Mon. Not. R. Astron. Soc. **265,** 271–300 (1993)

# Simulations of dissipative galaxy formation in hierarchically clustering universes – I. Tests of the code

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

Accepted 1993 May 11. Received 1993 May 6; in original form 1993 January 14



**ABSTRACT**

We present tests of a code designed to simulate the evolution of self-gravitating fluids in three dimensions. The code is based on the smoothed-particle hydrodynamics (SPH) technique for solving the hydrodynamical equations, together with a binary tree method for computing gravitational forces. Our tests are relevant to the evolution of non-linear structure in hierarchically clustering universes. In particular, we study the collapse and merger of quasi-spherical, slowly rotating clumps containing a mixture of gas and collisionless dark matter. For self-similar spherical collapse the detailed solution structure is already known, but we also study more realistic situations for which this is not the case. We analyse the behaviour of simulations with varying numbers of particles ($N$) in order to identify the minimum $N$ for which the results are meaningful and stable. We find that the adiabatic collapse and virialization of an isolated quasi-spherical clump requires more than $\sim 100$ gas particles in order to avoid substantial spurious energy transfers between the gas and the dark matter. Similar numbers are needed to obtain a qualitatively correct description of the merging of clumps. With $N_{\rm gas} > 300$, our simulations converge to give results with only small residual systematic deviations. However, in our experiments the gas always appears to gain energy at the expense of the dark matter. This transfer has a physical rather than a numerical origin. In a collision its magnitude increases with the initial orbital energy, and it can produce significant differences in the spatial distributions of gas and dark matter in the resulting merger remnant, even if the initial distributions are identical. Finally, we test the numerical reliability of several approximate schemes for modelling additional physical processes that are likely to play a key role in the evolution of the baryonic component of galaxies, namely radiative cooling, star formation, and the injection of energy by supernovae. In particular we study how the qualitative outcome of an experiment is influenced by the detailed manner in which such modelling is carried out. Unfortunately, we find a strong sensitivity to some aspects of what can at best be a very crude representation of these processes.

**Key words:** hydrodynamics – methods: numerical – galaxies: clustering – galaxies: formation – cosmology: theory – dark matter.

## 1 INTRODUCTION

*N*-body simulations have become a major research tool over the last decade, and have enabled quantitative testing of various theories of cosmogony. This became possible through the development of fast computers and the construction of efficient numerical techniques for solving the gravitational many-body problem. Until recently, however, simulations of the formation of structure in the Universe concentrated on the evolution of the 'dark' collisionless component, treating its relation to other components through rather simplistic models. This shortcoming resulted in substantial uncertainties when comparing results with observation.

Although *N*-body techniques allow the statistical properties of the dark matter distribution to be predicted in detail, a link to observation requires assumptions about how the luminous material (i.e. stars and gas) traces the underlying mass. The problem is that the baryons which make up this material might have followed a distinctly different

evolutionary path from the dark matter. This evolution can only be simulated with codes that solve the hydrodynamics equations governing the evolution of collisional fluids. In addition, star formation itself must be modelled, together with the source and sink terms it provides to the gas. This is a much more complicated problem than pure gravitational *N*-body dynamics, and codes capable of a proper treatment of the hydrodynamics have only recently become available and are at present still in the early stage of experimental testing (Evrard 1988; Hernquist & Katz 1989; Cen et al. 1990; Navarro & Benz 1991).

Eulerian grid-based techniques have traditionally been the most important tool in 3D simulations of fluid dynamics. Despite the sophistication attained by codes based on this approach, such codes still lack the resolution and flexibility necessary to tackle the formation of structure in a hierarchically clustering universe, where a high degree of substructure is present at all stages of the evolution. This property is a particularly serious limitation for models of galaxy formation, and it greatly reduces the usefulness of schemes where Poisson's equation is solved via fast Fourier transforms (FFTs) since these require periodic boundary conditions, and hence forces the simulation of large volumes surrounding the galaxy of interest.

Particle methods such as smoothed-particle hydrodynamics (SPH) appear better suited for these problems due to their Lagrangian nature, the lack of imposed symmetries, and their large dynamic range (for a review of SPH, see Benz 1990). In SPH, the fluid is divided up into particles which act as interpolation centres and is completely characterized by their position in phase-space, their mass, and their internal (thermal) energy. The remaining hydrodynamical variables can be calculated at any point in space through proper interpolation and the use of a chosen equation of state.

When coupled with a fast algorithm to compute the gravitational forces, this method provides one of the most promising ways to simulate galaxy formation. Tree data structures provide such an efficient and flexible Poisson solver, and are also well adapted for neighbour location, an essential aspect of SPH calculations. In combination with SPH they give a Lagrangian code that offers speed and accuracy, is free from grid-related limitations and has variable resolution both in space and in time. The fusion of one version of SPH with the Barnes–Hut tree algorithm was performed by Hernquist & Katz (1989), and shown to be potentially very powerful in subsequent applications (Hernquist 1989; Katz & Gunn 1991). The code used for the simulations described in this paper combines a mutual nearest-neighbour binary tree (Benz et al. 1990) with the formulation of SPH outlined in Section 2.

Simulation of the formation of structure in a hierarchical clustering model is a difficult challenge for these numerical techniques. Different levels of the hierarchy coexist at any given epoch, and the characteristic clustering scale increases rapidly with time (e.g. Efstathiou et al. 1988). Thus a simulation must be able to handle many different scales simultaneously if it is to follow the evolution of the hierarchy accurately. On scales smaller than a few particles, the approach surely breaks down because the internal structure of these clumps cannot be resolved. However, it is unclear what the consequences of these limitations are for the larger structures that form later. It is thus important to identify the lowest level of the hierarchy that can be considered adequately resolved, to study the systematic errors that are made on lower levels, and to investigate how these errors affect the formation of larger systems.

In this paper we restrict ourselves to tests that mimic, in a simplified way, the formation of individual non-linear clumps in a hierarchically evolving universe: the collapse, virialization and merger of rotating, near-spherical clumps. Several additional processes that are likely to be of crucial importance on galaxy scales are also explored. These include the influence on the gas of a collisionless dark matter halo, the effect of radiative cooling, the transformation of gas into stars, and energy injection from supernova explosions.

Unfortunately, detailed solutions are not known for analogues of most of the physical situations of interest. Thus we are forced to assess the accuracy of our simulations by comparing the results of equivalent experiments realized with different numerical parameters, primarily different particle numbers. A necessary condition for the simulations to be reliable is that they should converge to a unique solution as the number of particles increases. Self-similar spherical infall may be the only relevant situation for which the detailed behaviour is well known, thanks to the similarity solutions computed by Bertschinger (1985). A variety of other tests of SPH codes have been presented by Evrard (1988) and Hernquist & Katz (1989). The main contribution of this paper is to study situations that are directly relevant to hierarchical models for galaxy formation, and to clarify how simulations of this process will be affected by the inevitable lack of resolution on small mass-scales.

There are compelling reasons why star formation rates and energy injection by supernovae should play an important role in galaxy formation (e.g. Cole 1991; White & Frenk 1991). However, our poor understanding of star formation and of its effects on the surrounding gas forces us to adopt crude and simplistic representations of these processes, and so introduces substantial uncertainties into any attempt to make realistic models. Thus, although SPH simulations can greatly clarify some of the hydrodynamic and dissipative aspects of galaxy formation, it appears that such uncertainties may severely limit our ability to reach definitive conclusions about the origins of real galaxies.

This paper is organized as follows. In Section 2 we give a brief description of our SPH code and of the algorithms we use to implement radiative cooling, star formation, and feedback from supernovae. Sections 3, 4 and 5 then present our tests. Section 3 contains a comparison with the spherical-infall similarity solution. Section 4 studies the collapse and merger of adiabatic clumps of gas and dark matter. Section 5 explores how such collapse models are modified by the addition first of radiative cooling, then of star formation and feedback. In each case we study how results break down as the number of particles used in the simulation is reduced. Section 6 summarizes the results and presents our main conclusions.

## 2 THE CODE

### 2.1 Kernel estimates

SPH is an ingenious Monte-Carlo-like technique to solve the hydrodynamic equations. The system is divided up into

particles which carry thermodynamic information and whose locations are used to compute the values of the hydrodynamical quantities through kernel interpolation. A derivation of the relevant formulae will not be reproduced here, but may be found in the excellent reviews by Monaghan (1985) and Benz (1990). This subsection is largely based on the latter paper, which should be consulted for further details. Our aim here is to provide sufficient detail to allow a reconstruction of our specific method.

The smoothed density field at the position of particle $i$ is computed from the locations and masses of all the particles using

$$\langle \rho(\boldsymbol{r}_i) \rangle = \rho_i = \sum_{j=1}^{N} m_j W(|\boldsymbol{r}_i - \boldsymbol{r}_j|, h_{ij}), \tag{1}$$

where $W(r, h)$ is the kernel function,

$$W(r, h) = \frac{1}{\pi h^3} \times \begin{cases} 1-(3/2)(r/h)^2+(3/4)(r/h)^3, & \text{if } 0 \le (r/h) \le 1; \\ (1/4)[2-(r/h)]^3, & \text{if } 1 < (r/h) \le 2; \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

and $h_{ij} = (1/2)(h_i + h_j)$ is the pair-averaged smoothing length. The kernel function is normalized so that $\int_0^\infty W(r, h)\, \mathrm{d}^3 r = 1$. Since $W(r, h)$ has compact support, the sum (1) is actually performed over a limited number of neighbouring particles, namely those that are closer to $\boldsymbol{r}_i$ than $2h_{ij}$. The actual value of the smoothing length $h_i$ for each particle is determined by requiring that the number of neighbours remains between fixed limits. Experimentation has yielded $n_{\min} \sim 30$ and $n_{\max} \sim 50$ as good values for the lower and upper bounds for the number of neighbours, respectively.

Particle velocities, $\boldsymbol{v}_i$, are defined simply as the rate of change of their position. [Note, however, that if the continuity equation is to be satisfied the local momentum density must be defined through the equivalent of equation (1) with the particle velocities, $\boldsymbol{v}_j$, inserted into the sum on the right-hand side. This momentum density is then *not* equal to $\rho_i \boldsymbol{v}_i$. In practice the continuity equation is never considered explicitly, and conservation of mass is trivially implemented by keeping the particle masses constant.] The only other quantity that needs to be specified is the internal energy (per unit mass) $u_i$, since all remaining thermodynamic variables can then be computed using an appropriate equation of state. For the tests described in this paper, we chose an equation of state corresponding to an ideal gas, $P = (\gamma - 1)\rho u$, with $\gamma = 5/3$. The specific internal energy of a particle can be readily related to its temperature $T$ by the relation $u = kT/(\gamma - 1)\mu m_{\mathrm{p}}$, where $k$ is Boltzmann's constant, $\mu$ is the average molecular weight and $m_{\mathrm{p}}$ is the proton mass.

### 2.2 The hydrodynamical equations

The acceleration of particle $i$ is determined by Euler's equation,

$$\frac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} = -\frac{\boldsymbol{\nabla}(P_i + P_{i,\mathrm{visc}})}{\rho_i} - \boldsymbol{\nabla}\Phi_i,$$

where $\Phi$ is the gravitational potential, $P$ is the pressure, and $P_{\mathrm{visc}}$ is the artificial viscous pressure used to allow for the dissipation of kinetic energy into heat via shocks. If particle $i$ is a collisionless particle representing either dark matter or stars, the pressure gradient contribution is, of course, zero. The pressure gradient term can be written (to second order in $h$) as

$$\frac{\boldsymbol{\nabla}(P_i + P_{i,\mathrm{visc}})}{\rho_i} = \sum_{j=1}^{N} m_j \left( \frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij} \right) \times \boldsymbol{\nabla}_i W(|\boldsymbol{r}_i - \boldsymbol{r}_j|, h_{ij}), \tag{3}$$

where

$$\Pi_{ij} = \begin{cases} \dfrac{-\alpha c_{ij} \mu_{ij} + \beta \mu_{ij}^2}{\rho_{ij}} & \text{if } (\boldsymbol{v}_i - \boldsymbol{v}_j)\cdot(\boldsymbol{r}_i - \boldsymbol{r}_j) \le 0; \\ 0 & \text{otherwise} \end{cases}$$

is the viscous contribution. We have used $c_{ij} = (1/2)(c_i + c_j)$ to indicate the mean pair sound speed, and $\rho_{ij} = (1/2)(\rho_i + \rho_j)$ for the pair-averaged density. For our implementation of the artificial viscosity we need an estimate of the velocity divergence at particle $i$ due to the presence of particle $j$, which is given by

$$\mu_{ij} = \frac{h_{ij}(\boldsymbol{v}_i - \boldsymbol{v}_j)\cdot(\boldsymbol{r}_i - \boldsymbol{r}_j)}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^2 + \epsilon h_{ij}^2}. \tag{4}$$

Here, the term $\epsilon h_{ij}^2$ was introduced only to prevent divergences when $|\boldsymbol{r}_i - \boldsymbol{r}_j|$ becomes very small. The dimensionless constants $\alpha$ and $\beta$ have to be calibrated by numerical experiments with known analytical solutions. For the tests described in this paper, we have taken $\alpha = 1$, $\beta = 2.5$ and $\epsilon = 0.01$.

Equation (3) shows clearly one of the main advantages of the SPH technique. The spatial gradient calculation is reduced to a sum over the neighbours weighted by the gradient of the kernel, which can be easily tabulated. Thus, if an efficient way of finding particle neighbours is available, the calculation of the forces in an SPH code is straightforward. When the gravitational potential gradient is evaluated by searching a tree structure, the same tree search can be used to find the neighbours, a procedure that proves very efficient in practice.

The tree used in the present version of the code is a mutual-nearest-neighbour binary tree. The particles themselves constitute the lowest layer of nodes; the next layer is built from the particle pairs that are mutual nearest neighbours. Nodes are located at the centre of mass of a pair and quadrupole moments are stored. The computation of the rest of the tree is analogous, further nodes being constructed from particle–particle pairs or particle–node pairs, as appropriate. Ultimately, one node will sit at the top of the tree. For each node $i$ we also store a bounding radius $R_i$, defined as the radius of a sphere centred at the node's position, which is guaranteed to contain all the subnodes and particles that are constituents of the node. When computing gravitational forces on particle $j$, for example, nodes need only be 'opened' if they are closer to $j$ than a certain distance determined by the accuracy we would like to achieve. Speci-

fically, node $i$ is opened only if its distance to $j$ is less than $R_i/f_{\rm acc}$, where $f_{\rm acc}$ is a constant. Gravitational accelerations accurate to $10^{-4}$ are obtained when $f_{\rm acc}\approx 1$, while smaller values lead to even better accuracy at the expense of increased computing time. We use $f_{\rm acc}=0.4$ in all simulations reported in this paper. Details of the tree construction can be found in Benz et al. (1990).

Gravitational interactions between neighbours are smoothed using the same kernel (2), but with independently set smoothing lengths, $h_{\rm grav}$, which can have different values for each particle. Gravitational softening lengths do not change with time, and constitute a minimum value for the variable smoothing length, $h$, of the gas particles. The choice of a minimum value for $h$ is similar to setting a maximum value for the density and, indirectly, a minimum value for the time-step. The justification for such a procedure is that it would be wasteful to estimate the pressure gradients with higher resolution than the gravitational potential in regions where softening is important.

