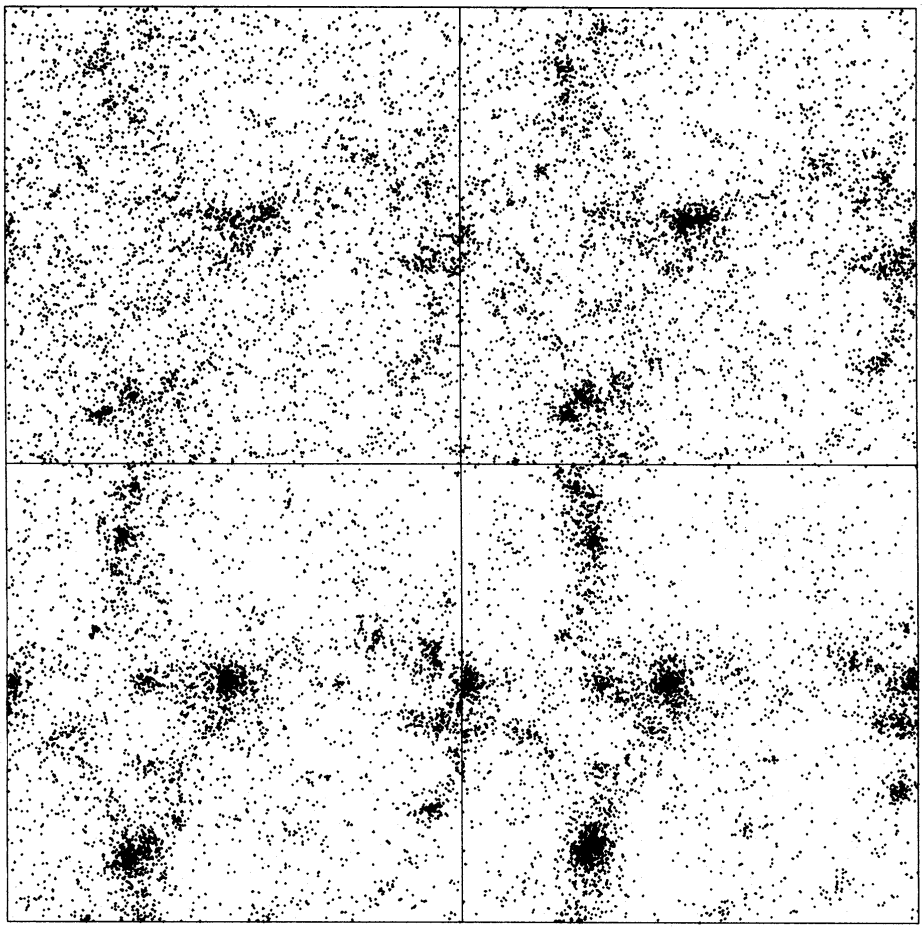

Figure 2. Projections through the gas distribution in simulation S1, showing one eighth of the particles, at times 0.35, 0.50, 0.72 and 1.00.

of the simulation as a whole. The main merger event can clearly be seen and comprises approximately half the total number of particles which eventually end up in the virialized core of the cluster. To investigate how this merger redistributes the mass, we also show in Fig. 4, for comparison, the evolution of the gas particles which lie in virialized regions just before the merger, at time $t=0.50$. After the merger many of these particles are ejected and their mean density drops substantially. The particles which were tightly clustered in the core at $t=0.50$ are, by the end of the simulation, distributed out to a distance of over 4 Mpc from the cluster centre, well outside the virialized region (see Fig. 5). We conclude that some sort of impulsive relaxation is the main process involved, for three reasons: most of the mass ejection occurs immediately following the merger event, the particles are ejected asymmetrically, and the number of particles involved greatly exceeds the number for which $t_r < t_0$ in the core (see Section 2.3). This is not violent relaxation in the usual sense, i.e. driven by a rapidly varying potential, because the dark matter is not affected as much as the gas. There is a small transfer of energy between particles but only one tenth of the dark matter particles escape. The main energy transfer comes about via shocks in the gas, with the quarter of the gas particles which escape being raised to significantly higher entropy than those which remain. The slow rate of mass ejection which continues after the merger until the end of the simulation may be artificial, as the two-body relaxation time is shorter than a Hubble time within the core of the cluster (see Section 2.3).

The core density of the central cluster is boosted greatly during the merger and the resultant density profile (Fig. 6a) already has a core radius which is much smaller than those observed in present-day clusters (typically 200–400 kpc, Jones & Forman 1984; Sarazin 1986). The subsequent slow evolution in density within approximately 100 kpc, which results in an increased core density and decreased core radius, may be due to artificial two-body relaxation.

The extent of the virialized region, as measured by the position of the entropy maximum (Fig. 6c), increases steadily with time after cluster formation, at a rate of about 230 km s^{-1} and reaching a total radius of 2–3 Mpc by the end of the simulation. Within the virialized region the entropy maintains a steeply rising entropy profile ($T/n^{2/3} \propto r$), which is much smoother than either the density or temperature profiles (Fig. 6b). Outside the virialized region the temperature of the gas drops rapidly, indicating that much of the infalling material is in systems of low virial temperature. However, even within this region the temperature is not constant, as is often supposed: it is low near the centre as the gas here was born within, and has not had to fall down into, the potential well of the cluster. On the other hand, the temperature declines outwards from about 200 kpc. It is of interest that the infalling gas does not appear to mix efficiently into the virialized region but maintains the low central temperature

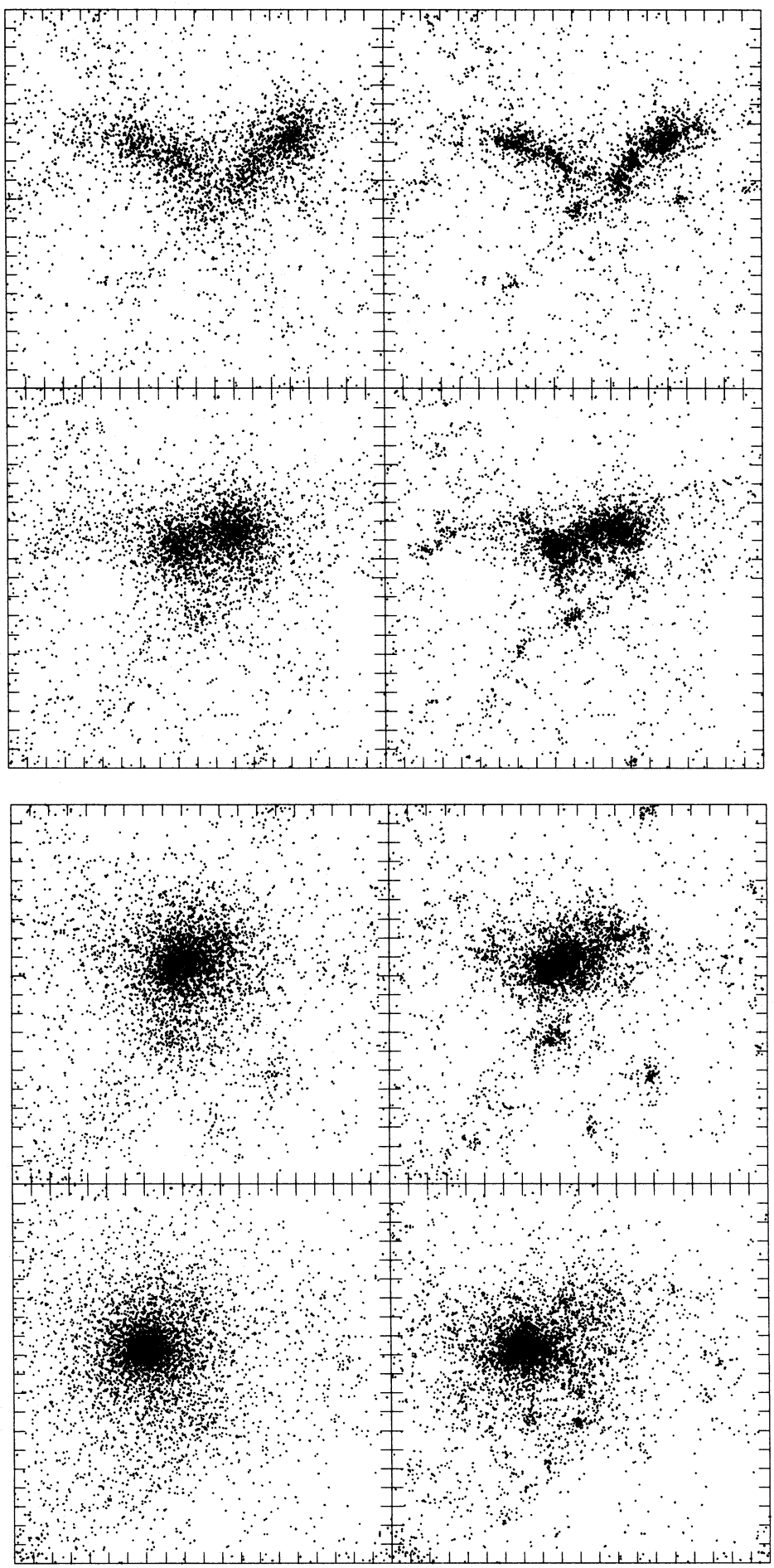

Figure 3. Projections through the central regions of simulation S1 at four different times: from top to bottom 0.35, 0.50, 0.72 and 1.00. On the left is the gas distribution, on the right the dark matter for comparison.

and the entropy gradient. Particles do not stay at a fixed radius once accreted, however, but make excursions out to larger radii (see Fig. 8) so that it is not true to say that material gets added in neat spherical shells.

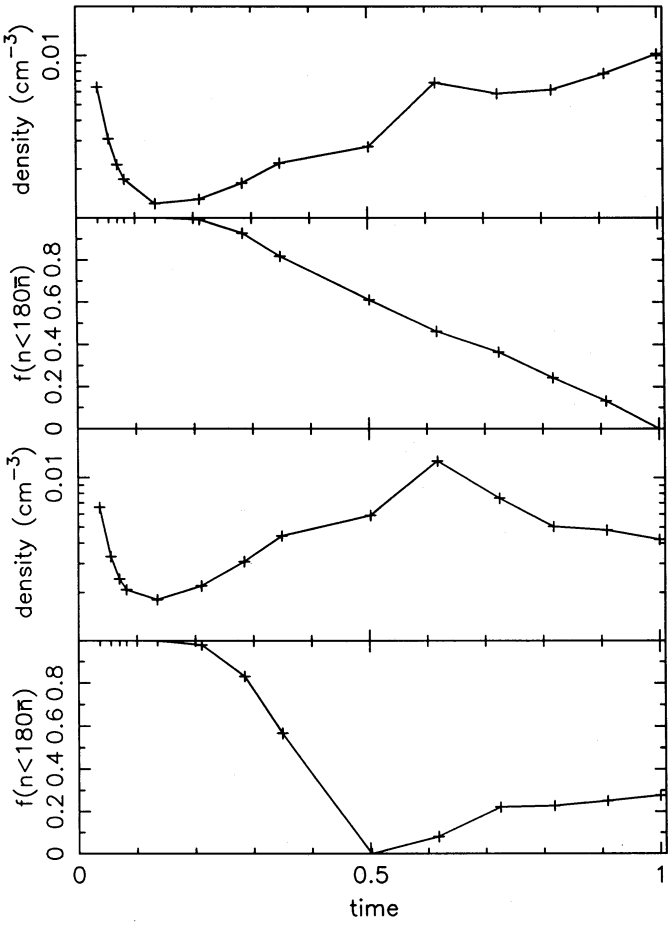

Figure 4. Evolution of the mean density and the fraction of particles outside the virialized region (i.e. overdensity less than 180). The upper and lower two panels are for particles which lie in virialized regions near the centre of the simulation at times 1.00 and 0.50, respectively.