The corresponding equation for the evolution of the internal energy is

$$\frac{{\rm d}u_i}{{\rm d}t}=\sum_{j=1}^{N} m_j\left(\frac{P_i}{\rho_i^2}+\frac{\Pi_{ij}}{2}\right)(\boldsymbol{v}_i-\boldsymbol{v}_j)\cdot\boldsymbol{\nabla}_i W(|\boldsymbol{r}_i-\boldsymbol{r}_j|,h_{ij}) \qquad (5)$$

$$-\frac{\Lambda(u_i,\rho_i)}{\rho_i}.$$

Here, $\Lambda(u_i)$ stands for the cooling (or heating) rates associated with non-adiabatic processes other than shocks, for example radiative losses. In this paper we approximate this function as $\Lambda(T)$, the cooling rate of an optically thin 'primordial' mixture of H and He in collisional ionization equilibrium at temperature $T$ in the absence of any significant background radiation fields (e.g. Fall & Rees 1985).

The last equation to be integrated corresponds to that governing the time evolution of the gas particles' individual smoothing lengths. Since $\rho$ roughly scales as $h^{-3}$, we write it as

$$\frac{{\rm d}h_i}{{\rm d}t}=-\frac{1}{3}\frac{h_i}{\rho_i}\frac{{\rm d}\rho_i}{{\rm d}t}=\frac{1}{3}h_i\boldsymbol{\nabla}\cdot\boldsymbol{v}_i, \qquad (6)$$

where we have used the continuity equation for the second equality. This equation is solved at each time-step together with Euler's equation and the energy equation. Almost no extra time is spent solving (6) because the divergence of the velocity is a quantity that has to be computed anyway (through equation 4). In practice, equation (6) performs very well and only occasional minor adjustments are needed to keep the number of neighbours between $n_{\rm min}$ and $n_{\rm max}$.

It can be shown that these equations conserve a suitably defined total energy, and, because forces between particle pairs are symmetric and parallel to their separation, the total linear and angular momenta are also conserved. It should be noted that equations (1)–(6) are not the only possible choice, and in fact several alternatives for the smoothing kernel and the specific form of the equations of motion can be found in the literature. The reviews by Monaghan (1985) and Benz (1990) list some of the possibilities, none of which appears to be clearly superior to the rest.

### 2.3 Time-stepping

The time integration is performed using a second-order Runge–Kutta method with individual time-steps $\Delta t_i$. The position, velocity, internal energy and smoothing length of particle $i$ are updated from $t$ to $t+\Delta t$ using

$$\boldsymbol{r}_i^{t+\Delta t}=\boldsymbol{r}_i^t+f\boldsymbol{v}_i^{t+\Delta t/2}\Delta t+(1-f)\boldsymbol{v}_i^t\Delta t,$$
$$\boldsymbol{v}_i^{t+\Delta t}=\boldsymbol{v}_i^t+f\dot{\boldsymbol{v}}_i^{t+\Delta t/2}\Delta t+(1-f)\dot{\boldsymbol{v}}_i^t\Delta t,$$
$$u_i^{t+\Delta t}=u_i^t+f\dot{u}_i^{t+\Delta t/2}\Delta t+(1-f)\dot{u}_i^t\Delta t,$$
$$h_i^{t+\Delta t}=h_i^t+f\dot{h}_i^{t+\Delta t/2}\Delta t+(1-f)\dot{h}_i^t\Delta t,$$

where $f$ is a constant that depends on the particular implementation of the Runge–Kutta method, and dotted symbols denote time derivatives. For the calculations in this paper we chose $f=0.996$. Note that, in order to advance a particle, we need to compute the value of the derivatives at half time-step, $t+\Delta t/2$. Since this requires the knowledge of all variables at $t+\Delta t/2$, we linearly predict them using the values of the derivatives at $t$. That is,

$$x_i^{t+\Delta t/2}=x_i^t+\dot{x}_i^t\Delta t/2,$$

where $x$ is $u_i$, $h_i$, or any component of the position or velocity.

A quantitative estimate of the error made at each time-step can be obtained from the relative size of higher order terms through the expression

$$\delta_x=\frac{|\dot{x}^{t+\Delta t}-\dot{x}^t|\,\Delta t}{|x|+\epsilon}\sim\frac{|\ddot{x}|\Delta t^2}{|x|},$$

where $\epsilon$ is a small positive number used to prevent divergences. At each time-step and for each particle we compute the maximum, $\delta$, of all the $\delta_x$s. We compare this with a predetermined tolerance, $\delta_{\rm m}$, in order to determine a new time-step, $\Delta t$, using

$$\Delta t_{i,\rm new}=\Delta t_{i,\rm old}(\delta_{\rm m}/\delta)^{1/2}. \qquad (7)$$

In practice we only allow time-steps that are a power of 2 subdivision of some fundamental time-step, $\Delta t_i=\Delta t_{\rm max}/2^n$. Thus we reduce $\Delta t$ by a factor of 2 whenever $\delta$ exceeds $\delta_{\rm m}$, but we increase it by a factor of 2 only when the result is less than the time-step suggested by equation (7). In addition, we only increase the time-step after an even number of smaller steps. In this way all particles are synchronized after each full time-step, $\Delta t_{\rm max}$.

Although all particles are given the same (small) time-step at the beginning of a calculation, a hierarchy of $\Delta t$ values soon develops. It is then necessary to keep track of the times, $t_i$, at which the variables of each particle were last updated. In order to identify at each time, $t$, those particles for which derivatives need to be computed, we build a list for which either $t_i=t$ (particles already advanced to the current time) or $t_i+\Delta t_i/2=t$ (particles for which mid-step derivatives are required at the current time). If the number of particles in this list is small, a full tree construction is unnecessary, and may be replaced by an update of node quantities; this significantly reduces the required CPU time. When calculating derivatives, the values of variables for particles not yet updated to time $t$ are linearly predicted as explained above. The use of individual time-steps in this way greatly improves

the performance of the code when there are large variations in time-step within the system. This is the normal situation when density and temperature span several orders of magnitude at any given time, as in the problems that interest us here.

A Runge–Kutta integrator has been preferred over the more popular leap-frog for two reasons. First, it provides us with a quantitative estimate of the error made at each time-step, which can be used to determine the time-step of a particle, as indicated above. Secondly, the SPH accelerations depend on the velocity through the artificial viscosity. Even if we used a leap-frog integrator we would therefore still need to compute accelerations at mid-step to maintain the second-order accuracy. Tests show that the Runge–Kutta scheme presented above is at least as fast as a leap-frog, with the extra advantage of being easier to implement, especially when individually adjusting time-steps are used.

If gas densities are high and radiative cooling is allowed, the time-steps computed from (6) for the internal energy, $u_i$, can be very small. In fact, they may prove to be several orders of magnitude smaller than the local dynamical time-scale. In such cases, we resort to an implicit integration of the energy equation (5) in which the internal energy of each particle is updated iteratively. Our method is based on that presented by Hernquist & Katz (1989).

### 2.4 Star formation

The treatment of star formation is obviously a fundamental and critical part of any simulation of galaxy formation, since it determines, to a very large extent, the structure of the final observable object. It provides a source term for stars, and both sink and source terms for the mass, momentum and energy of the gas. Unfortunately, it is unrealistic to hope to follow even the qualitative features of the small-scale hydrodynamics of dense star-forming regions. Our understanding of how large-scale conditions determine star formation rates, and of how stellar winds and supernovae modify the composition and thermodynamic state of gas in their environment, is so poor that only very simple schematic representations of these processes can be justified. Even the qualitative features of the resulting simulations must be regarded as suspect. However, we may hope to identify reliable features of the evolution by studying the sensitivity of results to the assumptions and parameters of the star formation and feedback models. The comparison of simulations with starbursts, merging galaxies and cluster cooling flows should constrain the kinds of star formation and feedback models that are consistent with these observed analogies of galaxies in formation. Clearly, much work will need to be done before simulations of this kind can be truly convincing.

Our procedures in the present paper are similar to those of Katz (1992). Once a region where star formation is likely has been identified, we 'create' new collisionless particles ('stars') in discrete steps and at a rate determined by the local properties of the gas. Thereafter, these stars affect the surrounding gas in ways that model the effects of exploding supernovae and of stellar winds. Once formed, stars are treated as collisionless particles; their motions are affected only by gravity.

To be more specific, a gas particle is considered to be a potential star-forming region if it is part of a sufficiently dense, collapsing subsystem. Such a subsystem is Jeans-unstable if its local sound crossing time is larger than its dynamical time. Pressure forces are then unable to halt the collapse. Particle $i$ is in a collapsing subregion if

$$\nabla\cdot\boldsymbol{v}_i<0. \tag{8}$$

This subregion will be Jeans-unstable on the smallest resolvable scale if

$$\tau_{\rm sound}=\frac{h_i}{c_i}>\tau_{\rm dyn}=\sqrt{\frac{3\pi}{16\,G\rho_i}}\,. \tag{9}$$

In addition, it will cool quickly and remain cool as it collapses provided that

$$\tau_{\rm cool}=\frac{3(\rho_{\rm gas}/\mu m_{\rm p})\,kT}{2\Lambda(T)}\ll\tau_{\rm dyn}. \tag{10}$$

In practice, condition (10) is readily satisfied if the gas density exceeds some threshold value, $\rho_{\rm crit}$ say, and its temperature is larger than the cut-off temperature of the cooling function, $T_{\rm cut}=10^4$ K. In the present work we therefore adopt a convergent flow and a density exceeding $\rho_{\rm crit}$ as the necessary and sufficient conditions for gas to be eligible to make stars. Choosing $\rho_{\rm crit}=7\times10^{-26}$ g cm$^{-3}$, the cooling time in regions with $\rho>\rho_{\rm crit}$ will satisfy

$$\tau_{\rm cool}(T,\rho)<\tau_{\rm cool}(10^{6.2}\ {\rm K},\rho_{\rm crit})=2.5\times10^7\ {\rm yr},$$

since $\Lambda(T)/T$ has a minimum for $T=10^{6.2}$ K. The dynamical time-scale is $3.6\times10^8$ yr for a region of density $\rho_{\rm crit}$, and decreases less rapidly with increasing density than does the cooling time. Thus the cooling time in such dense regions is always shorter than the dynamical time by more than an order of magnitude. Such regions cool rapidly to $T_{\rm cut}$, at which point they almost always satisfy condition (9).

The next step is to set the star formation rate, i.e. to decide how and how fast eligible gas particles are turned into stars. We adopt a rate of the form

$$\frac{{\rm d}\rho_{\rm gas}}{{\rm d}t}=-\frac{{\rm d}\rho_{\rm stars}}{{\rm d}t}=-\frac{\rho_{\rm gas}}{t_*}\,, \tag{11}$$

where $t_*$ is a characteristic time-scale chosen to be equal to the local dynamical time, $\tau_{\rm dyn}$, the time-scale for local dynamical collapse.

Numerical constraints force us to transform gas into stars in discrete steps. According to equation (11), the gas density is expected to decrease by a factor of 2 in a time $\tau\sim\ln 2\,t_*$. In order to mimic this evolution, a gas particle is assumed to give birth to a star particle containing half its mass at $t_i=t_{i,\rm stf}+\tau_i$, where $t_{i,\rm stf}$ is either the time at which it became eligible to form stars, or the time at which it last formed a star, whichever is latest. The mass of the parent gas particle is reduced accordingly, and we allow each gas particle to form a maximum of four stars. A newly born star particle inherits from its parent its phase-space coordinates and its gravitational softening length.

For a given initial mass function (IMF), we can calculate how much energy a star particle is expected to produce as supernovae explode. Adopting a power-law IMF with slope

1.5 and upper and lower limits of 40 and 0.1 $M_\odot$, respectively, and assuming that every star more massive than 8 $M_\odot$ ends as a supernova of energy $10^{51}$ erg, we find that $\sim 4\times 10^{48}$ erg is dumped into the surrounding gas for every $M_\odot$ turned into stars. Our code adds this energy bit by bit at each time-step according to the lifetimes of stars of different mass,

$$\log[\tau(m_*)] = 10 - 3.42\log(m_*) + 0.88[\log(m_*)]^2$$

(where $m_*$ is in solar masses and $\tau(m_*)$ in years; David, Forman & Jones 1990). Since only stars with $m_* > 8$ are assumed to explode, a star particle stops releasing energy $\tau(8\ M_\odot) = 4.3\times 10^7$ yr after its creation.

One of the most difficult and most critical processes to model in galaxy formation simulations is the way in which feedback from supernovae affects the surrounding gas. Unfortunately, there is no good understanding of how this coupling occurs or of how it should be modelled. In real star-forming systems supernovae are largely responsible for the multiphase structure of the interstellar medium. Although they are associated with the coolest and densest gas clouds, they can deposit much of their energy in a strong, hot, low-density wind. Our techniques cannot properly resolve this complex 'microstructure' in the gas, and we find below that qualitatively different results are obtained for different plausible, but essentially ad hoc, treatments of this problem. For simplicity, we assume that the energy produced by supernovae affects only the temperature and velocity field of the surrounding gas. We implement this assumption by increasing the internal and kinetic energy of the gas neighbours of each 'active' star by an amount corresponding to the energy released by supernovae during each time-step ($\Delta E_{\rm SN}$).

To be specific, after every time-step the velocities of the gas neighbours are altered by an amount $\Delta \boldsymbol{v}_{\rm SN}$, computed from

$$\Delta E_{\rm SN,vel} = f_v \Delta E_{\rm SN} = \sum_{\rm neigh} m_i (\boldsymbol{v}_i \cdot \Delta \boldsymbol{v}_{{\rm SN},i} + \frac{1}{2}\Delta v_{{\rm SN},i}^2), \qquad (12)$$

where $\boldsymbol{v}_i$ is the neighbour velocity relative to the star, $\Delta \boldsymbol{v}_{{\rm SN},i}$ is a velocity perturbation directed radially *away* from the star and assumed to have the same magnitude for all neighbours, and $f_v$ is an input parameter which controls the fraction of the available energy that we allow to perturb the gas velocity field. The rest of the energy of the supernovae, $(1-f_v)\Delta E_{\rm SN}$, is divided equally among the internal energies of these same neighbours. Although this implementation conserves energy, it has the disadvantage of reversing convergent gas flows for any value of $f_v > 0$. This induces large changes in the local gas velocity field, even for small values of $f_v$. We have therefore also explored a milder alternative in which the momenta of gas neighbours relative to the star are perturbed by the amount derived from equation (12) when the $\boldsymbol{v}_i$ are set to zero. The magnitude of the $\Delta \boldsymbol{v}_{{\rm SN},i}$ is then simply

$$\Delta v_{\rm SN}^2 = 2 f_v \Delta E_{\rm SN} \Big/ \left(\sum_{\rm neigh} m_i\right), \qquad (13)$$

and the velocity field perturbation goes to zero for small $f_v$. We will hereafter refer to these two alternatives as the 'energy' and the 'momentum' implementations of feedback, respectively.

Of course, these two cases with their single parameter, $f_v$, are far from exhausting all the plausible possibilities. We find that our results depend quite strongly on the particular choice made. It does turn out that, for the tests described in this paper, it is possible to find parameter values for which star formation rates in the simulations are roughly consistent with observational constraints. However, the treatment of star formation and feedback remains the most uncertain part of our code, and will require much further investigation.

## 3 ADIABATIC COLLAPSE

One of the most important requirements for a hydrodynamics code for cosmological simulations is an ability to simulate the shock-heating of cold gas in collapsing regions. This process occurs as non-linear structures detach from the overall expansion and come to equilibrium. While the collisionless dark matter virializes in a couple of free-fall times, the gas is expected to convert most of its kinetic energy of collapse into heat in a single accretion shock.

In order to analyse this process we have carried out two series of tests. The first studies self-similar accretion on to a point mass in an otherwise unperturbed Einstein–de Sitter universe. Similarity solutions for this problem have been presented by Bertschinger (1985) for the case of a universe filled with cold gas, and also for the case where the universe is filled with a mixture of gas and collisionless dark matter. These solutions provide a good check on the accuracy of our techniques. The second series of tests deals with the collapse of an initially rotating sphere of cold gas and dark matter. Again, most of the mass is assumed to be in the collisionless component. Unfortunately, an analytic solution is not available for this problem, but we compare the results of simulations realized with different numbers of particles to check that they converge to a unique solution as $N$ (and presumably the reliability of our results) increases. In both series of tests we treat the gas as adiabatic, postponing until Section 5 any consideration of the additional complications introduced by cooling, star formation and feedback.

### 3.1 Self-similar spherical accretion

The initial conditions of these tests assume that at some initial time $t_i$ a flat pressure-free universe is unperturbed except for a sphere of radius $R_i$ with overdensity $\delta\rho/\rho = \delta_i > 0$. The initial velocity field is pure unperturbed Hubble flow, $v = H_i r$, where $H_i = 2/3\, t_i$ and $r$ is measured from the centre of the sphere. Each fluid element, initially outside $R_i$, eventually stops expanding and falls back on to the perturbation. Its maximum radius and its turnaround time are related by

$$r_{\rm ta}(t) = R_i \delta_i^{1/3} \left(\frac{4t}{3\pi t_i}\right)^{8/9}.$$

If the fluid is collisional, this element does not fall back to the origin but is stopped by an outwardly moving shock wave soon after turning around. According to Bertschinger's (1985) similarity solution, this shock always lies at a fixed fraction of the current turnaround radius, $\lambda_{\rm s} = r_{\rm shock}/r_{\rm ta} =$

0.34. Indeed, the structure of the solution is fully characterized by the dimensionless functions $V$, $D$ and $P$ defined by

$$\lambda(r,t)=\frac{r}{r_{\rm ta}(t)},$$

$$v(r,t)=\frac{r_{\rm ta}}{t}V(\lambda),$$

$$\rho(r,t)=\rho_{\rm H}D(\lambda),$$

$$p(r,t)=\rho_{\rm H}\left(\frac{r_{\rm ta}}{t}\right)^2P(\lambda),$$

where $v$, $\rho$ and $p$ stand for the velocity, density and pressure, respectively, and $\rho_{\rm H}$ is the unperturbed density $\rho_{\rm H}=(6\pi Gt^2)^{-1}$. The functions $V(\lambda)$, $D(\lambda)$ and $P(\lambda)$ are given, for different values of the adiabatic constant $\gamma$, in tables 1 and 2 of Bertschinger (1985).

Note that, once $R_i$ and $\delta_i$ have been chosen, this problem has *no* free parameters, making it an ideal test of our numerical methods. As the shock moves outwards it encompasses more and more mass [$r_{\rm shock}\propto t^{8/9}$, $M(r<t_{\rm shock})\propto t^{2/3}$]. This is a desirable feature since it enables us to assess the performance of the code in the post-shock regime with different numbers of particles. If our expectations are correct, the simulations should approximate the similarity solution more and more accurately as more particles contribute to the determination of fluid properties inside the shock radius.

The initial conditions for the simulations are set up as follows. We place $N_{\rm part}\sim 1800$ particles on a grid inside a sphere of radius $R$. Vacuum boundary conditions outside $R$ are assumed. Each particle is assigned the same mass $M_{\rm total}/N_{\rm part}$. An external potential is added in order to provide a roughly spherical central perturbation of mass $M_{\rm pert}=0.05\ M_{\rm total}$ and radius $R_i=0.1\ R$. The gravitational softening is chosen to be equal to $R_i$. $R_i$ and $M_{\rm pert}$ are the only parameters needed to specify the solution completely. When a dark matter component is included, the collisionless particles are set on top of the gas particles and the masses are reassigned in order to have 90 per cent of the total mass in the dark component. Dark matter particles are thus nine times more massive than gaseous ones. In all cases we evolve the system until roughly half its mass turns around in order to avoid complications arising from boundary effects when the shock reaches the outermost shells.

Fig. 1 shows spherically averaged profiles corresponding to the pressure, velocity and density fields at three different times. These times can be completely characterized by the number of gas particles inside the theoretical $r_{\rm shock}$: 100, 250 and 490 at the three times chosen. The data corresponding to these times are shown by open triangles, open squares and filled circles, respectively. The overall spatial resolution of the simulation can be represented by the mean value of the smoothing length, $\overline{h}$. Since $\overline{h}\propto t^{2/3}$ (like the scale factor in an Einstein–de Sitter universe) while $r_{\rm shock}\propto t^{8/9}$, the resolution of the shock in the simulations (as measured by $\overline{h}/r_{\rm shock}\propto t^{-2/9}$) improves as the shock moves outwards. The radii corresponding to $(r_{\rm shock}+\overline{h})$ are shown in Fig. 1 by small vertical arrows. At small radii, the resolution is limited by the smallest value taken by the smoothing length. This only affects the inner regions at the earliest time, and is indicated by the arrow labelled $h_{\rm min}$. For the other two times shown in Fig. 1, the $h_{\rm min}$ values lie outside the figure.

As Fig. 1 shows, the agreement between the analytic solution (dotted line) and the results of the simulation is remarkably good. The only systematic deviations are found at early times, when the number of particles in the post-shock region is small, and the effects of the limited resolution of the experiments are larger. In softened regions ($r<h_{\rm min}$), the code underestimates the true values of the density and pressure. At later times, when $r_{\rm shock}\gg h_{\rm min}$, the profiles are a very good match to the similarity solution. They also show that the code can reproduce variations in density and pressure spanning $\sim 3$ orders of magnitude with as few as $\sim 250$ particles.

Post-shock oscillations are inevitable in any numerical scheme, and they are clearly visible in the velocity profiles. Nevertheless, the right limiting behaviour is clearly present, and velocity oscillations in the post-shock regime are less than 1/3 of the pre-shock infall velocity. In fact, more than 90 per cent of the collapse kinetic energy has been effectively transformed into heat. This is true even at the earliest time shown, where the whole spherical shock-heating process is being simulated with $\sim 100$ particles.

The inclusion of a gravitationally dominant dark component does not seem to change the results greatly, except, perhaps, for enhanced velocity oscillations at small radii. These are most probably the result of fluctuations in the gravitational field induced by the small number of particles used. The inner pressure and temperature profiles at the first time also indicate differences between 'Gas + DM' and 'Gas only' simulations. A careful analysis reveals that this is due to spurious energy exchange between the gas and dark matter, which artificially pumps energy into the gas component, making it slightly more spatially extended. The mechanism behind this energy exchange is discussed in detail in the following section.

We conclude that our code can follow the shock-heating of initially cold gas relatively accurately. Steep gradients in density and pressure can be realized with $\geq 250$ particles, but systematic deviations are to be expected if fewer than 100 particles are used. However, the lack of adequate resolution during the initial phases of these simulations does not seem to compromise their later evolution, which agrees very well with Bertschinger's solutions.

### 3.2 Adiabatic collapse of a rotating sphere

The second series of tests follows the collapse of a rotating, spherical 'protogalaxy' with a dominant amount of dark matter (90 per cent by mass). Initially, the particles are set on a grid, and are perturbed radially to give a density profile of the form $\rho(r)\propto r^{-1}$. Velocities are chosen so that the sphere will be in solid-body rotation around the $z$-axis. The initial angular momentum ($J_{\rm tot}$) can be characterized by the value of the spin parameter $\lambda=J_{\rm tot}|E|^{1/2}/GM_{\rm tot}^{2.5}\sim 0.1$ ($E$ stands for the total energy of the system). We adopt a higher spin than expected from tidal torques in a hierarchically clustering universe ($\lambda\sim 0.03\rightarrow 0.07$) in order to emphasize the systematic effects that may arise in small-$N$ systems when rotation is present. Simulations were carried out using three different numbers of particles, $N_{\rm gas}=N_{\rm DM}=90$, 368 and

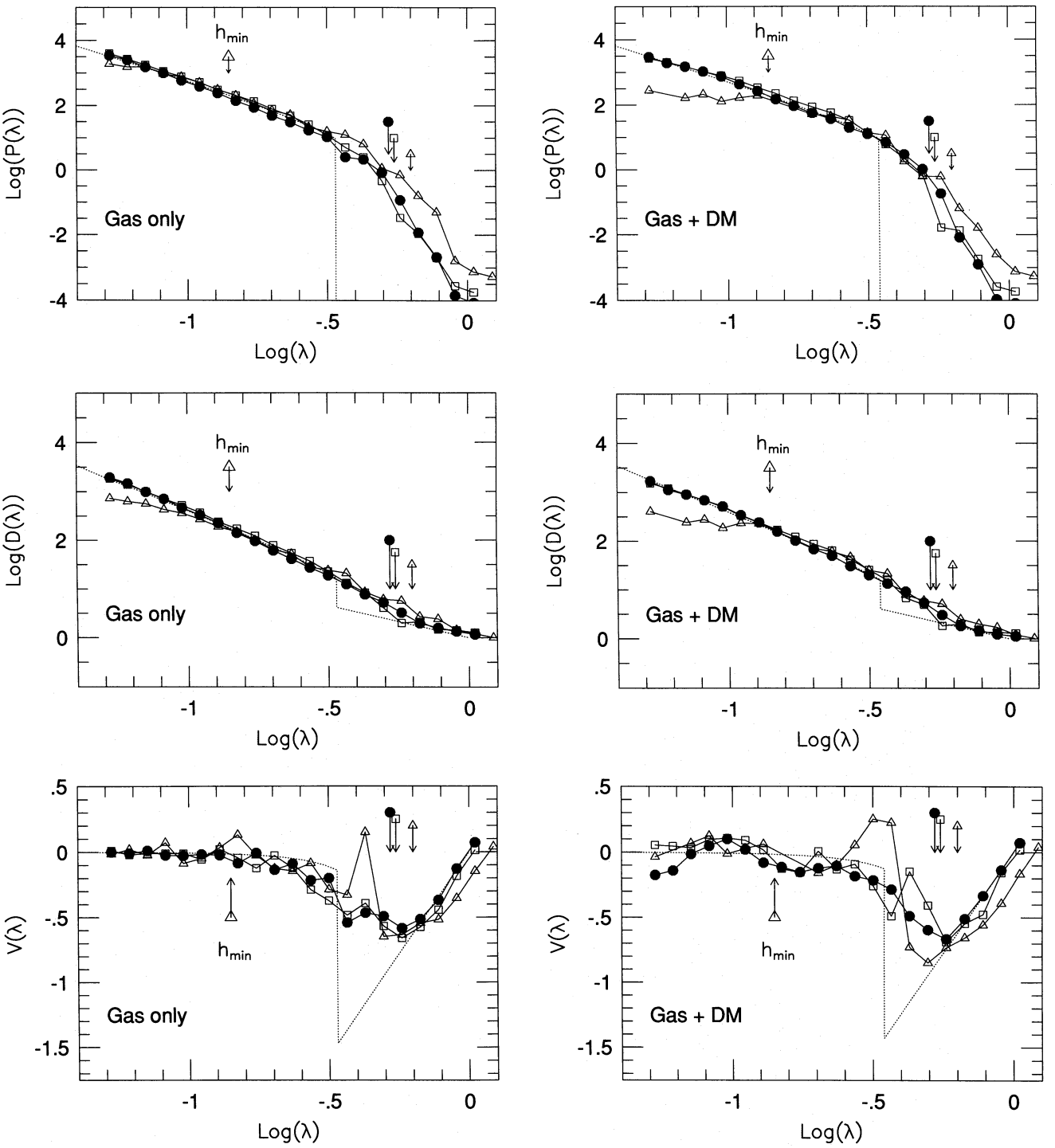


**Figure 1.** Dimensionless pressure, density and velocity profiles at three different times. These times are characterized by the number of particles inside $r_{\rm shock}$. Open triangles, open squares and filled circles correspond to $N(r<r_{\rm shock})$ equal to 100, 250 and 490, respectively. The profiles labelled 'Gas + DM' (where 'DM' denotes dark matter) refer to tests in which a dominant collisionless component was present. Vertical arrows to the right of the shock radius indicate the position of $r_{\rm shock}+\bar{h}$, as described in the text, and roughly indicate the resolution of the calculation at the shock location. The arrow labelled $h_{\rm min}$ indicates the resolution length in the inner parts of the model at the earliest time. For the other two times, $h_{\rm min}$ lies outside the figure.

1796. In each case the initial radius of the sphere was $R_{\rm tot}=100$ kpc, and its total mass was $M_{\rm tot}=10^{12}\,{\rm M}_{\odot}$, implying an rms velocity in virial equilibrium of 240 km $s^{-1}$. The dynamical time of the system is then $\tau_{\rm dyn}\sim 3.7\times10^{8}$ yr. For temperatures in the relevant range, $T\sim 10^{5}$–$10^{6}$ K, and for the densities attained during virialization, the cooling time of the gas is always substantially smaller than this. We therefore expect cooling to play an important role in the evolution of the system. For simplicity, however, we first treat a simplified adiabatic problem ignoring cooling. The gravi-

tational softening is taken to be 2 and 5 kpc for the gas and dark matter particles, respectively. The minimum time-step was typically $3 \times 10^5$ yr. The gas has a total mass of $10^{11}$ $M_\odot$ and is initially at a uniform temperature, $T = 10^3$ K, well below the equilibrium virial temperature of the system ($T_{vir} \sim 3 \times 10^6$ K).

Unless otherwise specified, we use units where

$G = 1$,

[mass] $= 10^{10}$ $M_\odot$,

[time] $= 4.71 \times 10^6$ yr,

[velocity] $= 207.4$ km s$^{-1}$,

[distance] $= 1$ kpc,

[density] $= 6.77 \times 10^{-22}$ g cm$^{-3}$,

[energy/mass] $= 4.3 \times 10^{14}$ erg g$^{-1}$.