The dark matter is clustered on a much smaller scale than the gas. This can be seen both in the comparison panels in Fig. 4 and in a plot of the high-density regions surrounding the cluster shown (for simulation S3) in Fig. 12, later. The subclumping is also apparent in the density profile across the cluster: the gas has very similar mass-weighted and volume-weighted density profiles, whereas for the dark matter these differ considerably. The difference is not caused by the elongated nature of the dark matter distribution, as the same result holds if we bin the density in ellipsoidal bins (see the next section), but is due to the presence of high-density substructure. Thus it is not true to say that the dark matter haloes merge to give a uniform distribution of matter; in this non-radiative model, it is the gas which is uniformly distributed.

The general picture of cluster evolution which emerges is slow accretion of small haloes which do not significantly disturb the cluster's structure, punctuated by large merger events of nearly equal-sized bodies which redistribute material by impulsive relaxation. When averaged in spherical shells, the cluster radius seems to increase uniformly in time by the accretion of gas which gets shocked to ever higher entropy as it impacts the outer regions of the cluster.

3.2 The identification of haloes

We find haloes by constructing a minimal-spanning tree of all baryonic particles which lie above some threshold density (usually $n_h = 180$ times the mean density $\bar{n}$). Individual condensed regions are then identified by cutting the spanning tree at a linking length of $2/n_h^{1/3}$ – in practice any length intermediate between the particle separation and the halo separation will give the same partition. It is possible to apply SPH to the dark matter also (but without imposing any gas forces) in order to calculate the density and then construct haloes in the same way. Once a halo has been identified, we can trace its evolution by looking at the properties of its constituent particles at different times. Also, the divergent evolution of dark and gas particles (which start with identical positions and velocities) can be followed.

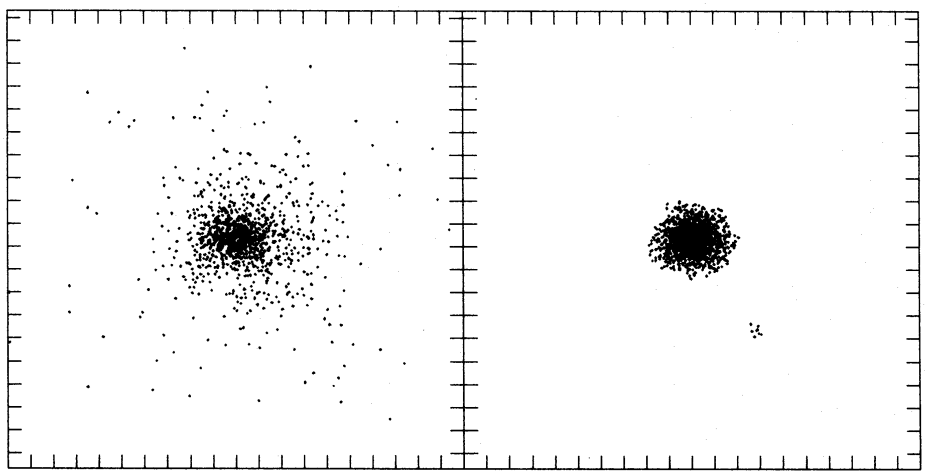

Figure 5. On the left is shown a projection at time 1.00 of those particles which were in virialized regions at the centre of simulation S1 at time 0.50. On the right for comparison is a projection of all the particles in virialized regions at the centre of simulation S1 at the final time.

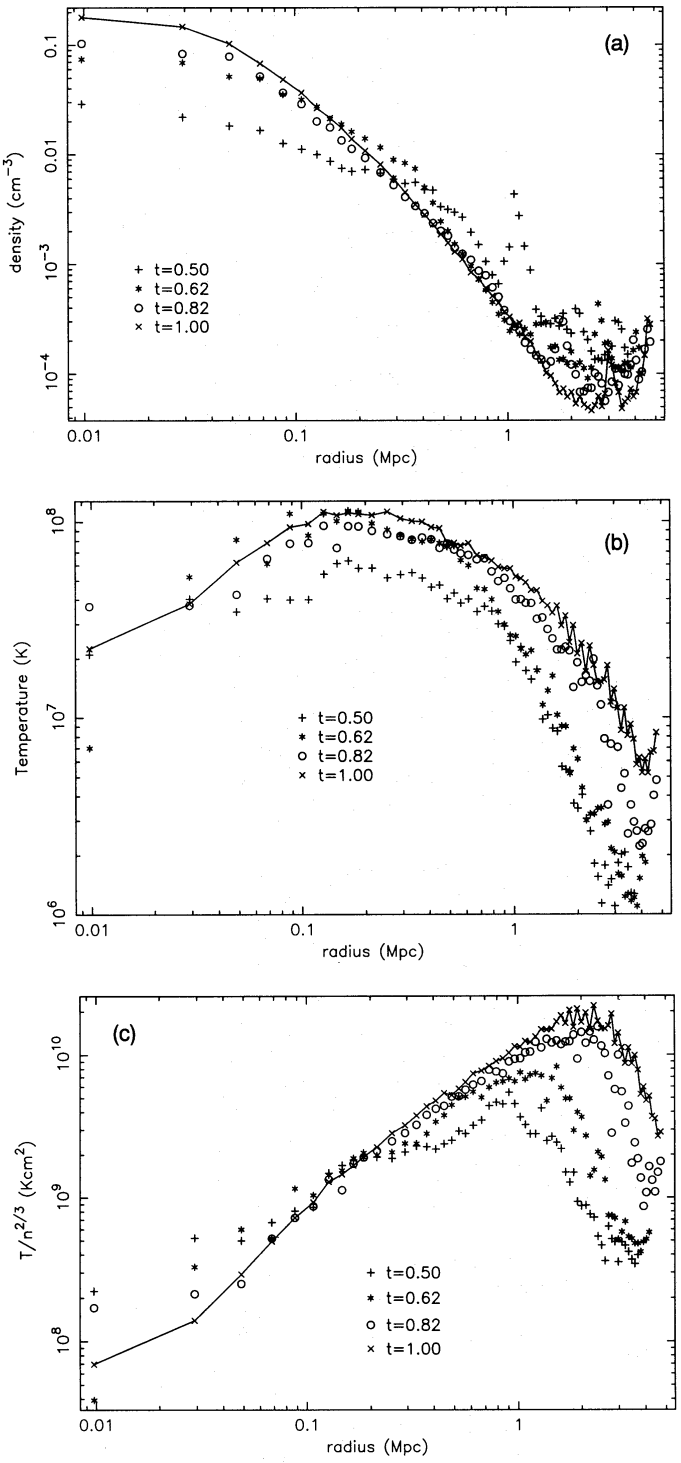

Figure 6. Gas profiles in spherical bins centred on the density maximum for simulation S1 at four different times: (a) density, (b) temperature, and (c) $T/n^{2/3}$ – an increasing function of entropy.

3.3 The geometry of the collapse

There has been some interest in the eccentricity of first-ranked cluster galaxies and the alignment of their major axes both with the cluster itself and with associated structure on larger scales (see, for example, Struble 1987; Lambas, Groth & Peebles 1988; Rhee & Roos 1989; Lambas *et al.* 1990; Fong, Stevenson & Shanks 1990). Such correlations are often interpreted as implying a 'pancake' model in which superclusters and clusters form before galaxies, however, this is not strictly required. An alternative explanation in the standard hierarchical cosmologies would be that the large-scale power determines the collapse rates in different directions via tidal fields, and that the final cluster retains a memory of this.

To measure the shape of a halo we first identify all constituent particles as described in Section 3.2. A mass-moment tensor,

$$I_{ij}=\sum r_i r_j - N_{\rm p} R_i R_j, \tag{19}$$

where $N_{\rm p}$ is the total number of particles and $R_i=N_{\rm p}^{-1}\sum r_i$ is the centre of mass, is then evaluated and the eigenvalues, $\lambda_1\le\lambda_2\le\lambda_3$, and orthonormal eigenvectors, $\boldsymbol{e}_l$ ($l=1, 2, 3$), of this real, symmetric matrix are found. For a uniform sphere of radius a, $I_{ij}=\frac{1}{5}N_{\rm p}a^2\delta_{ij}$: we therefore identify the semi-axes of the galaxy as $a_l=(5\lambda_l/N_{\rm p})^{1/2}$. Note that forming I_{ij} by summing inside isodensity surfaces gives a much more sensitive test (at early times) than the more usual procedure of summing over a spherical ball.

The axial ratios and directions for both the gas and dark matter in the cluster at the final time have been calculated for several different overdensity cuts, of which two are listed in Table 2. Note that the gas ellipsoids are less elongated because they are supported by isotropic gas pressure. Although we have only a single realization of a cluster here, it is prolate-triaxial, in agreement with the morphological analysis of clusters in the Lick catalogue by Plionis, Barrow & Frenk (1991). However, there is no obvious trend of increasing eccentricity with scale, such as that reported by Porter (1988) and reproduced by Rhee & Roos (1990) for dissipationless collapse from highly prolate initial conditions.

The principal axes of the gas and dark matter ellipsoids are not aligned, nor do the intermediate and minor axes always agree. The misalignment angles are about 20° in each of the principal directions at $n>180\overline{n}$, increasing to about 50° at higher overdensity as the intermediate and minor axes of the gas distribution swap roles. As this suggests, there is isophote twisting in that the principal axis directions rotate as the overdensity cut is varied. The maximum variations in direction between these two overdensities are approximately 30°, 30° and 20° for the major, intermediate and minor axes of the dark matter and 40°, 70° and 60° for those of the gas. The gas isodensity contours are close to spherical and so this rotation is perhaps not significant – it would be hard to detect in a projected surface density map. The most obvious interpretation of this rotation is that each shell of infalling material is subject to a different tidal torque at its period of maximum expansion: after collapse its size is much reduced and so is the tidal field across it.

The relatively small change in major axis direction for the dark matter distribution between overdensities of 180 and 1800 is in agreement with the observed alignment of brightest cluster galaxies with the major axis of the cluster. However, alignments are also seen on larger scales, out to 30 h_{50}^{-1} Mpc or more. It is hard to measure the tidal field in our simulation on such large scales, but that generated by the four most massive groups (clearly visible in Fig. 2) points predominantly in the y-direction and so is approximately aligned with the major axis of the cluster for an overdensity cut of 180. This result is no more than suggestive, as the

Table 2. Morphological parameters, evaluated as described in Section 3.3, for the cluster in simulation S1 at the final time. The columns give (i) overdensity selection criterion, (ii) type and number of particles, (iii) semi-intermediate axis length, (iv) axial ratios and (v)–(vii) direction cosines of the principal axes with the box sides.

$n/\bar{n}$	No. ptcles	a_2/Mpc	a_i/a_2	Direction cosines		
180	gas		1.16	0.165	0.915	0.367
	2037	0.66	1.00	-0.536	-0.229	0.813
			0.94	0.828	-0.331	0.453
	dark		1.49	0.468	0.828	0.309
	2623	0.75	1.00	0.558	-0.548	0.623
			0.83	0.685	-0.119	-0.719
1800	gas		1.16	0.712	0.691	0.125
	780	0.27	1.00	-0.702	0.699	0.134
			0.92	0.006	-0.183	0.983
	dark		1.55	0.682	0.641	0.352
	1012	0.26	1.00	-0.520	0.763	-0.383
			0.86	-0.514	0.078	0.854

separation of these groups is a significant fraction of the box size.

Next we smooth our initial conditions using a top-hat filter of the appropriate mass scale and look at the resulting density field. In general, we find that the density contours are highly anisotropic and do not correlate with those found at the end of the simulation, that is to say they do not contain those particles which end up in the cluster. For example, the density contour containing 2037 particles encloses only one peak, yet it extends all the way across the box, principally in the y–z plane, and the 2037 particles which end up in the virialized cluster at the final time are initially aligned with their long axis almost exactly perpendicular to this plane. Things are not quite so bad on smaller scales: for 414 particles these two distributions are aligned to within 18°. In linear theory the overdensity at each point in the density field (following the particles) grows at the same rate, so particles which are contained in the overdense region at one time remain so at later times. This will not hold true in the non-linear regime, as matter will fall in along the minor axis, following the collapse. Nevertheless we would expect these two distributions largely to agree. The difference we attribute to tidal forces which distort the isodensity surfaces. This should be more important on larger scales, as seen here, both because the intrinsic shape of overdense regions is more spherical and because there is longer for tides to act during their evolution (Binney & Silk 1979; Palmer 1983).

Fig. 7 shows the variation with time of the axial ratios of the ellipsoid defined by the particles which end up in the virialized cluster at various times. In each case the minor axis grows more rapidly than the intermediate axis, and on the two largest mass scales they exchange roles. this would be impossible for an isolated system (Icke 1984) and once again shows the importance of tidal forces. What is not shown in Fig. 7 is that the principal axis directions rotate under the influence of these tides. In all cases the long axis grows relative to the other two, by an amount which is relatively much larger for the smaller mass scales; these large eccentricities do not, however, seem to survive virialization which occurs first along the shorter axes causing a pancake and then a linear protocluster (see Fig. 3).

The angular momentum of the particles within a given isodensity contour should provide a clue to the dynamics. In all cases the specific angular momentum of the gas is greater

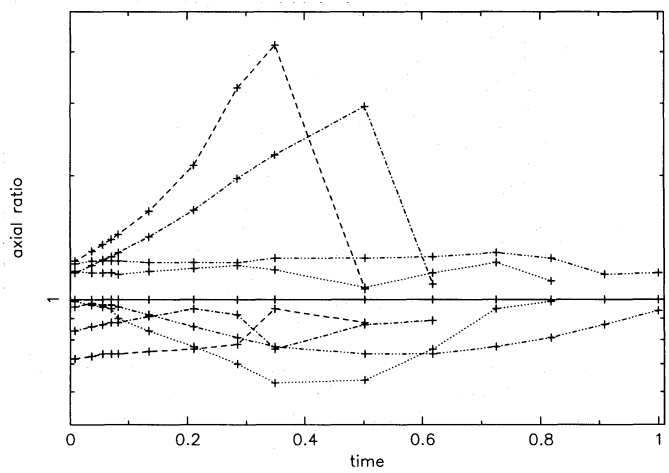

Figure 7. The variations of axial ratio with time for four different particle sets which correspond to virialized regions of the cluster in simulation S1 at times 0.35, 0.50, 0.82 and 1.00, containing 415, 1062, 1717 and 2037 particles, respectively.

than that of the dark matter, even though most of the random gas motions have been thermalized so that the specific kinetic energy is much lower. However, the specific angular momentum averaged within a spherical region is larger for the dark matter, and so this result merely indicates that the angular momentum of the dark matter is more nearly aligned with its major axis than that of the gas. In fact the direction of the angular momentum vector appears to range widely with no obvious pattern: there is no close alignment with any of the principal axis directions or of the gas and dark matter angular momenta with each other – in fact for an overdensity cut of 1800 the two angular momenta are almost exactly perpendicular! The rotational kinetic energy of the gas, assuming solid-body rotation with the mean angular momentum, is approximately one fifth of the total kinetic energy. This is sufficient to cause weak anisotropy, but is not aligned in such a manner as to explain the observed axial ratios. For the dark matter rotation is negligible. We conclude that the flattened shapes are maintained by anisotropic velocity dispersions as in giant elliptical galaxies. The low angular momentum found for this high-sigma cluster is consistent with the analysis of Hoffman (1986) and once again suggests alignment between the major axis of the cluster and the large-scale tidal field.

3.4 The formation time of the cluster and pre-virialization

The analytic models of Press & Schechter (1974, hereafter PS) and Peaks theory (Bardeen *et al.* 1986) are often used to estimate the number of haloes of a given mass in hierarchical clustering cosmologies. These can give reasonable fits to the number of haloes corresponding to large galaxies or small groups (e.g. Carlberg & Couchman 1989), however we are concerned here with the more massive objects arising from rare, highly overdense perturbations. It is such massive clusters which dominate the X-ray luminosity function and the Abell catalogue, and yet the two models predict highly differing number densities. We wish to investigate this discrepancy. We will first review the basic idea behind each scheme and then compare each with the collapse rate of haloes in our simulations.

Both PS and Peaks theory first smooth the initial Gaussian random field on the mass scale, m. PS then counts the fraction, $f(m, \nu)$, of the universe exceeding a particular density threshold $\delta = \nu\sigma$, where σ is the rms fluctuation on that scale. For a particular choice of δ this gives the fraction of the universe which has 'turned around' in haloes of mass m by a redshift z_t, where for spherical 'top-hat' perturbations $1 + z_t \approx \delta/1.06$. Higher mass fluctuations are assumed to contain those of lower mass and so we must differentiate with respect to mass to find the number of haloes in the range $m \mapsto m + dm$:

$$f = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\delta}{\sqrt{2}\sigma}\right), \tag{20}$$

$$dN_{\mathrm{PS}} = n_{\mathrm{PS}}(m, \nu)\, dm = -(2\times)\frac{\bar{\rho}}{m}\frac{\partial f}{\partial m}\, dm. \tag{21}$$

With a strict application of this method, the normalization should be chosen so that exactly half the universe is in overdense regions. However, we expect that fluctuations on widely different mass scales will be independent and so all matter is taken to be in fluctuations on some scale. More sophisticated treatments (e.g. Cole & Kaiser 1988) give the same result without resort to this fudge.

Peaks theory, on the other hand, counts the number density of peaks, $\varpi(m, \nu)$ which lie above the threshold. Each peak has a mass m associated with it and these masses are assumed not to overlap. In this case the number density of haloes is given by (Bardeen *et al.* 1986)

$$\varpi = \left(\frac{3+n_k}{6}\right)^{3/2}\frac{\bar{\rho}}{\sqrt{2\pi}m}\,\nu^2 \mathrm{e}^{-(\nu^2/2)}, \tag{22}$$

$$dN_{\mathrm{Peaks}} = n_{\mathrm{Peaks}}(m, \nu)\, dm = -\frac{\partial \varpi}{\partial m}\, dm, \tag{23}$$

where n_k is the slope of the power spectrum of the density fluctuations. For comparatively common (low ν) objects, Peaks theory will overestimate the number density since the haloes associated with each peak will tend to overlap. On the other hand, for rare (high ν) objects PS will underestimate the number density since, for example, an overdensity of mass m which peaks above the threshold for just one position of the smoothing centre will give zero contribution to $f(m)$. A further problem with these theoretical models is that they assume a spherical collapse. With substructure, the collapse may take much longer than predicted by the top-hat model, by up to a factor of 1.5 for reasonable mass distributions.

We are interested in the number density and collapse rates of Abell clusters which arise from high-ν fluctuations. In this regime, we have

$$n_{\mathrm{Peaks}} \sim \frac{1}{2}\left(\frac{3+n_k}{6}\right)^{3/2}\nu^3 n_{\mathrm{PS}}, \tag{24}$$

where the constant factor depends upon the power spectrum and the choice of smoothing filter. This difference is explained by the fact that the PS counts the fraction of space where a top-hat centre will give $\delta > \delta_{\mathrm{cut}}$, whereas Peaks just counts a mass m for every peak which exceeds the threshold. For isolated, high-ν peaks the curvature of the density field at the peak, r_m, is of order the smoothing length on that mass-scale. It is then easy to show that the volume occupied by a peak of overdensity ν is proportional to r_m^3/ν^3, in agreement with the above result.

The simplest way to characterize the growth rate of fluctuations in our simulations is to fit the change of radius with time to a simple top-hat collapse model:

$$\theta - \frac{1}{2}\sin 2\theta = \frac{\pi}{2}\left(\frac{t}{t_t}\right), \tag{25}$$

$$r = \begin{cases} r_t \sin^2\theta, & \theta < \dfrac{3\pi}{4} \\ \dfrac{1}{2} r_t, & \theta \geq \dfrac{3\pi}{4}, \end{cases} \tag{26}$$

where θ is a parameter and the subscript t refers to the turnaround time, or time of maximum expansion.

Fig. 8 shows the time evolution of various characteristic radii for three different sets of cluster particles. These are identified, as in Section 3.2, by finding all particles which lie above an overdensity threshold of 180 at three different times. Once labelled, the distribution of these particles can be followed in time. As well as the semi-axes of the particle distribution (defined in Section 3.3), we also plot the mean of these and the top-hat radius, centred on the mean particle position, which contains the same mass. The rates of growth of fluctuations up to turnaround vary slightly along the principal axis directions but are reasonably well represented by the harmonic mean. Note that on the largest mass-scale the top-hat radius is initially smaller than the semiminor axis; this shows that the simple ellipsoid model is not a good representation of the particle distribution and that there is a lot of substructure.

In Table 3 we give the initial overdensity, δ_{turn}, of the top-hat model which best fits the data before turnaround of the halo, and compare this with that given by PS and Peaks theory. It can be seen that in each case δ_{turn} is slightly less than δ_{Peak} but substantially larger than δ_{PS}. Hence Peaks theory will slightly overestimate the number of haloes which have turned around on a given mass-scale, whereas Press–Schechter will greatly underestimate them.

After turnaround the collapse of haloes is much slowed. We attribute this principally to the non-spherical nature of the collapse, although it has been suggested (Cavaliere *et al.* 1986) that subclustering is the primary cause. For a planar perturbation, the growth rate of fluctuations is $r = Rt^{2/3}[1 - \frac{2}{3}(t/t_t)^{1/3}]$, where R is the unperturbed comoving size. This model underestimates the growth rate at early times, giving $\delta \propto t^{1/3}$, whereas $\delta \propto t^{2/3}$ for a spherical perturbation. It gives a better fit, with its longer collapse time, after turnaround. Peebles (1980, section 20) and others have discussed the collapse of an *isolated* ellipsoid which follows the same general form as in Fig. 8. Note, however, that the presence of surrounding material will significantly slow the collapse along the minor axis. A useful measure, from the point of view of measuring collapse rates, is the equivalent initial overdensity in a simple top-hat model which would give the same elapsed time before virialization, at $r \approx \frac{1}{2} r_t$. This quantity, δ_{vir}, is also shown in Table 3. This time the values are closer to the PS estimate than to Peaks theory.