As shown by Evrard (1988, 1990) and by the similarity solutions discussed above, if cooling effects are negligible the gas evolution in a quasi-spherical collapse is dominated by an outwardly moving shock which efficiently transforms most of the radial kinetic energy into heat. After collapse, the gas settles to a spatial distribution that closely resembles that of the underlying dark matter. This is also a good description of what we see in the present simulations. Fig. 2 shows the

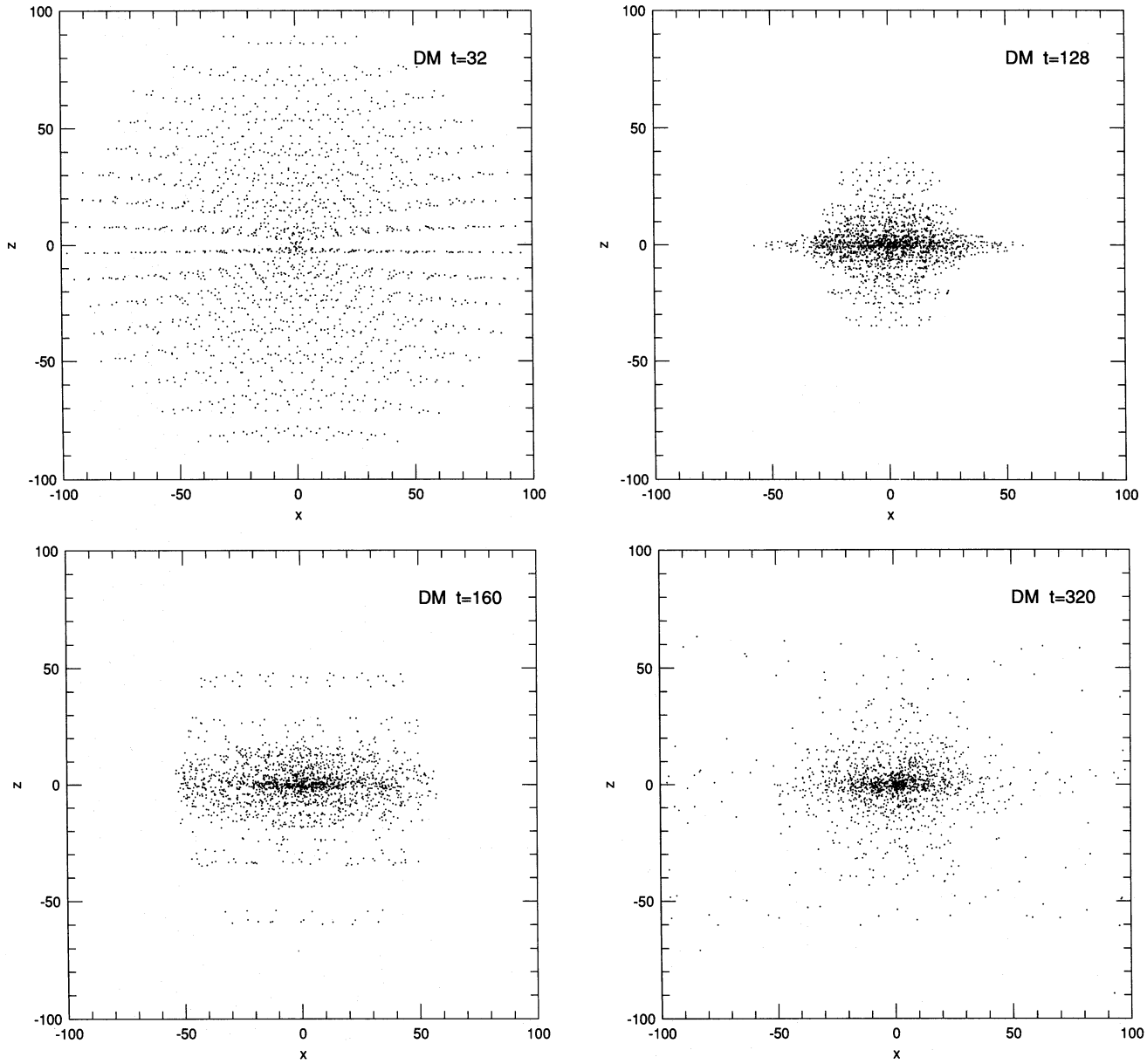


**Figure 2.** 'Adiabatic' collapse of a rotating clump of gas and dark matter. The left and right panels show the gas and dark matter components, respectively. Only the $x$–$z$ projection is shown at four different times, indicated on the figures. The frames are 200 units across. The system collapses at $t \sim 100$. By the end of the simulation, both the gas and dark components are in equilibrium. The gas is more nearly spherical than the dark matter, an effect due to the fact that equidensity contours of the gas must coincide with equipotentials which are considerably rounder than contours of equal total density.

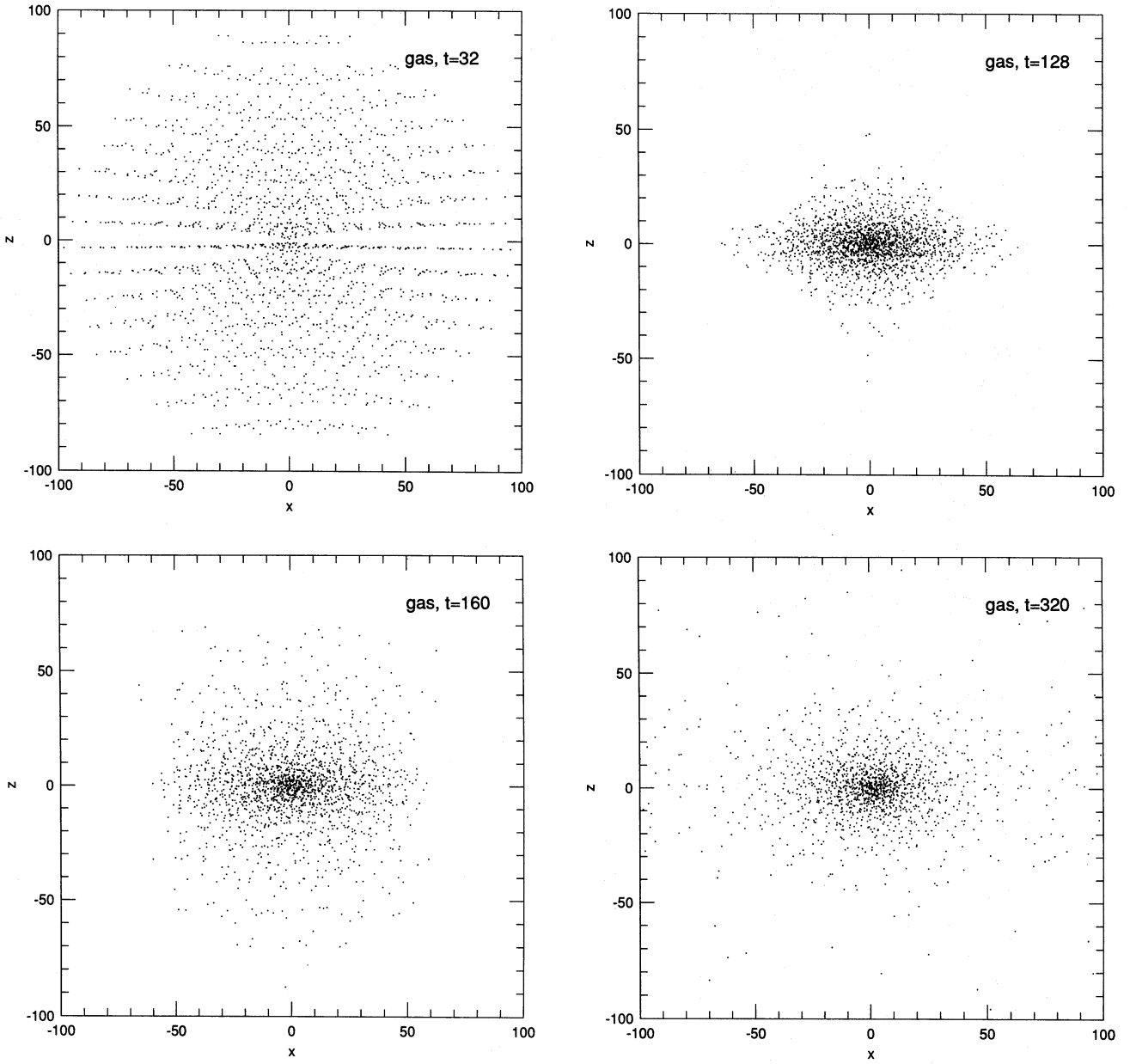


**Figure 2** – *continued*

evolution of the gas and dark matter distributions in the largest simulation ($N_{\rm gas} = N_{\rm DM} = 1796$). The two components develop similar structures, the gas being slightly more spherical in the final state. This difference is probably due to the fact that gas must be stratified on equipotential surfaces in an equilibrium system, and these are always rounder than equidensity surfaces. A similar difference was noted by Evrard (1990) in his own simulations.

Fig. 3 shows the final projected radial density profiles of the systems. In this plot and all the following ones, the profiles corresponding to runs realised with 90, 368 and 1796 particles will be shown with open triangles, open squares and filled triangles, respectively. Figs 2 and 3 confirm our expectation that the gas and dark matter reach similar equilibrium configurations, except, perhaps, that the gas appears somewhat less concentrated, in particular for the case with $N_{\rm gas} = 90$. Fits of the form

$$\rho(r) = \rho_0 / [1 + (r/r_{\rm core})^2]^{3\beta/2} \qquad (14)$$

are shown as dashed lines for the system with $N_{\rm gas} = 1796$. Core radii, central densities, and values of $\beta$ corresponding to the best fits can be found in Table 1 for every profile. The data in this table confirm the visual impression that, although all the dark matter profiles are remarkably similar, the gas profile differs substantially when only 90 gas particles are used. In particular, the $\beta$ value is substantially smaller than in the other two cases, indicating that the gas is more extended than the dark matter. A more sensitive indicator of this difference is a comparison of the half-mass radii of the two components. The evolution of these radii is shown in Fig. 4.

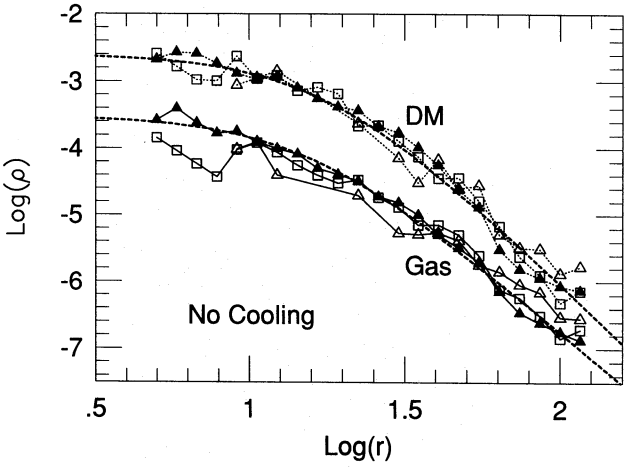


**Figure 3.** Final projected density profiles for the gas and dark matter in an adiabatic collapse model. Different symbols indicate different numbers of particles: open triangles, open squares and filled triangles correspond to 90, 368 and 1796 particles of each type. Note that the gas and the dark matter have similar spatial distributions, except, perhaps, that the gas is slightly less concentrated. The $N_{gas}=90$ case is the only clearly discrepant one. Fits to the filled-triangle profiles using equation (12) are shown as dashed lines. Parameters of similar fits to all profiles are given in Table 1.

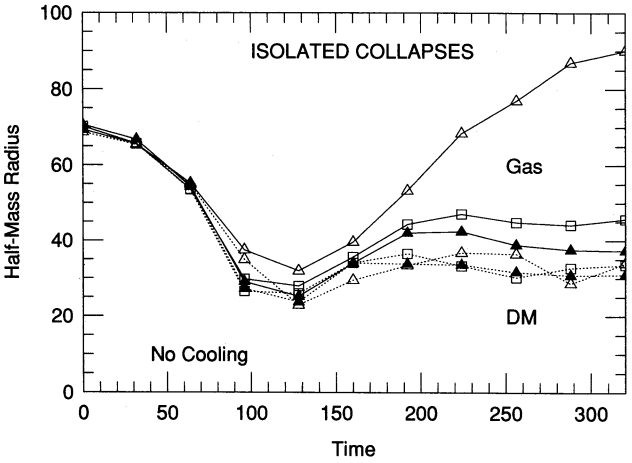


**Figure 4.** The evolution of the half-mass radii of the gas and the dark matter in an adiabatic collapse. Solid and dotted lines correspond to the gas and dark matter, respectively. The symbols indicate the different numbers of particles used, as in Fig. 3. Gas and dark matter evolve quite similarly, except in the $N_{gas}=90$ test, which is steadily expanding by the end of the simulation. This expansion is due to the enhanced effects of viscosity in small-$N$ systems, which cause a transfer of energy between the two components at collapse.

**Table 1.** Fit parameters for isolated collapse and multiple merger models.

| $N_{gas}$ ($=N_{DM}$) | $\beta$ ($DM$) | $\beta$ (gas) | $\rho_0$ (DM) | $\rho_0$ (gas) | $r_{core}$ (DM) | $r_{core}$ (gas) | $r_h$ (DM) | $r_h$ (gas) |
|---|---|---|---|---|---|---|---|---|
| 1796 | 1.50 | 1.30 | 2.5e-3 | 3.0e-4 | 17.5 | 15.0 | 31.0 | 37.3 |
| 368 | 1.55 | 1.25 | 2.5e-3 | 1.5e-4 | 18.0 | 17.0 | 33.6 | 45.7 |
| 90 | 1.35 | 0.95 | 2.0e-3 | 1.5e-4 | 17.0 | 13.0 | 34.0 | 90.0 |
| 6 × 90 | 1.25 | 1.15 | 2.5e-3 | 8.0e-5 | 27.0 | 35.0 | 80.7 | 111.5 |
| 6 × 368 | 1.30 | 1.27 | 2.3e-3 | 2.5e-4 | 28.0 | 28.0 | 85.1 | 94.3 |
| 6 × 6 × 29 | 1.10 | 1.00 | 1.7e-3 | 1.2e-5 | 45.0 | 90.0 | 246.3 | 579.3 |
| 6 × 6 × 90 | 1.10 | 1.10 | 1.9e-3 | 6.0e-5 | 46.0 | 65.0 | 238.8 | 407.2 |

In the $N_{gas}=90$ case the gas rebounds substantially more than the dark matter after collapse.

Since our code conserves energy to a high accuracy (better than 1 per cent), the expansion seen in Fig. 4 must be driven by energy that the gas gains at the expense of the dark matter. What drives this energy transfer and why does it appear less important when more particles are used? Fig. 5 shows the evolution of the thermal (Th), potential (W) and 'total' energies of the gas (solid lines) and contrasts them with the kinetic (K), potential and 'total' energies of the dark matter (dotted lines). Filled circles and open triangles indicate the amount of gas kinetic energy invested in motion about the rotation axis ($K_{rot}$) and in radial motions ($K_{rad}$). $K_{rot}$ is defined as

$$K_{rot}=\frac{1}{2}\sum_{i=1}^{N} m_i \frac{(\boldsymbol{J}_i\cdot\hat{\boldsymbol{J}}_{tot})^2}{[r_i^2-(\boldsymbol{r}_i\cdot\hat{\boldsymbol{J}}_{tot})^2]}, \qquad (15)$$

where $\boldsymbol{J}_i$ is the specific angular momentum of particle $i$, and $\hat{\boldsymbol{J}}_{tot}$ is a unit vector in the direction of the total angular momentum of the system. The sum in equation (15) is actually carried out only over those particles that satisfy the condition $(\boldsymbol{J}_i\cdot\hat{\boldsymbol{J}}_{tot})>0$, in order to exclude counterrotating particles.