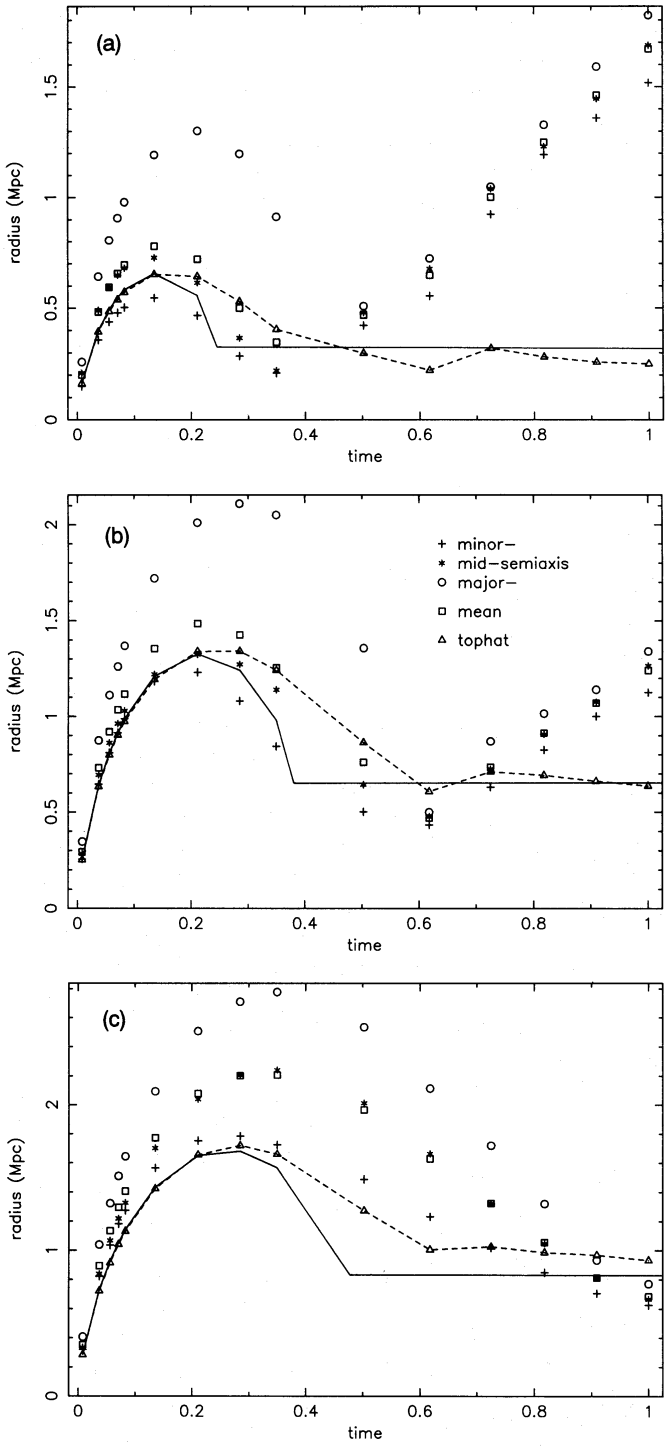

Figure 8. The variations of semi-axis length with time for three different particle sets which correspond to the virialized regions of the cluster in simulation S1 at times 0.35, 0.62 and 1.00, containing 415, 1455 and 2037 particles, respectively.

Peebles (1990) has suggested that the presence of substructure gives rise to 'pre-virialization' which slows the collapse of protoclusters and lowers the virial density significantly. This can be seen by a simple application of the virial theorem to cluster formation. Energy conservation gives

$$(PE)_t+(KE)_t=(PE)_v+(KE)_v, \tag{27}$$

Table 3. Measured overdensities and those arising from fitting top-hat models to the collapse rates of peaks for four different masses of the central cluster in simulation S1. The columns give (i) number of gas particles, (ii) the radius (today) containing this mass, (iii) the initial CDM rms fluctuations on each scale, (iv) the initial value of δ which best fits the collapse up to maximum expansion, (v) the initial value of δ which best fits the collapse up to virialization, (vi) the measured peak value of δ after smoothing on the halo mass-scale, and (v) the value of δ which bounds the mass of the halo after smoothing on the halo mass-scale.

No. ptcles	R/Mpc	σ	$\delta_{\rm turn}$	$\delta_{\rm vir}$	$\delta_{\rm Peak}$	$\delta_{\rm PS}$
415	7.2	0.042	0.16	0.12	0.186	0.124
1062	9.9	0.034	0.13	0.10	0.144	0.087
1455	11.0	0.031	0.12	0.09	0.125	0.077
2037	12.3	0.028	0.10	0.08	0.108	0.069

where (PE) is the potential energy, (KE) is the kinetic energy and the subscripts t and v refer to turnaround and virialization, respectively. The virial theorem applied to the final, supposedly steady, cluster gives

$$(PE)_v+2(KE)_v=0. \tag{28}$$

The standard argument then states that the kinetic energy is zero at turnaround so that, taking $(PE)\propto 1/r$, we have $r_v=\frac{1}{2}r_t$, i.e. collapse by a factor of 2. This collapse can be increased slightly if the final cluster is centrally condensed.

On the other hand, if there is significant internal motion at turnaround, $(KE)_t=-\kappa(PE)_t$, then the final radius is larger, $r_v=\frac{1}{2}r_t/(1-\kappa)$. In his simulations Peebles found typically $r_v\approx\frac{2}{3}r_t$, corresponding to $\kappa\approx\frac{1}{3}$. The mean-square velocity perturbation within a top hat of comoving radius R can be evaluated in the same way as the mean-square density fluctuations (see, e.g. Peebles 1980). It is simply

$$\langle v^2\rangle=\frac{1}{(2\pi)^3V_u}\int \dot{a}^2a^2\frac{P(k)}{k^2}[1-W(kR)]\,d^3k, \tag{29}$$

where V_u is a normalizing volume, a is the expansion factor, P is the power spectrum and W is the top-hat window function. We have assumed here that the peculiar velocity grows as $\dot{a}a$; this will only be true until the perturbation goes non-linear, when $\dot{a}a=H_0a_0a$. Let us suppose that the mean-square velocity saturates at λ times this value. Then, normalizing to unit rms fluctuations on a scale R_0, we have

$$\langle v^2\rangle=\lambda(H_0R_0)^2\frac{\int a_tP(k)\,W(kR)\,dk}{\int a_0P(k)(kR_0)^2\,W(kR_0)\,dk}, \tag{30}$$

which gives, for a power-law power spectrum $P\propto k^{n_k}$,

$$\langle v^2\rangle\approx\lambda(H_0R_0)^2\frac{2(3+n_k)}{1+n_k}\left(\frac{R}{R_0}\right)^{(1-n_k)/2}. \tag{31}$$

This has the same scaling with mass as the virial temperature which is proportional to the potential

$$\frac{3GM}{5r_v}\approx(H_0R_0)^2\left(\frac{R}{R_0}\right)^{(1-n_k)/2}. \tag{32}$$

The above discussion suggests that, for $\lambda\approx1$, internal motions can be important and lead to pre-virialization. However, in our simulations this does not occur. The col-

lapse, although slowed after turnaround, does continue to approximately half the turnaround radius. The reason is simply that the measured values of κ in our simulations, i.e. the ratio of kinetic to potential energies at turnaround, is small. This is true both for the dark matter and gas particles and including the kinetic energy dissipated as heat in the gas. Typically we find $\kappa \approx 0.1$ which gives a negligible effect. Larger values of κ might be found for power-law fluctuation spectra which have larger values of n_k and hence more substructure on small scales than cold dark matter.

We have dwelt on the topic of collapse rates at some length, as they determine the number of high-mass clusters which dominate both optical and especially X-ray catalogues. The overdensity of a top-hat model which matches the collapse is slightly lower than that given by Peaks theory before turnaround, but lies closer to the Press–Schechter value at the time of virialization. Unfortunately, we are again limited by our single realization; it is probable that there is a wide range of collapse times after virialization depending upon the symmetry of the system. However, a naive application of the Press–Schechter model is likely to underestimate the number of rich clusters. The internal kinetic energy at turnaround of the cluster in our simulation is small and does not lead to pre-virialization.

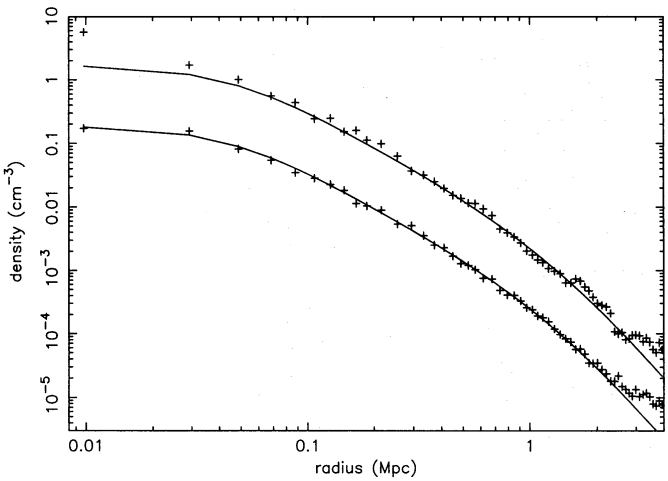

Figure 9. The mean density profiles, in spherical bins centred on the respective density maxima, for the gas and dark matter in the central cluster in simulation S1 at the final time. The density unit for the gas is the number of particles per cubic centimetre assuming cosmic abundances. For the dark matter the number of particles is not well defined and the density is normalized so that the gas and dark matter scales represent equal mass densities. The line drawn through the gas data is model 2 in Table 4. An identical model, apart from a factor of 9 in normalization, fits the outer part of the dark matter distribution very well.

3.5 Density and temperature profiles

The haloes produced in our simulations are not steady, spherically symmetric structures; they are continually growing by accretion. Nevertheless, except during mergers with other large haloes of similar mass, the quasi-static approximation should be a reasonable one.

We identify haloes as in Section 3.2 and calculate the mean and median particle positions and the position of the maximum density: the difference between these can be as large as 200 kpc (showing that the halo is not perfectly symmetric) which is greater than any core structure we might hope to see. It is important therefore to use the density maximum to define the halo centre. This makes such a large difference to the results that we choose to measure the density of the dark matter in the same way as the gas, i.e. using SPH, in order to find the centre of the dark matter distribution separately. We bin all particles (not just those which exceed the density cut defining the halo) about the chosen centre in spherical shells of thickness equal to half the smoothing length (we have also tried using ellipsoidal bins aligned with the moment of inertia ellipsoid but it makes little difference). The density profiles of the gas and dark matter in the central cluster at the final time are shown in Fig. 9. Note that the density maxima of the two distributions are offset by 50 kpc; the core structure visible in the dark matter would not be seen if it were binned about the centre of the gas distribution. It is the volume-weighted density, as measured by the number of particles in a spherical shell, which is shown here; the mass-weighted density profile, obtained by averaging the density on particles within each shell, is flatter, especially in the dark matter, because of subclustering.

For comparison with other results and with theory we fit the central gas density profile to an 'isothermal β-model'

Table 4. Fits to the density profiles of both the gas and dark matter of the cluster and the four largest groups in simulation S1 at the final time. Model 1 is given by equation (33) where 'limit' gives the outer radius beyond which the fit become a poor one. Model 2 is given by equation (34); where this is a poor fit we list instead the asymptotic slope of the density profile beyond the limit of validity of model 1.

		Model 1				Model 2			Model 3
Object		$n(0)/\mathrm{cm}^{-3}$	a/kpc	s	limit/kpc	$n(0)/\mathrm{cm}^{-3}$	a/kpc	b/Mpc	s
Cluster	gas	0.19	50	2.1	1000	0.19	50	1.25	
	dark	9.	11	1.5	250	3.*	35	1.25	
Group 1	gas	0.04	80	2.0	600				3.0
	dark	2.5	30	2.0	700				4.6
Group 2	gas	0.04	70	2.0	1000	0.04	70	1.7	
	dark	1.1	40	1.9	1000				5.1
Group 3	gas	0.10	40	1.9	1000				3.0
	dark	1.8	25	2.0	1000	1.7	30	1.3	
Group 4	gas	0.014	160	2.6	1000	0.014	140	1.0	
	dark	2.1	30	2.0	1000				4.2

*Good fit except for central bin.

(Cavaliere & Fusco-Femiano 1976),

$$n(r) = n(0)(1+x^2)^{-s/2}, \tag{33}$$

where $x = r/a$, the distance from the centre of the cluster divided by the core radius, and the asymptotic slope $s = 3\beta_{\rm fit}$. $\beta_{\rm fit} = 2/3$ corresponds to the outer parts of an isothermal sphere, $\beta_{\rm fit} = 1$ gives a Hubble profile and larger $\beta_{\rm fit}$ give well-defined haloes of finite mass. This fit is often a poor one in the outer regions of a halo as the decline in density is not a simple power law. In many cases the profile gradually steepens in the outer regions, while in others there seems to be a break to a second power law. The cluster profile is well fitted over a large range of radii by the function

$$n(r) = n(0)\frac{[1/(a^2+r^2)]-[1/(b^2+r^2)]}{(1/a^2)-(1/b^2)}, \qquad a<b. \tag{34}$$

This resembles equation (33) above for a, $r \ll b$, but with a fixed slope, $s = 2$, corresponding to the outer regions of an isothermal sphere. At large radii, $r \gg b$, the slope increases to $s = 4$ although the correct asymptotic value is not well defined by the data and any fitting function with an increasing gradient at large radii would probably do equally well. Note that these two models have the same number of free parameters and yet the second gives an accurate fit over a much larger range of radii.

The results for the central cluster and the other four largest groups in the simulation at the final time are given in Table 4. There are several points of interest here. First, the dark matter slopes are all close to $s = 2$; that of the cluster is 1.5 within 250 kpc but is nearer to 2 when averaged out to 1 Mpc. This slope is indicative of the outer regions of an isothermal sphere, as would be expected if relaxation were to go to completion. The smoothing length in our simulations is 40 kpc, so it would seem that none of the group haloes has a significant core in the dark matter distribution. As discussed in Section 3.1 earlier, the emergence of a central peak in the density field of the cluster does seem to be associated with impulsive relaxation during its formation, although the subsequent core evolution may be dominated by numerical artefacts. Similarly the gas, which sits in the potential well of the dark matter, also has a small core radius (except for group 4). Until recently it had been thought that clusters of galaxies possess substantial cores in their galaxy distributions and catalogues typically list core radii of 200–400 kpc (e.g. Sarazin 1986). Their measurement is hampered by the small number of galaxies present in any one cluster and the difficulty of identifying the centre. Recently, however, combining data from many clusters and centring on the central dominant galaxy, an exponential profile, increasing all the way into the centre, has been found (e.g. Merrifield & Kent 1989). This result holds only for clusters which possess a massive central galaxy, but suggests that this aspect of our simulations may be correct.

The gas distribution in real clusters must possess a finite core radius or the X-ray emission would be divergent towards the centre. Observed cluster X-ray core radii are typically 400 h_{50}^{-1} kpc (Jones & Forman 1984) and so there must be some element of missing physics in our simulations. The most obvious is heating by, for example, supernovae or galaxy motions; the pressure is constant within 100 kpc so that making the gas isothermal would lead to a measurable core radius. Even in the absence of heating, however, a small core would soon establish itself: the central gas would cool leading to inflow of more matter of higher entropy. Cooling time is proportional to $T^{1/2}/n \propto (T/n^{2/3})^{3/2}/T$, so that at a given (virial) temperature the cooling time of the gas would increase (or correspondingly its density would be lower). This effect is present in the cooling simulation S3 (see Section 4.2) but is only of minor importance, as in rich clusters such as this one the cooling time is less than a Hubble time only in the very centre, within a couple of hundred kiloparsecs; it is the less massive clusters, born early on in the Universe when the background density was higher, which have the most massive cooling flows. The strong dependence of X-ray emission on cluster morphology bodes ill for attempts to constrain cosmological models using the X-ray cluster luminosity function. It has been shown by Edge *et al.* (1990) and Pearce & Thomas (1991) that the strong evolution found by Edge *et al.* in the number density of high-luminosity clusters may reflect an evolution in core radius as much as a change in the cluster mass function.

In the outer regions of the cluster the density profiles steepen. The exact form of the turnover varies between clusters and with time, however, the dark matter has a consistently steeper slope than the gas; this perhaps reflects the pressure support of the intracluster medium as the gas falls into the outer regions of the cluster.

A similar procedure to the above can be used to derive the temperature profile of the haloes; that of the central cluster is shown in Fig. 6(b). All the haloes show a similar effect: the temperature is low in the centre, rises steeply out to about 100 kpc, then remains constant out to 500 kpc or so and finally falls rapidly in the outer regions of the halo where virialization is not complete. The low central temperatures arise because gas initially near the middle of the protohalo does not have to fall through a large potential difference during halo formation. In practice we expect the heating mechanisms described above to create an isothermal core and this would lead to a larger core radius for the gas. Some, but not all, of the decline of temperature with radius between 500 kpc and the edge of the cluster may be due to incomplete thermalization of kinetic energy for the gas in our code, as discussed in the next section.

3.6 The β problem

Much play has been made in the literature of the fact that the ratios of halo and gas temperatures (energies) deduced from fits to isothermal models, $\beta_{\rm fit}$ above, do not agree with direct measures of the gas temperature and galaxy velocity dispersion, $\beta_{\rm spec} = \sigma_{\rm gal}^2/(kT_{\rm gas}/\mu m_{\rm H})$ (e.g. Mushotzky 1984, 1988; Evrard 1990), whereas it is easy to show (see equation 38, below) that in a steady cluster in which an isothermal gas and galaxies sample the same gravitational potential, and in which the galaxies follow a Hubble density profile with a constant velocity dispersion, $\beta_{\rm spec}$ should equal $\beta_{\rm fit}$,

$$\beta_{\rm spec} = \beta_{\rm fit}. \tag{35}$$

It is not our intention to bore the reader with yet another overview of the topic, but nevertheless our interpretation of this discrepancy, guided by the simulations, differs from most previous work. We suggest that the non-specialist skip to the summary at the end of this section.

The current status of the observations has been summarized by Edge & Stewart (1991). The best fits to imaging data still come from *Einstein* data and give $\beta_{\rm fit} \approx 0.5$–$0.8$ (Jones & Forman 1984), although *ROSAT* should provide more accurate results. $\beta_{\rm spec}$ has come down in value over the years primarily as a result of lowered determinations of the velocity dispersion of clusters: both observations (summarized by West & Bothun 1990) and numerical simulations (Frenk *et al.* 1990) have shown that substructure and projection effects can lead to substantial overestimates. The best current value is $\beta_{\rm spec} \approx 0.8$–$0.9$ which is still significantly larger than $\beta_{\rm fit}$. Failing any further significant revision of these numbers, it is clear that one of the basic assumptions in the simple model must be wrong.

There are no galaxies in our simulations and so we cannot test the isothermal β-model directly. Instead we compare the gas against the dark matter density distribution and show below that there is good agreement with a hydrostatic model. Any discrepancy which arises when applying the model to the observations must therefore be telling us something interesting about the relative distribution of *galaxies* and dark matter.

From Table 4 we see that $\beta_{\rm fit} \approx 0.63$–$0.7$ in the inner regions of four of our haloes; the fifth, with $\beta_{\rm fit} \approx 0.87$, has a larger core radius. These values are consistent with the observations (Jones & Forman 1984). In the outer regions the density profile steepens and $\beta_{\rm fit}$ increases, but we are only interested in the centre where most of the flux originates and where the cluster is in virial equilibrium.

The virial temperature of the haloes can be estimated from the velocity dispersion of the dark matter particles. The specific energy of the gas, $\varepsilon_{\rm gas}$, is also easily evaluated but differs from the total energy due to incomplete virialization. We find $\frac{1}{2}\langle v_{\rm gas}^2\rangle/\varepsilon_{\rm gas} \approx 1/8$ in the central 600 kpc, increasing to unity at 3 Mpc and with negligible thermal energy in the intergalactic medium. This central value is similar to, though slightly lower than, that of Evrard and suggests that there is residual kinetic energy in the cluster gas: this may, however, be a numerical artefact. Thermalization of kinetic motions is achieved via artificial viscosity which works well in one-dimensional shock tube experiments, although even in this circumstance the shock is about 6 particles wide. In our simulations the shock is non-steady, non-planar (so that particles may be able to slip around the side) and we have only a few $\times 6^3$ particles in the cluster core. Accordingly, we have calculated the fraction of kinetic energy which arises from differential motion between each particle and its neighbours (within the range of the SPH smoothing kernel): within 2 Mpc of the cluster centre approximately two thirds of the kinetic energy is local to each particle, in the intercluster region this fraction decreases dramatically. While incomplete thermalization may be present in clusters, we cannot rule out the possibility that much of the residual gas motion seen in cluster simulations is artificial.

Assuming that most of the kinetic energy of the gas should really be thermalized, we then measure the ratio of gas and dark matter energies to be

$$\beta = \frac{\langle v_{\rm dark}^2\rangle}{\langle v_{\rm gas}^2\rangle + 2\varepsilon_{\rm gas}}. \qquad (36)$$

The variation of β with radius in the cluster is plotted in Fig. 10. It averages to about unity within 1 Mpc, then drops to 0.5 at the edge of the virialized region before rising again on the outskirts of the cluster. Averaged over the whole cluster, then, we find $\beta \approx 0.9$–1.0. This is slightly lower than the value of 1.06 found by Evrard (after allowance for his slightly different definition of β): we do not know the reason for this difference.

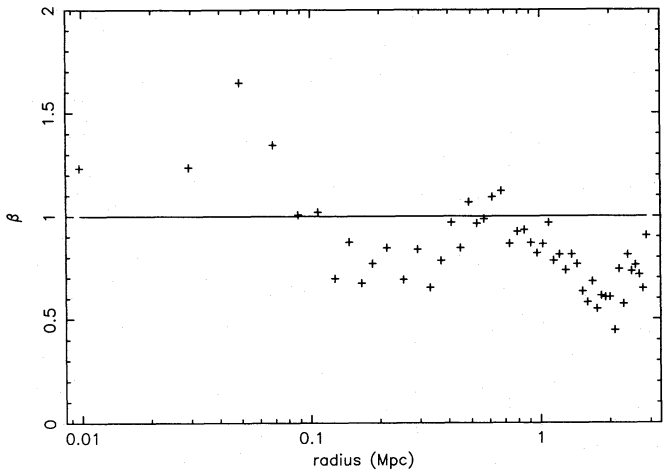

Figure 10. The ratio of dark matter to gas energy, as defined by equation (36), in spherical bins about the cluster centre in simulation S1 at the final time.