'Total' energies refer to the sum of thermal, kinetic and potential energies for each component. The dark matter's thermal energy is, of course, zero. A constant value of the total energy for each component would mean that no energy is being transferred between components. This is the case when 1796 particles are used, but a sudden jump is clearly noticeable in the total energy of the gas at the time of collapse when $N_{gas}=90$. The transfer ends right after the collapse, and there is no appreciable further exchange. This process can be interpreted as follows. When $N_{gas}$ is small, a gas particle will have neighbours located far away and so probing a rather different local velocity field. The artificial viscosity terms tend to damp the relative motion of neighbours and become important earlier in the collapse when a smaller number of particles is used. As a result the gas collapse is spuriously retarded, allowing the dark matter to overtake the gas and exert a larger inward acceleration on it. This 'phase lag' between the collapse of the two components provides a mechanism for transferring energy between them. As shown in Fig. 5, the dark matter ends up more tightly bound, although the relative change in energy is smaller because the total mass of dark matter is 9 times greater. Although the process described here operates to some extent in all three cases, its effects are already small for $N_{gas}=368$ (see Table 1 and Figs 4 and 5), indicating that with more than ~300 particles the spherical collapse of a mixture of gas and dark matter can be followed rather accurately.

It is important to prove that 'phase' differences during collapse can produce an energy transfer of the kind seen here, and to exclude differences in particle mass or softening, or the presence of rotation, as possible sources for the effect. To this end, we ran a 'toy' model consisting only of collisionless particles. The initial system is similar to the models used

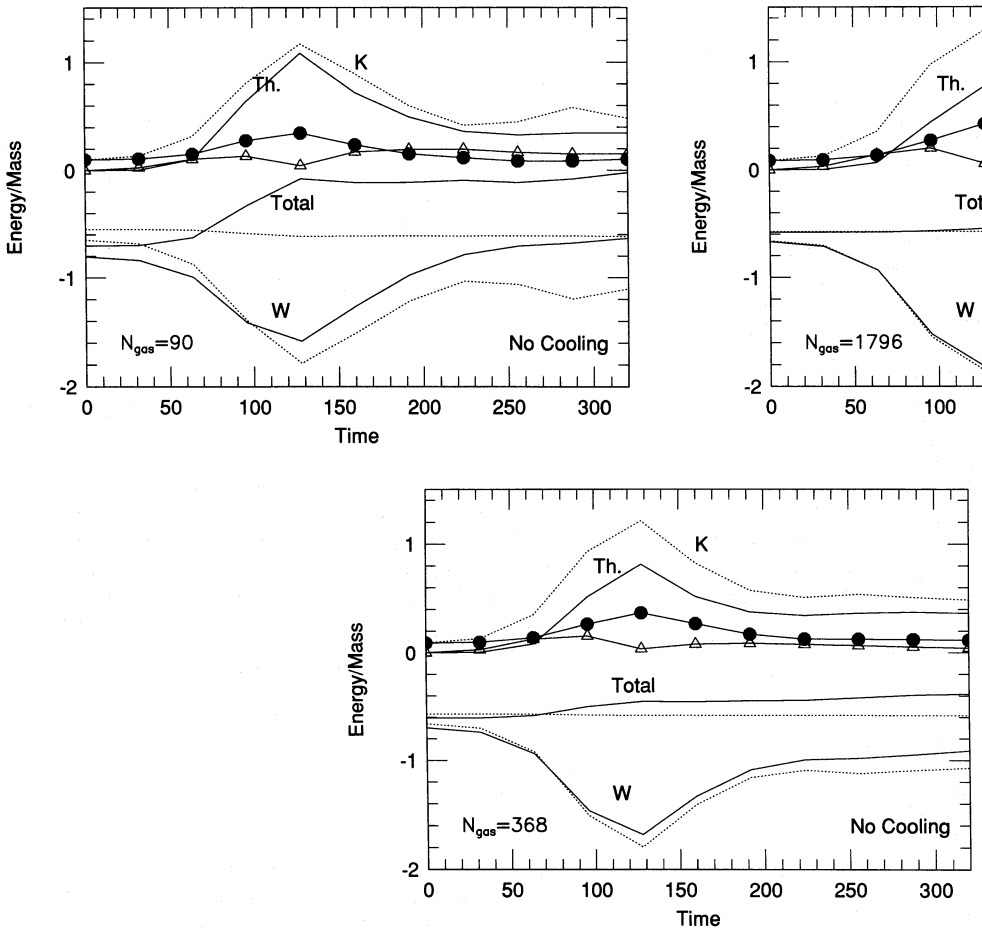


**Figure 5.** The time evolution of various specific energies in adiabatic collapses. Dotted lines refer to dark matter energies and solid lines to gas energies. K labels the total kinetic energy of the dark matter, Th. the total thermal energy of the gas, filled circles and open triangles the gas kinetic energy in rotation and in bulk radial flows, respectively, W the gravitational potential energies, and Total the sum of all kinetic, internal and gravitational energies for each component. The three panels refer to similar models realized with different numbers of particles. Note the significant differences in the evolution of the two components in the case with $N_{\rm gas} = 90$.

above, but its gas particles are replaced by collisionless particles. Their properties are altered so that every particle now has the same mass and gravitational softening. The velocities are such that one component is initially at rest (the 'rest' component), and the other is endowed with inward radial velocities of magnitude proportional to radius (the 'accelerated' component). The kinetic energy associated with these motions is negligible and, therefore, both components initially have almost the same total energy. However, the accelerated component collapses slightly faster in an analogous manner to the dark matter in the $N_{\rm gas} = 90$ simulation. This 'toy' system has total mass $M_{\rm tot} = 2$, initial radius $R_{\rm tot} = 1$, and collapse time around unity. Three different values for the initial velocities of the 'accelerated' component were tried, $\alpha = 0.1$, 0.2 and 0.4 in the relation $\boldsymbol{v} = -\alpha \boldsymbol{r}$. The evolution of the total energy of each component is shown in Fig. 6. Energy is clearly transferred between components; the transfer is now symmetric in magnitude because the components have the same mass, and it ceases once the collapse is complete. The net result is that the 'accelerated' component ends up more tightly bound than the 'rest' component, supporting the interpretation of the energy

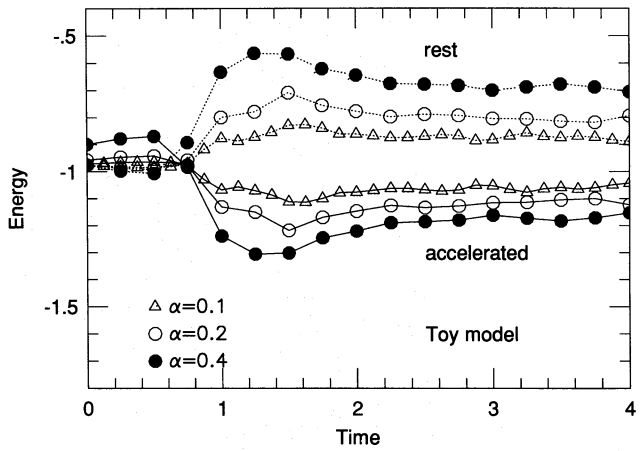


**Figure 6.** Energy evolution of the two components for the 'toy' model described in the text. The 'rest' component lags behind the 'accelerated' one during the collapse. This timing difference allows energy to be transferred between components. The rest system always gains energy at the expense of the accelerated component. This transfer becomes more efficient as the timing difference increases. See the text for further details.

transfer process described above. The mechanism seems to depend on the magnitude of the 'phase difference', since systems with larger $\alpha$ give rise to larger energy transfers.

A further point to note about our three protogalaxy collapse simulations is that the final thermal energy of the gas is very similar in all three cases (i.e. the average gas temperature is the same). This indicates that the extra energy gained by the gas is used primarily to drive the rather slow expansion seen in Fig. 4, and does not produce much change in gas temperature. The larger $K_{\rm rad}$ and smaller $K_{\rm rot}$ in the $N_{\rm gas}=90$ model can also be traced to this expansion (a larger radius at fixed total angular momentum necessarily requires a reduced $K_{\rm rot}$). Thus the effects of small $N$ are mainly apparent as a change in the spatial distribution of the gas relative to the dark matter.

To summarize, in the absence of radiative cooling the gas and dark matter settle to very similar configurations after a quasi-spherical collapse. The gas is shock-heated to the virial temperature of the system, and very quickly reaches approximate hydrostatic equilibrium. Artificial viscosity can induce spurious systematic effects in small-$N$ systems. These tend to reduce the similarity between the gas and dark matter distributions, especially in the central regions, making the gas less concentrated than the dark matter.

## 4 ADIABATIC MERGERS

Although the results of the previous section show that in a spherically symmetric collapse no substantial difference arises between the spatial distributions of gas and dark matter, this conclusion cannot simply be extrapolated to the formation of clumps in a hierarchically clustering universe. Numerical experiments show that in such a universe the formation of virialized clumps is far from spherical, and is characterized by mergers and by the accretion of previously collapsed structures. Our results already indicate that energy is readily transferred between components when 'phase differences' are present during collapse. In the spherical collapses discussed above, only numerical artificialities gave rise to such phase differences. However, it is easy to imagine less symmetric situations where these differences might be quite common. For example, in a head-on merger of two similar clumps we expect the dark matter to pass through freely at closest approach, whereas the gas will be stopped by a shock. Depending on how the shock propagates, 'phase differences' can easily arise between the evolution of the gas and that of the dark matter. The outcome of such encounters is best studied with numerical experiments, and constitutes a first step towards understanding the more complicated process of the hierarchical formation of structure. We begin by studying mergers between pairs of similar objects, and then proceed to more complicated models of hierarchical merging.

### 4.1 Binary mergers

Merging encounters between clumps of similar size are perhaps the most extreme dynamical evolution that virialized clumps can undergo. As described above, significant energy transfer may occur between gas and dark matter during this process. Motivated by this expectation, we set up initial conditions representing nearly head-on encounters between the models described in Section 3. The total energy per unit mass of each individual model is $E_{\rm tot}/M_{\rm tot}=-0.67$. We tried encounters with specific orbital energies in the range $(-0.9, +0.46)$. The initial separations were chosen so that each object can collapse and virialize before the collision. Unless otherwise specified, the initial internal spins can be assumed to be parallel to each other and to the (small) orbital angular momentum.

Pictures of the gas and dark matter distributions projected on to the orbital plane are shown in Fig. 7 for a model with specific orbital energy $-0.8$. The systems are approaching along the $x$-axis. They collapse individually at $t\sim100$, and reach pericentre at $t\sim150$ with a core separation of about 20 units. The dark matter haloes pass through each other freely, but the $x$-motion of the gas is almost stopped by a shock. Since the internal spins are parallel, the gas components deviate from their original direction of motion, rolling slightly along the $y$-axis so that they can make their way past each other. Separate plots of the gas from each initial clump reveal no interpenetration along the shock front. Once past pericentre, the gas components quickly rejoin their dark matter haloes, and start their way back for a second collision. After this the systems soon merge. This process is shown more directly in Fig. 8, where we have plotted the path of the centres (i.e. the most tightly bound subsystems) of the dark matter and gas components of one of the two colliding objects. Symbols are plotted at time intervals $\Delta t=16$, starting at $t=112$. The last symbol plotted corresponds to $t=352$. The $y$-offset between the components, which is quite noticeable at closest approach, gives rise to an effective torque which transfers angular momentum between them. As a result, the specific angular momentum of the gas is 35 per cent larger than that of the dark matter in the final remnant.

The evolution of the total energies of the gas and of the dark matter is shown in Fig. 9 for mergers with initial specific energy $-0.8$. The two cases correspond to each initial object having either 368 (open squares) or 1796 (filled triangles) particles of each type. Clearly, the offset between components at pericentre leads to a sudden increase in the energy of the gas component with a corresponding drop in that of the dark matter. The energy jump is similar for the two cases, although the total energy of the gas in the simulation with $2\times368$ gas particles is a little higher, a difference that can be traced to spurious energy transfer during the collapse of the individual systems before the collision. If we assume that the run with $2\times1796$ gas particles gives the 'correct' result, we see that the gas component during a merger can increase its total energy by $\sim40$ per cent, implying that its half-mass radius will also be almost twice that of the dark matter.

It is important to understand what determines this apparently large factor. If the energy transfer mechanism is similar to the one described in Section 3, we expect the net transfer to depend on the magnitude of the 'phase differences', which in turn is likely to be related to the impact velocity. In fact, the energy change in the gas component, normalized to the specific energy of the individual pre-merger systems, turns out to be an increasing function of the encounter energy, as we show in Fig. 10. This figure also shows the energy changes for similar encounters where the individual spins are anti-parallel, or are zero. The changes

appear to be minimum when the spins are parallel, perhaps reflecting the fact that the gas components can roll past each other in this case with relative ease. If the pre-merger systems are not rotating, the gas components are effectively stopped at the centre in a frontal encounter, presumably leading to an increased 'phase lag' and a more efficient transfer of energy to the gas.

In summary, significant energy transfer is to be expected during mergers of clumps containing both gas and dark matter. In the absence of radiative losses, shocks tend to pump energy from the dark matter into the gas, which can end up significantly less bound as a result. The magnitude of this effect depends rather strongly on the encounter parameters. In our experiments it is smallest when the internal spins of the colliding clumps are parallel.