The main reason why β as defined by equation (36) exceeds $\beta_{\rm fit}$ is that the dark matter is not distributed with a density proportional to r^{-3}, as is usually assumed for the galaxy distribution, but as r^{-2}. This gives

$$\beta = \tfrac{2}{3}\beta_{\rm fit} \qquad (37)$$

in the simple model, or $\beta \approx 1$ as obtained in the inner parts of our cluster.

A more sophisticated relationship between $\beta_{\rm fit}$ and β can be calculated from the spherically symmetric form of the Jeans equation for hydrostatic support (Binney & Tremaine 1987, p. 198):

$$\beta = \frac{\langle v_{\rm dark}^2\rangle}{2\varepsilon_{\rm gas} + \langle v_{\rm gas}^2\rangle}$$

$$= \frac{f_T(3\beta_{\rm fit} + \theta) + (1 - f_T)[3/(3-2\alpha)](3\beta_{\rm fit} - \phi - 2\alpha)}{[3/(3-2A)](3B_{\rm fit} - \Phi - 2A)}, \qquad (38)$$

where $f_T = 2\varepsilon_{\rm gas}/(2\varepsilon_{\rm gas} + \langle v_{\rm gas}^2\rangle)$ and $\theta = -d\ln T/d\ln r$. $\alpha = 1 - \sigma_t^2/\sigma_r^2$ is a measure of the anisotropy of the velocity dispersion in the gas and $\phi = d\ln\sigma_r^2/d\ln r$ the gradient of radial velocity dispersion; A, B and Φ are the corresponding quantities for the dark matter.

Note that this equation is a little misleading because it is not $\beta_{\rm fit}$ which determines β but rather the other way around. Small changes in the parameters on the right-hand side of the equation could lead to large variation in β but this merely indicates that $\beta_{\rm fit}$ is well constrained. The narrow range of slopes $s \approx 2$ for the density profile of the gas is a consequence of the mean gas and dark matter energies being approximately equal. That being said, the general trends apparent in Fig. 10 can be explained as follows: in the centre the gas is cool, as it has not fallen through a large potential gradient but was initially positioned at the centre of the cluster potential; between radii of 100 and 300 kpc the dark matter has a

slightly steeper density gradient but beyond this the particle orbits become highly radial and provide much less support; beyond 1 Mpc the isotropic gas pressure diminishes and the increasing radial gradient of the gas velocity dispersion provides less support than the decreasing dark matter velocity dispersion; finally, in the outskirts of the cluster, the properties of the gas and the dark matter become equal and β rises again.

The dark matter shows a central core in velocity dispersion within about 700 kpc, presumably due to efficient mixing of material, whereas the gas thermal energy keeps increasing in to a radius of 200 kpc. Beyond 2 Mpc the dark matter density and energy decline slightly faster out to the background values. These differences are sufficient to explain the variation which is seen in Fig. 10.

It would be of great interest to know whether in real clusters the dark matter and gas are at the same temperature (i.e. $\beta \approx 1$). This is the case in our simulations when averaged over the whole cluster, and indeed run S2 in which the cluster is more isolated and can be run on to later times shows $\beta \approx 1$ throughout the cluster except in the very centre. Our simulations neglect the effects of heating associated with metal enrichment of the intracluster medium or galaxy motions, and radiative cooling, all of which will leave their imprint on the gas.

Note that, although there is no sign of the β-discrepancy in our results, this does not mean that it has gone away or been explained; it is merely that there are no galaxies in our simulations. If we put $\beta_{\rm fit} = 1$ in equation (38) as is reasonable for the galaxy distribution, and $f_T = 1$, $\theta = \Phi = 0$, then we recover $\beta_{\rm spec} = \beta_{\rm fit}$ independent of the anisotropy of galaxy orbits. In other words the relative density profiles of gas and galaxies should reflect the relative energies available for support, contrary to the observations. There are many possible causes for this discrepancy: the gas temperature may decline with radius as found in our simulations, the observed velocity dispersions may decrease still further, the Hubble law may be a poor fit to the density profile, or the radial component of the velocity dispersions may increase rapidly with radius – in both of these latter two cases anisotropic orbits would help.

The steeper decline of galactic number density with radius than either the gas or dark matter suggests that galaxies have smaller kinetic energies than the other two components. This hypothesis has gained support from recent simulations (Carlberg, Couchman & Thomas 1990; Carlberg & Dubinski 1991) where the effect, brought about by dynamical friction, is called velocity bias. Unfortunately, this once again exacerbates the β-discrepancy and it remains to be seen whether velocity bias is present in real clusters. This is discussed in the next section.

3.7 The dynamics of galaxy haloes

In this section we discuss the dynamics of galaxy haloes, identified with substructure in the cluster at the final time. In run S1 there are many high-density regions in the dark matter but few in the gas distribution. When cooling is added, however, the dark matter haloes become sites for accretion of cooled gas and hence, presumably, galaxy formation – see the discussion of simulation S3 in Section 4.2, below. Unfortunately, there are many factors which may complicate a simple relation between dark matter haloes and galaxies: the different disruption rates during clustering is the most obvious, plus the relative effects of dynamical friction and ram pressure from the background. It has even been suggested by Carlberg & Couchman (1989) that there may not be a simple one-to-one association of galaxies with haloes. Nevertheless, in order to define a galaxy sample in our simulations with which to test the effectiveness of galaxies as tracers of the mass, we assume here that large haloes are sites of massive galaxies and that there is a monotonic relation between halo size and galaxy mass – all weighting of mean quantities will be the same for each halo and not by galaxy mass.

Substructure is identified as in Section 3.2. In the outskirts of the cluster between radii of 1 and 3 Mpc we use a minimum density cut of 180 times the mean density, which should identify all virialized haloes. Within 1 Mpc the mean cluster density is higher and so we raise the density cut to 1800 – if applied throughout this would not lead to splitting of any of the outer haloes into subcomponents but would merely result in a lowered detection rate. In both cases the minimum number of particles is fixed at five to ensure a reasonable estimate for the mean velocity and internal velocity dispersion; raising this minimum number makes very little difference to the results.

32 haloes are detected altogether in addition to the central dominant (cd) galaxy, 12 in the outer region beyond 1 Mpc and 20 within this radius. The mean internal velocity dispersion (three-dimensional) of the haloes is approximately 1200 km s^{-1} which is equivalent to 3×10^7 K, so the haloes correspond to subgroups rather than individual galaxies. The mean external velocity dispersion of the haloes (i.e. treating each halo as a single galaxy) is 1100 km s^{-1}, with a velocity relative to the cluster mean of 360 km s^{-1} which is just within the acceptable error if the haloes are drawn from a random population centred on the cluster. The cd lies within 100 kpc of the cluster centroid and has a peculiar velocity of 130 km s^{-1}.

The above external velocity dispersion of the halo population is much less than the rms velocity of 1600 km s^{-1} for all the dark matter particles within the cluster, but becomes equal to it if we add, in quadrature, the internal velocity dispersion of the haloes. This remains true for our two subsamples: the velocity dispersion between haloes is almost unchanged although the outer sample has negligible internal velocities, whereas in the inner haloes these dominate. In both cases, however, the total internal plus external velocity dispersion is equal to the cluster rms velocity in the same region: 2000 km s^{-1} in the centre and 1200 km s^{-1} further out, reflecting a sharp decline with radius. There is thus no evidence for velocity bias in the halo population which we have identified. Our sample is not ideal for the study of this effect, however, since it is defined by massive agglomerations of dark matter particles at the present time; ideally we should select baryonic galaxies at high redshift and follow their evolution through to the present, as in the work of Carlberg & Couchman (1989).

The mean velocity anisotropy parameter (see equation 38) is $A \approx 0.63$, but varies strongly as a function of radius, being highest in the outer regions where infalling matter is on highly radial orbits. We have looked for a trend with mass, expecting higher mass galaxies to be on more circular orbits, and find the opposite to be true; this may be because low-mass haloes on highly radial orbits are preferentially dis-

rupted. If anything, high-mass haloes have slightly larger velocities than low-mass ones and so there is no evidence in velocity space for either dynamical friction or mass segregation.

The dynamics of the galaxy haloes can be used to estimate the cluster mass. There are many different ways of doing this; here we just concentrate on two. The virial theorem, when applied to test particles in orbit about a point mass, leads to the estimator $V1=(1/G)\sum v^2/\sum(1/r)$, which when projected on to the sky becomes $V2=(3\pi/2G)\sum v_p^2/\sum(1/R)$. Here v_p is the line-of-sight component of the velocity, v, relative to the cluster mean, and R is the projected component of the distance, r, from the halo to the cluster centre. The reciprocal of the radius in these expressions leads to a large (infinite!) variance and means that the estimator is dominated by objects close to the cluster centre. A more stable statistic is the combination v^2r (Bahcall & Tremaine 1981) which gives

$$B1=\frac{2(2-A)}{(3-2A)\,G}\frac{1}{N}\sum v^2r$$

and

$$B2=\frac{32(2-A)}{\pi(4-3A)\,G}\frac{1}{N}\sum v_p^2R$$

for the two estimators, where N is the total number of objects in the sample. There are various alternatives to these statistics which it is not worth examining in detail here: for example, one can weight the halo contributions by mass, or use interhalo separation in place of the radius within the cluster. The values of these statistics when applied to the cluster as a whole and to the two radial divisions are shown in Table 5.

The first thing to note is that the scatter in the projected mass estimators is large and can be a factor of 3 or more even with as many as 20 galaxies. The virial estimators have a larger scatter than the Bahcall & Tremaine ones, and we would expect this difference to grow as the number of galaxies is increased. We will use the mean value given by the three coordinate directions, below.

Table 5. Estimates of the mass of the cluster in simulation S1 at the final time from the four statistics described in Section 3.7. The results are expressed in terms of the number of particles of each kind (assuming their spatial distribution is the same). There are three entries for each of the projected statistics, one for each coordinate direction. Also, there are two columns for each of the B-estimators; the first assumes an anisotropy parameter $A=0$ whereas the second uses the measured value from the simulation.

Statistic	Whole cluster		Outer region		Inner region	
True	4459		4459		2454	
V1	962		2187		782	
V2	745		2880		468	
	712		1779		617	
	1814		2998		1588	
B1	1931	2400	3010	3968	1283	1593
B2	1722	2437	3473	5444	672	811
	1422	2012	2112	3311	1009	1218
	2247	3449	2674	4192	1991	2403

All the estimators greatly underestimate the mass when applied to the cluster as a whole. $B1$ and $B2$ are about a factor of 2 too small. The reason for this is simply that the assumption of test particles in orbit about a point mass is a poor one and only measures the mass within a typical galaxy radius. If the haloes are drawn uniformly from the density distribution within the cluster, then a neat little calculation by Heisler, Tremaine & Bahcall (1985) shows that, regardless of the mass distribution within the cluster, the B statistics underestimate the mass by exactly a factor of 2 (they actually show this only for radial and isotropic orbits in $B2$ but it holds for both $B1$ and $B2$ whenever A is constant throughout the cluster). The corrected B statistics therefore lie close to the true cluster mass.

The virial estimators are more than a factor of 4 too low. This again is due to the high weighting given to small separations, which effectively leads to a measurement of the mass only within the central regions of the cluster. Also it is not clear how to correct the virial statistics for velocity anisotropy, which appears to be a 20 per cent effect here. In view of this and the large variance, we do not recommend the use of the virial theorem in this way to measure cluster masses.

In the outer regions of the cluster the B statistics (once corrected for velocity anisotropy) provide a reasonable fit to the mass without the need for an extra multiple of 2, whereas in the centre this factor is required. This difference arises because in the former case most of the mass does indeed lie interior to the galaxy distribution, and shows that the statistic must be tailored to the particular system under study and not just applied blindly. It is interesting that the maximum possible correction for radial orbits ($A=1$), as opposed to isotropic ones ($A=0$), is also a factor of 2 for statistic $B2$ (1.5 for $B1$), and that this compensates for the uncertainty in the mass distribution.

In conclusion, the virial estimators have large scatter and tend to grossly underestimate the mass. The projected mass statistic, $B2$, is probably best for most applications (for which three-dimensional information is not available), although even here the scatter is large with 32 galaxy haloes in our sample, and incorrect modelling of the velocity anisotropy or mass distribution can each introduce a factor of 2 into the result. In the absence of any specific model we agree with Heisler *et al.* (1985) that taking $B2$ with $A=1$, i.e. $(32/\pi G)(1/N)\sum v_p^2R$, is probably the best option for estimating the masses of clusters of galaxies from galaxy motions.

4 DISCUSSION

4.1 Comparison with flat box

Runs S1 and S2 differ only in the manner in which the box is allowed to expand. In S1 the expansion factor matches a cycloid with initial overdensity 0.0312 and final overdensity 1.87, whereas in S2 the expansion factor is taken proportional to $t^{2/3}$ as it would be in a flat universe in which the overdensity is identically zero ($\Omega_{box}=1$). Thus S2 represents a more isolated cluster which forms from fluctuations on scales of less than 50 Mpc only. The variance of the cluster on a scale of 2000 particles (i.e. a comoving top-hat radius of 12.2 Mpc), which is 3.1σ in the box, is equivalent to 2.7σ overall, as compared to S1 where the extra power on large scales lowers the fluctuation probability to 3.7σ.

Because the same waves are present in both S1 and S2, we expect to see the same overall structures in the latter as in the former. However, the time-scale for the growth of large-scale structures will be slowed by the decreased power. This effect is clearly visible in Fig. 11 which shows the appearance of simulation S2 after 500 time-steps ($t=1.38$). The positions of the largest structures closely resemble those in Fig. 3 at time 0.723 for run S1, whereas the condensed nature of the groups and clusters is similar to that at $t=1.00$. This is confirmed by an analysis of the substructure which shows little difference between the number and size of haloes surrounding the central cluster in the two simulations at the same time.

The central cluster at $t=1.38$ contains 1346 particles with overdensity greater than 180. This is rather less than we would expect: taking the ratio of the initial overdensity on this scale in the two boxes, the collapse time-scale should vary as

$$\frac{t_2}{t_1}=\left(\frac{\delta_{1i}}{\delta_{2i}}\right)^{3/2}=\left(\frac{\delta_{2i}+\delta_{bi}}{\delta_{2i}}\right)^{3/2}\approx 1.5, \tag{39}$$

where the subscripts 1, 2 and b refer to simulations S1, S2, and the box, respectively. At time $1.38/1.5\approx 0.92$ in S1, the size of the virialized region was 1800 particles. This suggests that the effective size of the cluster is greater than just the virialized region – there are 3000 particles within the position of the entropy maximum and the effect of the missing large-scale power on this scale is much more pronounced. Once again we see that the naive application of a simple top-hat model to cluster formation can lead to large errors in an estimate of the collapse time.

At equal times the cluster mass in run S2 is reduced relative to that in S1, but the total extent of the shocked region is much the same. Comparing the final state of S1 with the closest analogue we have for S2 ($t=1.06$), we find that the density normalization in the latter is lowered by a factor of 2.3 and the temperature by a factor of 1.6, leaving the entropy normalization almost unchanged. Thus the entropy profile is almost identical at equal times in each simulation; it is not clear why this should be so.

The ratio of dark matter to gas energy, β, appears to follow that for run S1 in the centre of the cluster but is much more uniform in the outer regions, perhaps reflecting the more isolated environment.

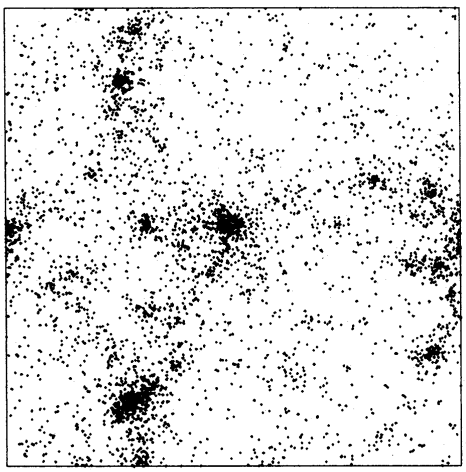

Figure 11. A projection of the distribution of gas particles in simulation S2 at time 1.38.

4.2 The cooling run

In this section we look at the difference that cooling makes to the basic simulation by comparing simulations S1 and S3. The cooling time of gas in protogalactic haloes was first studied by Rees & Ostriker (1977) and Silk (1977), who concluded that it would be less than the dynamical time in galactic systems but longer in higher mass ones. More detailed models, including the presence of dark matter (White & Rees 1978; Thomas 1988) fix the transition mass between these two regimes as approximately $M_G=3\times10^{13}(\Omega_b/\Omega)\ M_\odot$, corresponding to the mass of a giant elliptical galaxy. Thus it is reasonable to suppose that the ability of gas to cool and form stars regulates the formation of galaxies.

In our simulations M_G corresponds to just 10 gas and 10 dark matter particles, so that we are unable to resolve galaxies in any detail. Nevertheless, the general features of the model are reproduced well. The most obvious difference between S1 and S3 is the presence of dense substructure in the gas distribution of the latter. This can be seen by comparing the right-hand panel in Fig. 5 with the left-hand panel of Fig. 12 where we show the central gas particles with over-

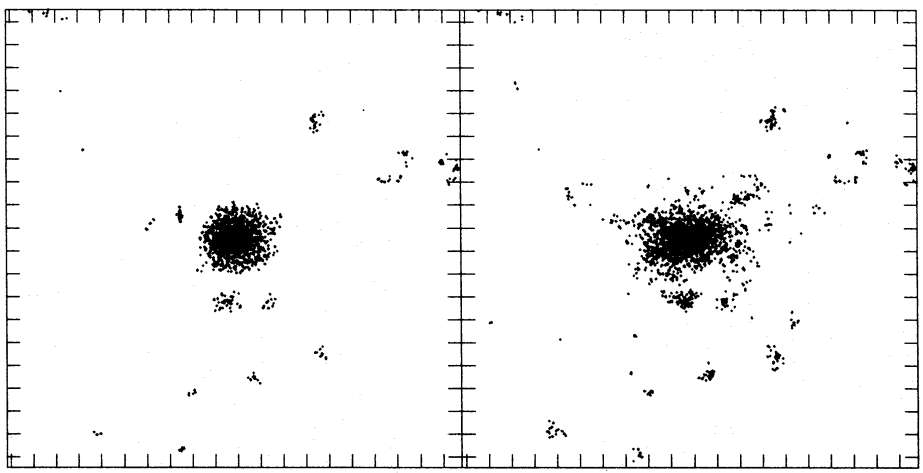

Figure 12. Projections through the centre of simulation S3 at time 0.89 of all particles with measured overdensity exceeding 180. The gas is pictured on the left, the dark matter on the right.

density greater than 180 at the final time for run S3 (this corresponds to $t = 0.89$ rather than $t = 1.00$ as each time-step is shorter and takes longer to evaluate than in S1). Apart from these dense regions, which contain only a small fraction of the total mass, the overall structure of the cluster is unchanged.

The temperature of the gas in the haloes depends upon their size. With up to 10 gas particles (exceeding the density cut) the temperature is close to 10^4 K, the minimum for which cooling is effective (at lower temperatures hydrogen is effectively neutral). As the halo mass increases, cooling rapidly becomes negligible and the temperature rises to the virial temperature, in complete agreement with the theory.

In the real universe we would expect cool gas to form stars and hence galaxies. There would be sufficient dissipation for the galaxy to become self-gravitating and subsequently to survive any evolution in its environment. In our simulations, collapse is limited by the softening which is fixed at 40 kpc and there is no star formation. Nevertheless, many subgroups do appear to survive in the outskirts of the cluster at radii greater than about 1 Mpc. These gas clouds are associated with dark matter haloes which are present even in the absence of cooling: so the gas does not condense far enough in our simulations to confer extra stability to the haloes. In order to model galaxy formation, we would have to increase the spatial resolution and add a star formation algorithm to our code.

Most clusters of galaxies exhibit cooling flows in their cores, the exceptions being those which show evidence of a recent merger (for the latest review of cooling flows see Fabian, Nulsen & Canizares 1991). The cooling time at constant pressure via bremsstrahlung radiation is

$$t_{\rm cool} \approx \frac{5kT/2}{3n\Lambda/2} \approx 1.05 t_0 \left(\frac{T}{10^8\,\mathrm{K}}\right)^{1/2} \left(\frac{n}{10^{-2}\,\mathrm{cm}^{-3}}\right)^{-1}, \qquad (40)$$

which gives a cooling time which is less than the age of the Universe for gas within about 170 kpc of the cluster centre – the cooling flow would have been more important had the cluster been less rich but formed at an earlier stage. Outside the cooling region the temperature and density profil[illegible] simulations S1 and S3 are identical, but inside 70 kpc the temperature and entropy in S3 fall dramatically whereas the density continues to rise, with no sign of a central core. The reason for this can clearly be seen in Fig. 13 which shows a histogram of the gas particle densities within the inner 150 kpc: there is a much wider range of densities present in the cooling run. The sharp peak of 70 or so particles with overdensity greater than 10^5 represents cooled gas which has collected at the centre of the cluster, but there is also a range of high-density gas throughout the central region of the cluster. Cooling has thus produced a multiphase medium and would give rise to distributed deposition of cooled material, as is observed in nearby cooling flows (Thomas, Fabian & Nulsen 1987; Thomas 1988). It is expected that there would be a range of entropies in the gas throughout the cluster, but the resolution in our simulations is not sufficient to show this. That we have obtained the correct qualitative behaviour at the cluster centre, however, augurs well for the future of SPH to model cooling in cluster and galaxy simulations.

The amount of cooled gas which has accumulated in the cluster core agrees well with the simple estimate of the amount of gas within the cooling radius, where the cooling time equals the age of the system, about 100 particles or $3 \times 10^{12}\,M_\odot$. Plotting density and temperature profiles of the high-entropy gas only reveals that the density and temperature of the hot phase resemble those in the non-cooling run, but with a slightly lower central density because higher entropy gas has been drawn into the central regions.

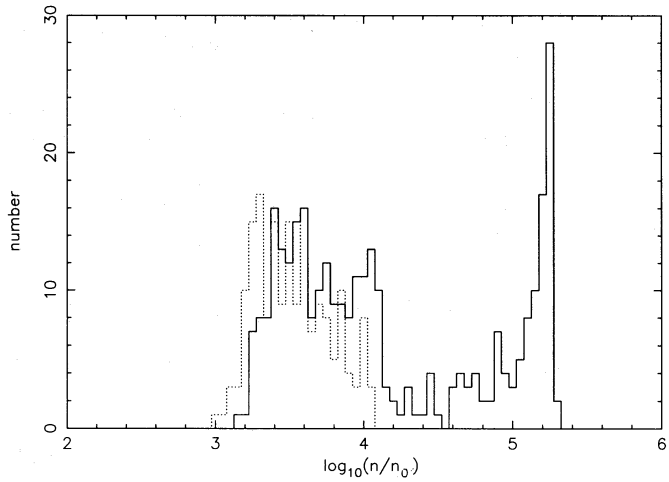

Figure 13. Density histograms of gas particles within the central 150 kpc of the simulation S3 at time 0.89 (solid line) and simulation S1 at time 0.82 (dotted line).