### 4.2 Hierarchical aggregation of clumps

The results of the previous sections raise a serious question regarding the reliability of cosmological simulations of hierarchical clustering. Although systems that contain several hundreds of particles, and are thus apparently free from the unwanted effects plaguing small-$N$ systems, will eventually build up, to do so they must pass through a stage

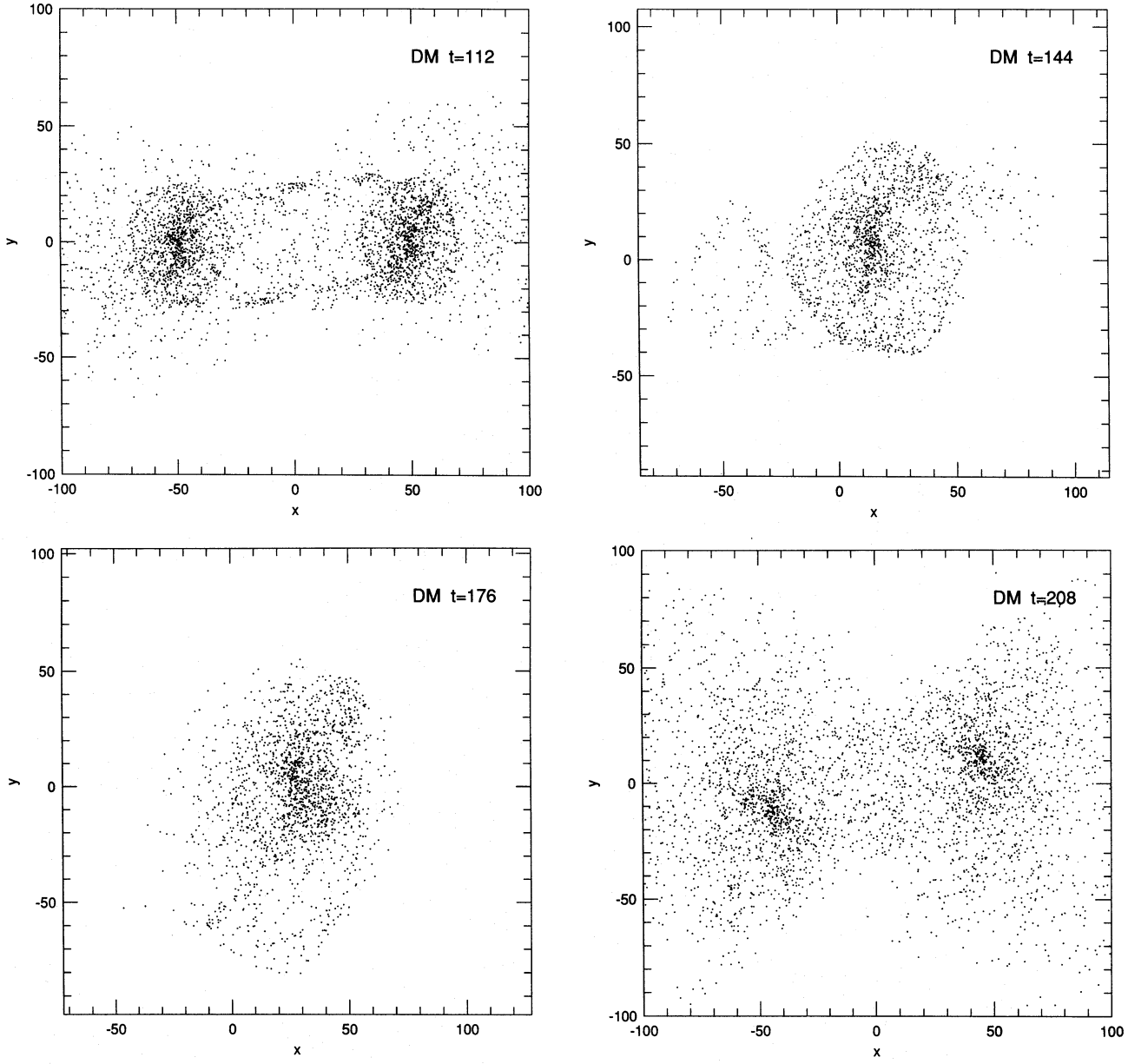


**Figure 7.** Gas and dark matter particle positions during an almost head-on merger. The system is viewed in projection on to the orbital plane. The individual systems collapse at $t \sim 100$ and reach the pericentres of their relative orbits at $t \sim 150$. They merge after the second approach, by $t \sim 500$. The frames corresponding to times 144 and 176 show only *one* of the colliding systems. The system moving towards $-x$ (from right to left) is shown at $t = 144$ while the other is pictured at $t = 176$. The shock front can be clearly seen in the gas, where no interpenetration occurs. The dark matter systems pass through each other freely. This difference induces a 'phase lag' between the gas and the dark matter which is responsible for the exchange of energy between components.

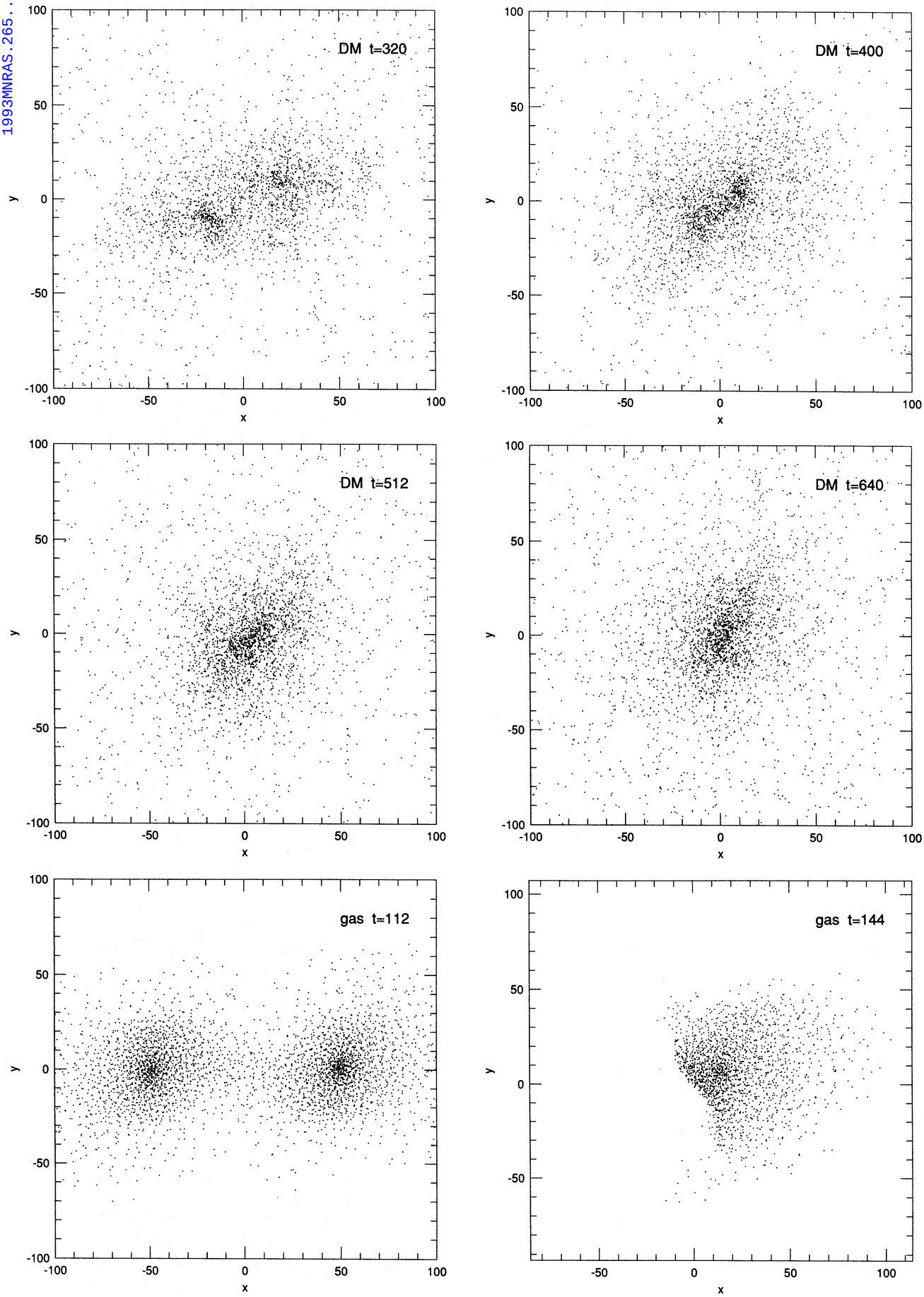


**Figure 7** – *continued*

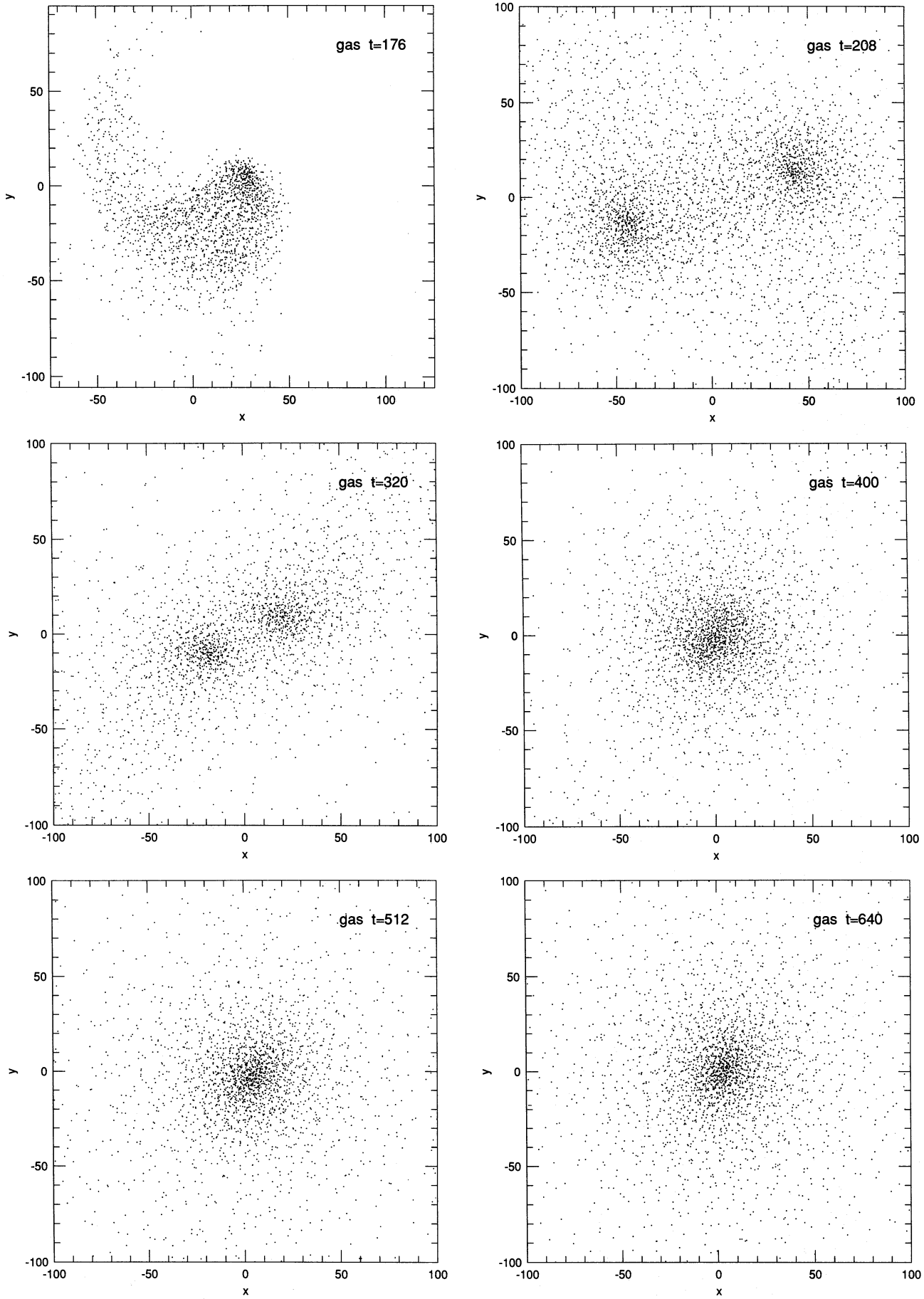


**Figure 7** – *continued*

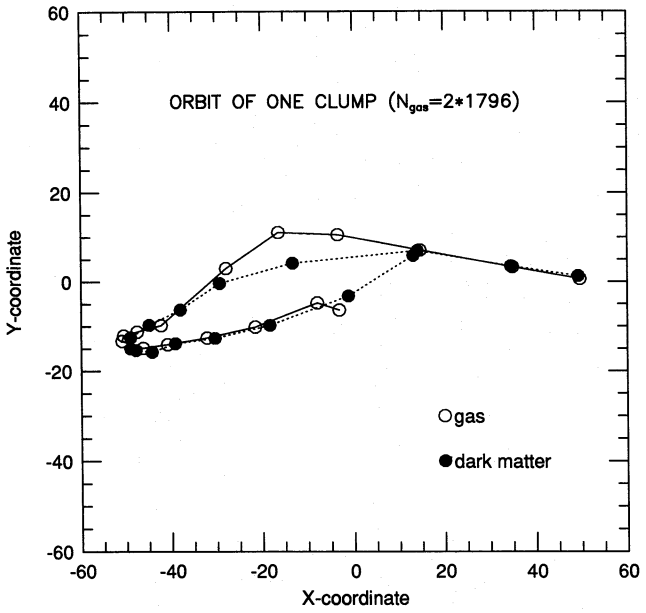


**Figure 8.** Location of the centres of the gas and dark matter distributions of one of the 'galaxies' during the merger simulation shown in Fig. 7. Symbols are plotted at equal time intervals, $\Delta t = 16$, starting at $t = 112$. The time sequence is from right to left and counterclockwise. The spatial offset between components at the pericentre is quite clear. This offset results in energy and angular momentum being transferred from the dark matter to the gas.

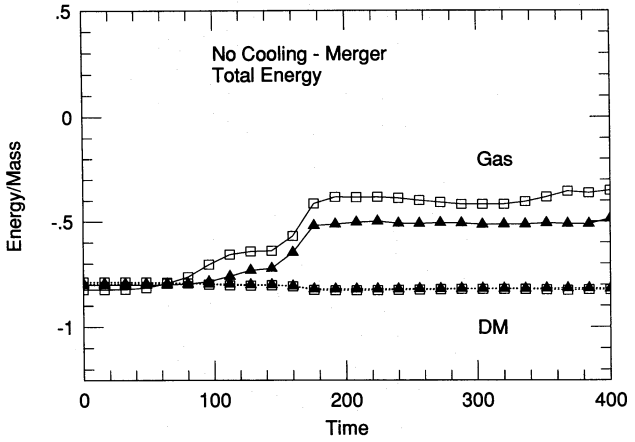


**Figure 9.** Energy evolution in the merger simulation shown in Fig. 7 (filled triangles). Total numbers of gas and dark matter particles are $N_{gas} = N_{DM} = 2 \times 1796$. Open squares correspond to an equivalent experiment carried out with $N_{gas} = N_{DM} = 2 \times 368$.

in which most of the mass is in clumps of no more than a few tens of particles. This is especially true for simulations with initial conditions for which small-scale power in the perturbation spectrum is strong. How will errors in the early evolution influence the final results of a simulation?