4.3 The X-ray luminosity–mass relation for clusters

Clusters of galaxies are powerful emitters of X-radiation from the intracluster medium (ICM) and can be detected at high redshift. It is to be hoped, therefore, that surveys such as the *Einstein* Medium Sensitivity Survey (Gioia *et al.* 1990), or a smaller but more complete sample by Edge *et al.* (1990) and the future *ROSAT* (almost) all-sky and deeper surveys will be able to determine the cluster X-ray luminosity function to cosmologically interesting distances. By relating this to the predicted mass function it should be possible to constrain different cosmologies. Such a comparison has already been attempted by Cavaliere & Colafrancesco (1984). Unfortunately the translation from X-ray luminosity to mass is strongly dependent upon the state of the ICM in the centre of the cluster, as in the absence of a core the emission would be divergent towards small radii. While cooling will tame this divergence, it is clear that there is a lot of missing physics in our simulations: there is no metal enrichment of the ICM together with the associated heating, there are no galaxies which would heat the gas via dynamical friction or dissipative sound waves (Pringle 1989) and there is inadequate modelling of its multiphase structure. No reliable conversion from mass to X-ray luminosity can be made, then, until the models become more sophisticated. The temperatures of clusters would be much more directly computable from their masses, but unfortunately can only be determined with any accuracy for a small number of bright, nearby clusters (Henry & Arnaud 1991).

To compare observations with cosmological models we need also to determine the predicted mass function of clusters. As was discussed at some length in Section 3.4, the two standard models, Press–Schechter and Peaks theory, give widely differing predictions for the number of high-mass

clusters. In general, Peaks theory performs better and gives a larger number of clusters at turnaround. However, departures from sphericity and substructure can slow the subsequent collapse by a factor of 2, and there is likely to be a wide range of masses for a given initial overdensity. *It is not sufficient to apply the Press–Schechter model blindly* as this will underestimate the number density of massive, X-ray luminous clusters.

4.4 Mass determinations from gas profiles

The equation of hydrostatic equilibrium for gas in a spherically symmetric cluster,

$$\frac{1}{\rho}\frac{dP}{dr} = -\frac{GM(<r)}{r^2}, \tag{41}$$

can be used to determine the cluster mass provided that two independent measures of density and temperature are available across the cluster. One of these is the emissivity, $n^2\Lambda$, which is readily obtained by deprojection of the surface brightness profile (Fabian *et al.* 1981). In the future the temperature may be determined as a function of position in the cluster, but this is not yet possible both because of the poor spatial resolution of the observations and the slow variation of spectral shape with temperature in the observed energy range. In practice, then, some assumption relating density and temperature is made, varying from a $\Gamma = \frac{5}{3}$ polytrope, $P \propto \rho^{\Gamma}$, to isothermal, $\Gamma = 1$. This gives a wide range of possible masses (e.g. for application to the Coma cluster see Cowie, Henriksen & Mushotzky 1987; Hughes *et al.* 1988).

Our simulations suggest an alternative postulate: although the density and temperature profiles vary in slope we find that $P \propto \rho^{5/3} r$ to fairly high accuracy in the shocked gas. This gives masses within, but towards the upper end of the above range (e.g. for $\rho \propto r^{-2}$ we would have $\Gamma = \frac{7}{6}$). Additional heating processes which have been neglected in our simulations will tend to make the gas more isothermal, whereas cooling in rich clusters is likely to be important only in their centres. Therefore the two assumptions of $P \propto \rho^{5/3} r$ and $P \propto \rho$ are likely to bound the true solution.

4.5 Future improvements

There are at least three ways in which the present work can be extended. First, in order to study the dynamics of the collapse and the morphology of the end-state we need to perform a large number of simulations with a variety of initial conditions, both different realizations of the CDM power spectrum and also other spectra with different amounts of power on small and large scales. In this way the statistical variations in subclumping, the collapse rate, mass estimators, etc. can be found. Unfortunately, even though a wide variety of overdensities can be taken for the box, the extra power on large scales can only be added in a symmetric manner – the effect of large-scale tidal fields on cluster morphology cannot be followed.

A second extension would be to do simulations with a larger number of particles. This would enable much higher resolution of the cluster and group cores. To make such simulations viable, given the strong clustering which is present, a variable-resolution code such as TREESPH (Hernquist & Katz 1989) or adaptive P^3M (Couchman 1991) is required. Such simulations are in progress.

Finally, the most interesting aspect of the simulations from the point of view of the intracluster medium is the missing physics. We would like to include galaxy formation, metal enrichment, heating, magnetic fields, and any other important physical processes. Ultimately we hope to model the full multiphase structure of the ICM with the hot phase, cooling dense blobs and cold H I gas.

ACKNOWLEDGMENTS

PAT would like to thank the CITA for support during the initial program development from which the code used in these simulations evolved. HMPC would like to thank the NSERC for financial support.

REFERENCES

Bahcall, J. N. & Tremaine, S., 1981. *Astrophys. J.*, **244,** 805.
Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. S., 1986. *Astrophys. J.*, **304,** 15.
Binney, J. & Silk, J., 1979. *Mon. Not. R. astr. Soc.*, **188,** 273.
Binney, J. & Tremaine, S., 1987. *Galactic Dynamics,* Princeton University Press, Princeton, New Jersey.
Bond, J. R. & Efstathiou, G. E., 1984. *Astrophys. J. Lett.*, **285,** L45.
Carlberg, R. G. & Couchman, H. M. P., 1989. *Astrophys. J.*, **340,** 47.
Carlberg, R. G. & Dubinski, J., 1991. *Astrophys. J.*, **369,** 13.
Carlberg, R. G., Couchman, H. M. P. & Thomas, P. A., 1990. *Astrophys. J. Lett.*, **352,** L29.
Cavaliere, A. & Fusco-Femiano, R., 1976. *Astr. Astrophys.*, **49,** 137.
Cavaliere, A. & Colafrancesco, S., 1984. *Astrophys. J.*, **331,** 660.
Cavaliere, A., Santangelo, P., Tarquini, G. & Vittorio, N., 1986. *Astrophys. J.*, **305,** 651.
Cole, S. & Kaiser, N., 1988. *Mon. Not. R. astr. Soc.*, **233,** 637.
Couchman, H. M. P., 1991. *Astrophys. J. Lett.*, **368,** L23.
Cowie, L. L., Henriksen, M. & Mushotzky, R., 1987. *Astrophys. J.*, **317,** 593.
Edge, A. C. & Stewart, G. C., 1991. *Mon. Not. R. astr. Soc.*, **255,** 428.
Edge, A. C., Stewart, G. C., Fabian, A. C. & Arnaud, K. A., 1990. *Mon. Not. R. astr. Soc.*, **245,** 559.
Efstathiou, G., Davis, M., Frenk, C. S. & White, S. D. M., 1985. *Astrophys. J. Suppl.*, **57,** 241.
Evrard, A. E., 1990. *Astrophys. J.*, **363,** 349.
Fabian, A. C., Nulsen, P. E. J. & Canizares, C. R., 1991. *Astr. Astrophys. Rev.*, **2,** 191.
Fabian, A. C., Hu, E. M., Cowie, L. L. & Grindlay, J., 1981. *Astrophys. J.*, **248,** 47.
Fong, R., Stevenson, P. R. F. & Shanks, T., 1990. *Mon. Not. R. astr. Soc.*, **242,** 146.
Frenk, C. S., White, S. D. M., Efstathiou, G. E. & Davis, M., 1990. *Astrophys. J.*, **351,** 10.
Gingold, R. A. & Monaghan, J. J., 1977. *Mon. Not. R. astr. Soc.*, **181,** 375.
Gioia, I. M., Henry, J. P., Maccacaro, T., Morris, M. L., Stocke, J. T. & Wolter, A., 1990. *Astrophys. J. Lett.*, **356,** L35.
Heisler, J., Tremaine, S. & Bahcall, J. N., 1985. *Astrophys. J.*, **298,** 8.
Henry, J. P. & Arnaud, K. A., 1991. *Astrophys. J.*, **372,** 410.
Hernquist, L. & Katz, N., 1989. *Astrophys. J. Suppl.*, **70,** 419.
Hoffman, Y., 1986. *Astrophys. J.*, **301,** 65.
Hughes, J. P., Yamashita, K., Okamura, Y., Tsunami, H. & Matsuoka, M., 1988. *Astrophys. J.*, **327,** 615.
Icke, V., 1984. *Mon. Not. R. astr. Soc.*, **206,** 1P.

Jones, C. & Forman, W., 1984. *Astrophys. J.*, **276,** 38.
Lambas, D. G., Groth, E. J. & Peebles, P. J. E., 1988. *Astr. J.*, **95,** 996.
Lambas, D. G., Nicrota, M., Muriel, H. & Ruiz, L., 1990. *Astr. J.*, **100,** 1006.
Lucy, L., 1977. *Astr. J.*, **82,** 1013.
Merrifield, M. R. & Kent, S. M., 1989. *Astr. J.*, **98,** 351.
Mushotzky, R., 1984. *Phys. Scripta,* **T7,** 157.
Mushotzky, R., 1988. In: *Hot Astrophysical Plasmas,* p. 273, ed. Pallavicini, R., Kluwer, Dordrecht.
Palmer, P. L., 1983. *Mon. Not. R. astr. Soc.*, **202,** 561.
Pearce, F. R. & Thomas, P. A., 1991. *Mon. Not. R. astr. Soc.*, **250,** 607.
Peebles, P. J. E., 1980. *The Large-scale Structure of the Universe,* Princeton University Press, Princeton, New Jersey.
Peebles, P. J. E., 1990. *Astrophys. J.*, **365,** 27.
Plionis, M., Barrow, J. D. & Frenk, C. S., 1991. *Mon. Not. R. astr. Soc.*, **249,** 662.
Porter, A. C., 1988. *PhD thesis,* California Institute of Technology.
Press, W. H. & Schechter, P., 1974. *Astrophys. J.*, **187,** 425 (PS).
Pringle, J. E., 1989. *Mon. Not. R. astr. Soc.*, **239,** 479.
Rees, M. J. & Ostriker, J. P., 1977. *Mon. Not. R. astr. Soc.*, **179,** 541.
Rhee, G. F. R. N. & Roos, N., 1989. *Astrophys. Space Sci.*, **157,** 201.
Rhee, G. F. R. N. & Roos, N., 1990. *Mon. Not. R. astr. Soc.*, **243,** 629.
Sarazin, C. L., 1986. *Rev. Mod. Phys.*, **58,** 1.
Silk, J., 1977. *Astrophys. J.*, **211,** 638.
Struble, M. F., 1987. *Astrophys. J.*, **317,** 688.
Thomas, P. A., 1988. *Mon. Not. R. astr. Soc.*, **235,** 315.
Thomas, P. A., Fabian, A. C. & Nulsen, P. E. J., 1987. *Mon. Not. R. astr. Soc.*, **228,** 973.
West, M. J. & Bothun, G. D., 1990. *Astrophys. J.*, **350,** 36.
White, S. D. M. & Rees, M. J., 1978. *Mon. Not. R. astr. Soc.*, **183,** 341.