In order to investigate this issue we simulate the formation of a system through a series of mergers of small 'building blocks'. As before, the results of equivalent realizations with different numbers of particles are compared in order to search for systematic errors in the small-$N$ simulations. In

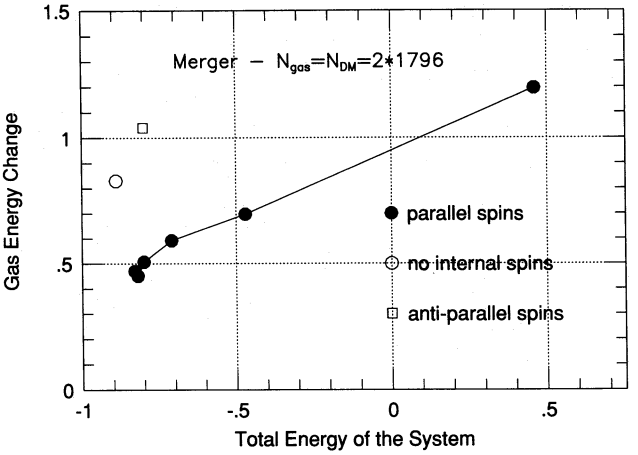


**Figure 10.** Energy gained by the gas in different merger simulations as a function of the orbital energy of the encounter. Energy changes are normalized to the specific energy of an individual pre-merger system. Filled circles refer to colliding clumps which are initially rotating around parallel axes. Empty squares indicate the case in which their spin vectors point in opposite directions. The orbital angular momentum is very small in all cases. There is a tendency for more energetic encounters to result in larger amounts of energy being transferred between components. At a fixed orbital energy, a frontal encounter between non-rotating systems seems to induce the largest energy exchange.

most hierarchical clustering scenarios, as virialized systems form through the amalgamation of smaller units the typical object evolves towards increasingly deeper potential wells and lower mean densities. Our simulations try to mimic this process in the following way. We use as 'building blocks' the same models as described in Section 4. These will constitute the first level of the hierarchy. A second level is constructed by distributing six such models at the corners of an octahedron with spacing such that the collapse time of the system as a whole is twice that of each individual block. This ensures that the blocks collapse before merging, and results in a virial temperature for the final system that is about twice that of the individual blocks. This is the expected scaling between mass and collapse time for a hierarchy that develops from a power-law spectrum of initial density perturbations ($|\delta_k|^2 \propto k^n$) with index $n \approx -1.5$, approximately the value expected on galactic scales in a cold dark matter (CDM) universe. The directions of the internal spins are chosen so that they will cancel out in pairs. A small amount of coherent solid-body rotation is added to the system as a whole in order to give a spin parameter $\lambda = 0.05$. This smaller value of $\lambda$ is similar to that found in cosmological calculations of hierarchical clustering. A third level of the hierarchy can be added in a similar way. CPU limitations determine how many particles we can use in each of these simulations. 'Building blocks' with 90 and 368 gas particles were used for the two-level hierarchies, while 29 and 90 gas particles were used in the fundamental units of the three-level hierarchies. These choices give totals of 540, 2208, 1044 and 3240 gas particles in the four cases. Particle plots illustrating the evolution of the larger three-level hierarchy are shown in Fig. 11.

As expected, the gas is shock-heated to approximately the virial temperature of each level of the hierarchy as it

collapses. The temperature and density profiles of the gas components in the final systems are shown in Fig. 12. There are clearly only rather small differences between simulations run with different numbers of gas particles, except near the centre, where discrepancies arise because of differing resolution. Arrows in the figure show the minimum value of the gas smoothing length at the centre of each final configuration. Outside these regions the density and temperature profiles agree very well. It seems then that, even if small-$N$ effects compromise the early evolution of these simulations, the spurious effects they cause have little influence on the structure of the more massive (and better resolved) levels of the hierarchy which collapse later.

The reason for this is clarified by studying the energy evolution of the gas component. The major deviation caused by small-$N$ effects is the spurious transfer of energy to the gas during the collapse of the initial clumps. The perturbations caused by this effect are, however, small compared to the energy jumps that occur later as larger clumps merge and virialize. It thus seems unlikely that they will much affect the final equilibrium structure of the gas. As shown in the last section, the amplitude of the energy transfer between gas and dark matter in a merger depends mainly on the energy of the encounter. Fig. 13 shows the evolution of various energies for the gas and for the dark matter in the two-level and three-level hierarchy simulations. The early energy transfer corre-

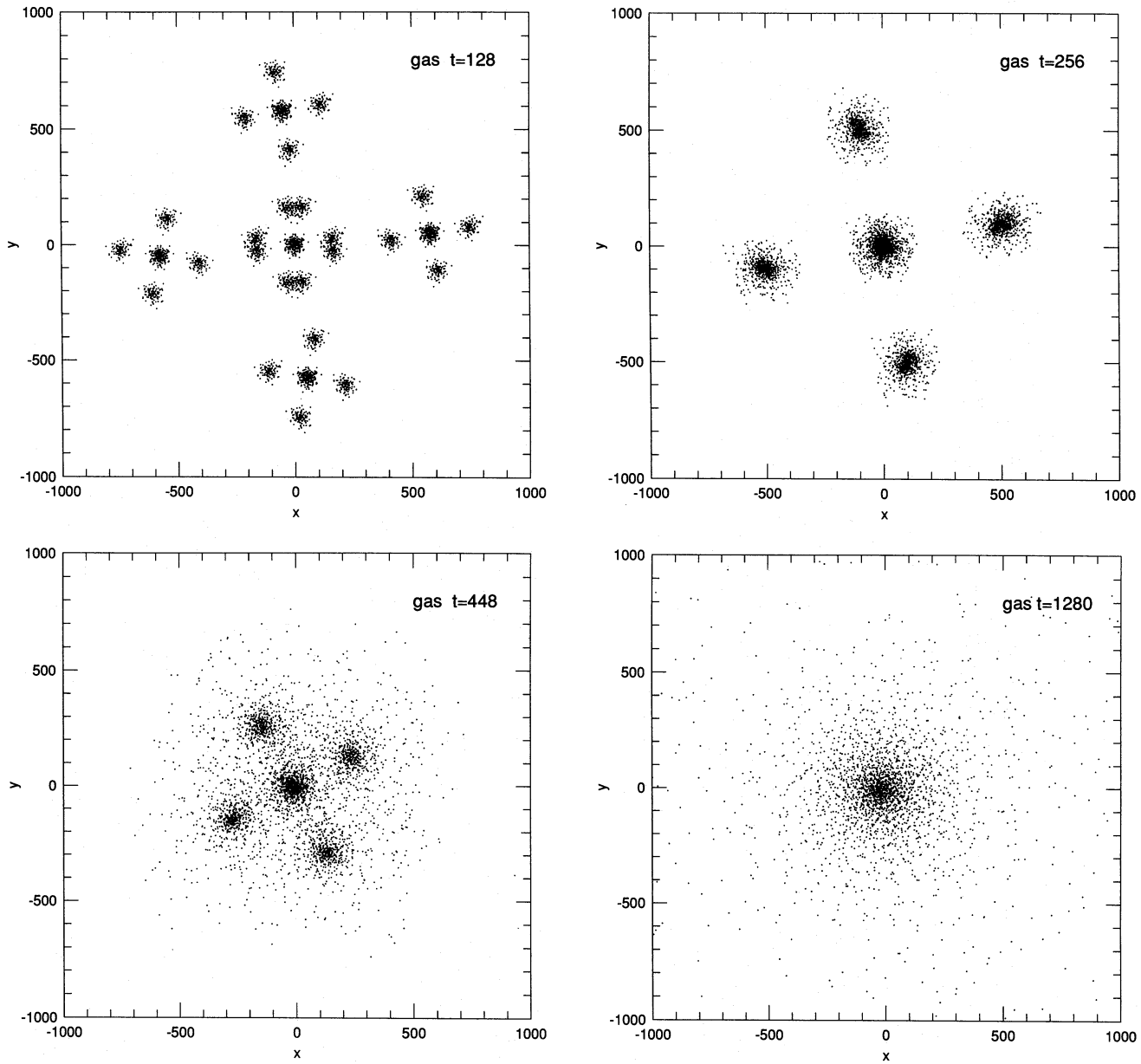


**Figure 11.** Different stages of the evolution of a three-level hierarchy. Left and right panels show the gas and dark matter components, respectively. The first level of the hierarchy consists of 36 identical clumps with 90 gas particles each (the 'building blocks'). These collapse individually at $t \sim 125$. The second level collapses at $t \sim 250$, and the final collapse of the system occurs at $t \sim 500$. The spatial distributions of gas and dark matter are quite similar at every stage of the calculation.

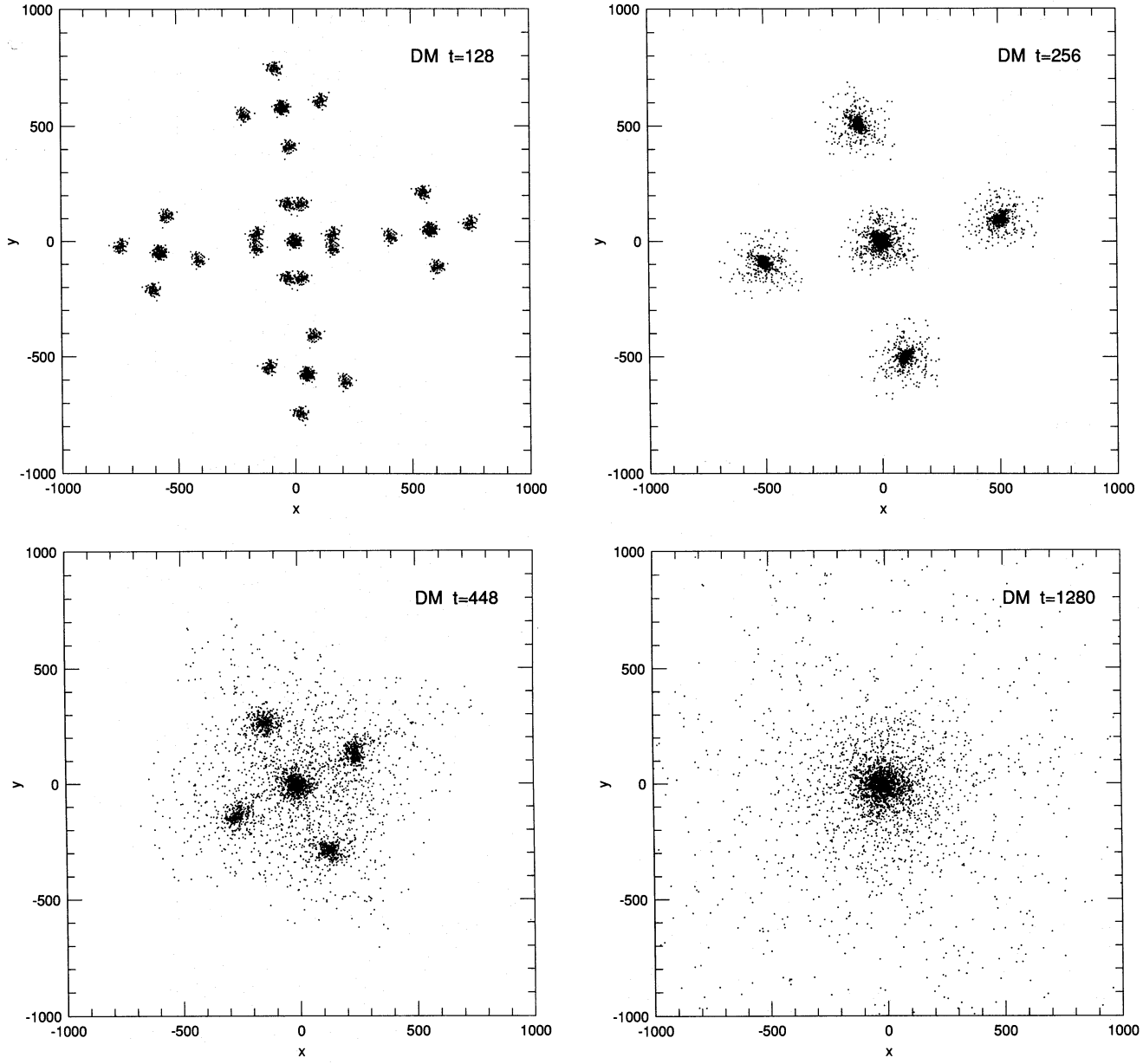


**Figure 11** – *continued*

sponding to the collapse of the poorly resolved first level of the hierarchy is almost completely washed out by the subsequent energy transfers that occur during mergers. As in the binary merger models of the last section the final thermal energies are very similar in simulations with different numbers of particles, even though the total gas energies can differ significantly, by almost 30 per cent for the case where the first level of the hierarchy was realized with only 29 gas particles.

Notice that the total transfer of energy between components is quite large in these models. This is perhaps best quantified by noting the different half-mass radii of the gas and dark matter distributions. These are given in Table 1, which also shows parameters for fits of equation (12) to the final density profiles. There are no significant differences in the parameters for the dark matter distribution as a function of the particle number used. For the gas, however, the core radius is substantially larger and the power-law index, $\beta$, is significantly shallower for runs with smaller $N_{\rm gas}$. Thus poor resolution seems to lead both to an over-diffuse core and to an over-extended envelope. This is a serious problem if simulations of this type are used in order to predict quantities, such as the X-ray luminosity of a galaxy cluster, that are crucially dependent on the central properties. However, in the simulations of the present section the gas core radius estimates are not much larger than the resolution of the simulations, and should be regarded as suspect anyway.

In summary, multiple mergers, like binary mergers, tend to transfer energy from the dark matter to the gas. If mergers cause the energy of the gas to grow relatively quickly as hierarchical clustering progresses, numerical artefacts intro-

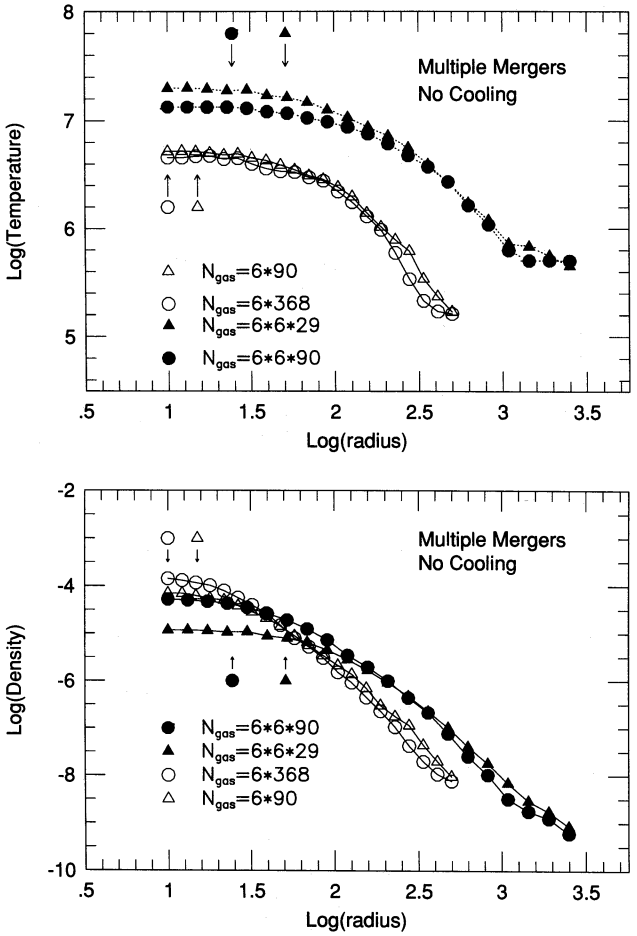


**Figure 12.** Equilibrium temperature (upper panel) and spatial density (lower panel) profiles of the gas in the final objects formed in the two- and three-level hierarchies. The total number of gas particles used in each run is shown in the figure. Estimated values in the central regions are affected by the limited resolution of the calculations. The radius corresponding to the minimum gas smoothing length is shown by the arrows. Outside the softened regions the profiles agree well with each other, irrespective of the number of gas particles used for the first level of the hierarchy.

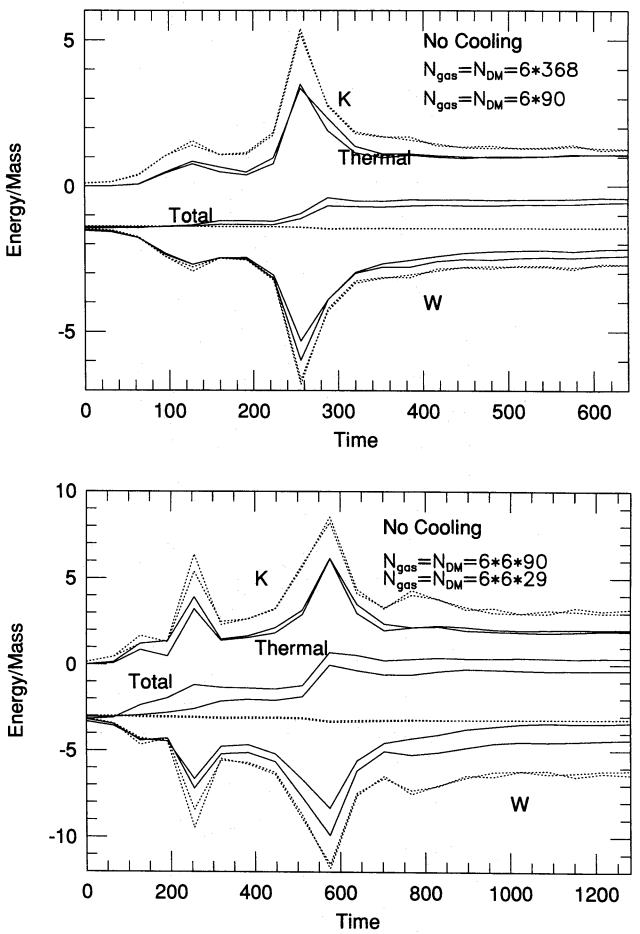


**Figure 13.** Energy evolution during the two- and three-level hierarchy simulations. The gas seems always to end up less tightly bound than the dark matter after major merger events. The amount of energy transferred is such that it washes out the spurious energy exchanges that occur when small-$N$ systems collapse (at $t \sim 100$). By the end of the runs, the differences in energy between systems built up from different small-$N$ clumps are relatively minor.

duced at early times may not greatly affect the equilibrium structure of the final product. Such 'large' energy increases are expected during the formation of galaxies and galaxy clusters in a CDM-like universe. The values of thermodynamic variables near the centre of a clump are quite sensitive to residual spurious effects from the early evolution of small-$N$ clumps, and they should therefore be treated with particular caution.

## 5 SIMULATIONS INCLUDING COOLING

As discussed above, the effects of radiative cooling may be very important in systems of the kind we are studying. Since the efficiency of gas cooling increases with density, $\tau_{\rm cool}$ can become much smaller than the dynamical time-scale in a protogalactic collapse, so that any thermal energy gained through shocks is quickly radiated away. Without pressure support, collapse of the gas can proceed unimpeded until it becomes centrifugally supported, presumably in a thin disc-like structure. Indeed, this has been suggested as the primary mechanism for forming galaxy discs in dark haloes (Fall & Efstathiou 1980; White 1991).

We can envisage two different ways in which this process might occur. During the collapse, shocks may drive the gas to a quasi-hydrostatic equilibrium configuration, much as if cooling effects were negligible (Section 3.2). The formation of a gaseous disc would then largely result from the subsequent flow of cooling gas towards the centre (a 'cooling flow' model). On the other hand, the gas may radiate energy so fast that it never gets heated to the virial temperature. The disc would then form almost completely at the collapse time, and might evolve little thereafter.

Which of these alternatives is at work in real dark haloes is an important question because the alternatives are likely to predict different disc structures and evolutionary paths, leading to rather different theories of galaxy formation. In the 'cooling flow' model we might expect larger amounts of gas to remain in the hot phase at the present day, and therefore higher present X-ray luminosities (e.g. White & Frenk 1991).

The same model predicts a significant delay between the collapse of the halo and the assembly of the bulk of the disc. This delay is not expected if the second alternative holds.

For the parameters we adopt in our standard protogalactic collapses the 'cooling flow' alternative does not provide a good description of what happens. Fig. 14 shows the evolution of the gas component when radiative cooling is switched on in one of the collapses of Section 3.2. A thin gaseous structure forms at the centre of the dark halo immediately after collapse ($t \sim 120$) and evolves little thereafter, apart from a few particles near the edge of the disc which are lost by the end of the simulation. The thermal energy of the gas never plays an important role in the dynamical evolution since the average temperature stays well below the virial temperature of the system (near $T \sim 10^4$ K). As before, shocks are able to dissipate the energy in radial motions very effectively, and soon after the collapse almost all of the remaining kinetic energy is invested in the rotational motions that support the disc. The efficiency of this process improves considerably with increasing $N_{\rm gas}$. By the end of the simulation, only 12 and 6 per cent of the kinetic energy is in radial motions for $N_{\rm gas} = 90$ and 1796, respectively.

Throughout the main body of the cold gaseous discs that form in these experiments, Toomre's (1964) stability parameter, $Q$, is less than 1. As a result, the discs break up into clumps which orbit around a slowly rotating central blob (Fig. 15). To test the sensitivity of this behaviour to our gas parameters we tried a lower value for the gas mass in our

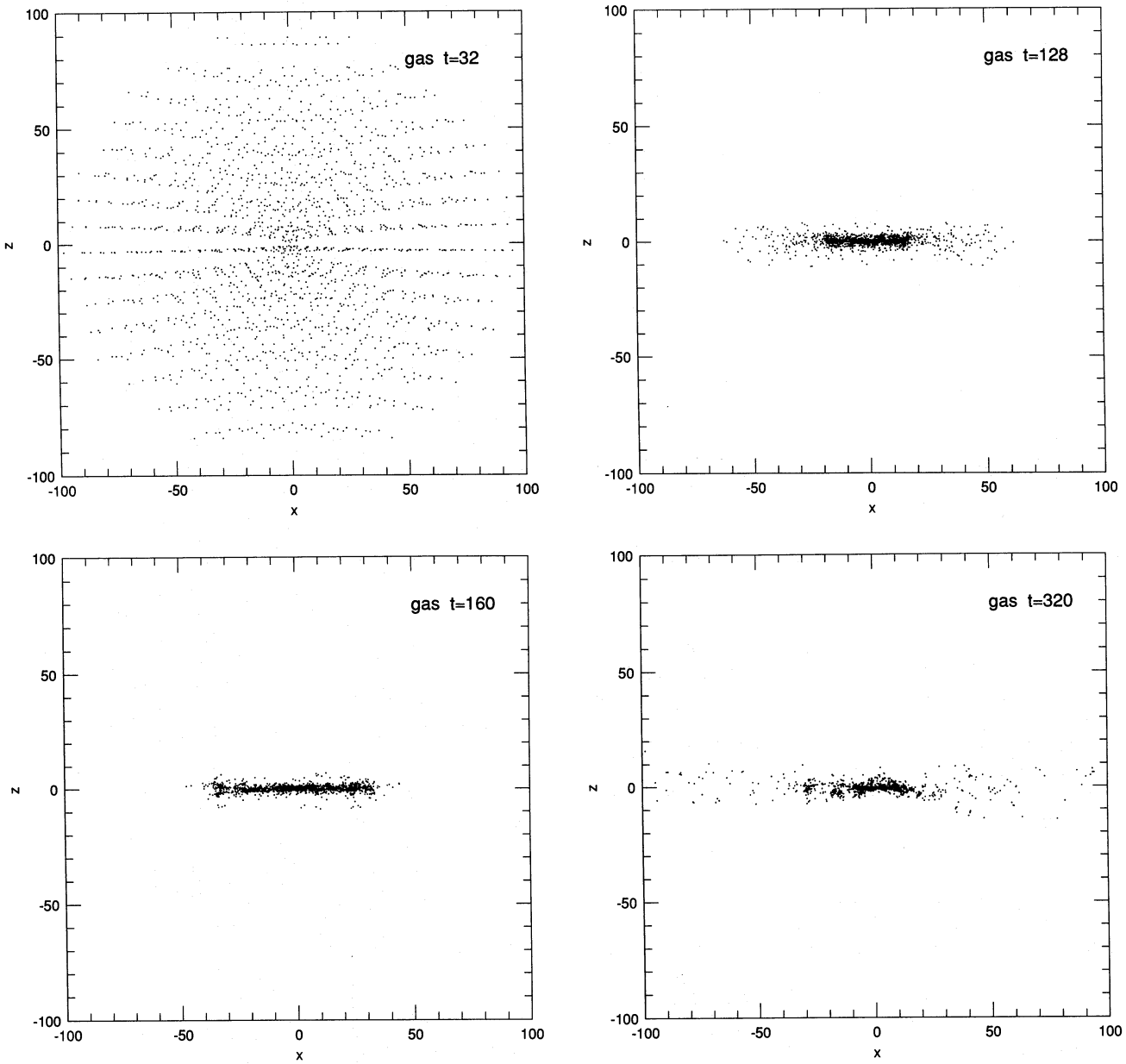


**Figure 14.** Different stages of the evolution of the gas component when radiative cooling is present. A disc forms immediately after collapse, and its $z$-structure remains almost unaltered thereafter. The times shown are the same as those in Fig. 2. Only gas particles are shown because the dark matter evolution is quite similar to the adiabatic case.

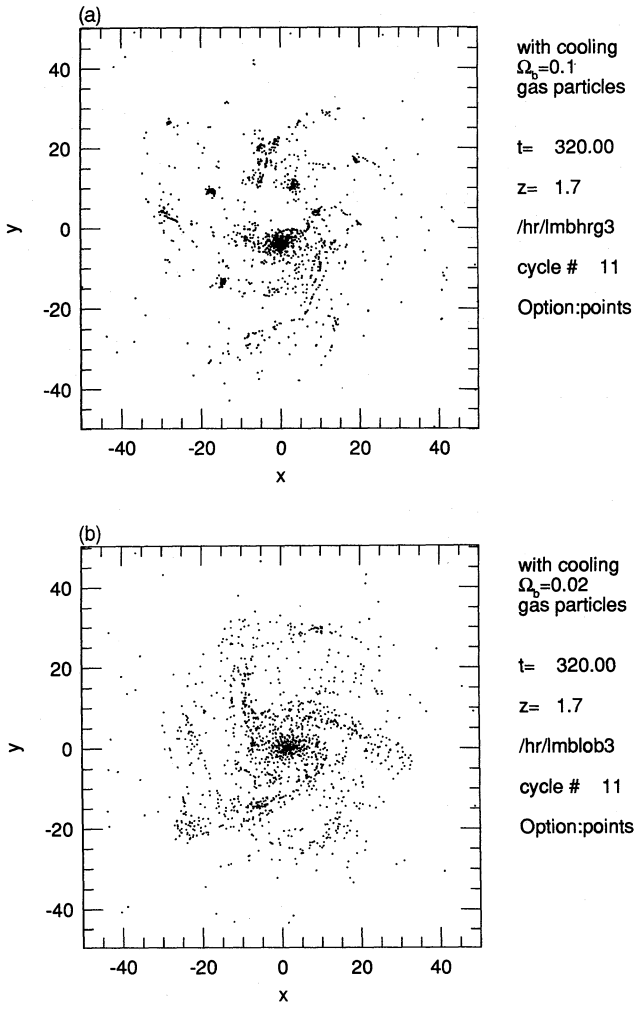


**Figure 15.** Face-on view of the gaseous disc formed in the tests with cooling. When the disc is Toomre-unstable ($Q<1$, $f_b=0.1$), it rapidly breaks up into clumps (a). When the gas density is lowered enough to make the disc stable ($Q>1$, $f_b=0.02$), the prominent clumps disappear and the disc shows a nice spiral structure (b).

collapse models, $2\times10^{10}$ $M_\odot$, giving a gas fraction $f_b=0.02$. Although this choice results in longer cooling times because the gas density is typically 5 times smaller, the change is insufficient to produce qualitative changes in behaviour during the collapse. However, the instabilities in the disc are now much weaker since $Q$ is larger than 1 everywhere. Fig. 15 compares face-on views of the discs in simulations with $f_b=0.1$ and 0.02. The prominent clumps that mar the $f_b=0.1$ case are not present when $f_b=0.02$; the latter model shows a nice spiral structure.

Viscous effects must, of course, be present in these discs, and will tend to transform kinetic energy into heat. Such additional thermal energy is radiated away very efficiently, and so does little to heat the discs. This energy loss is, however, artificial, and is expected to be substantially larger in small-$N$ systems where numerical viscosity is more effective. This is indeed seen in the energy-loss rates plotted in Fig. 16. From equations (3) and (4) it is clear that viscous effects scale as $h^2$ or, approximately, as $N_{\rm gas}^{2/3}$. We therefore expect the energy-loss rate of a 90-particle disc to be roughly

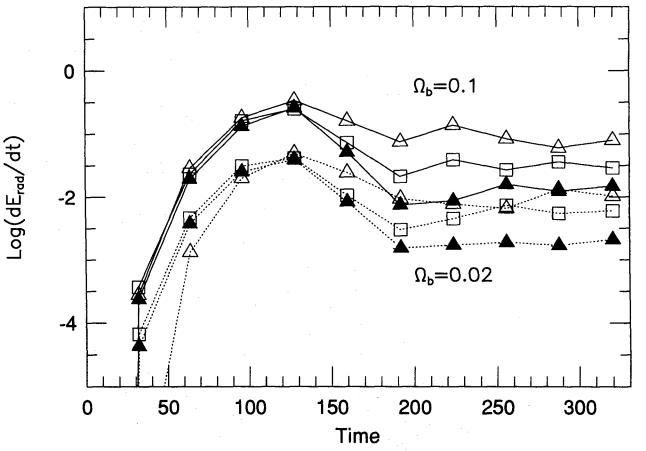


**Figure 16.** The evolution of the cooling rates of disc systems formed with different numbers of particles and with different $f_b$. Discs realized with small $N$ radiate energy at higher rates due to the increased effects of artificial viscosity. Although the disc structure is not affected because the additional heat is radiated away quickly, such small-$N$ effects lead to systematic overestimates of the total radiative energy loss. Cooling rates are in units of $1.5\times10^{10}$ $L_\odot$ and time is in units of $4.71\times10^6$ yr. Symbols are as in Fig. 3.

7 times that of a similar system realized with 1796 particles, a factor that agrees well with the data shown in Fig. 16. Very similar behaviour is seen in models where the gas component is much less important in mass ($f_b=0.02$), suggesting that viscous losses are due mainly to the differential rotation of the disc and are not much affected by the local instabilities so evident in Fig. 15.

When local instability leads to the formation of clumps, significant angular momentum exchange may occur between the inner and outer parts of the disc. Such 'physical' angular momentum transport may need to be distinguished from the artificial transport induced by numerical viscosity. Since the latter depends strongly on the number of particles used, this point can be checked by examining the angular momentum transfer process in our different tests. As angular momentum is transported outwards, we expect mass to be drawn inwards. The gas half-mass radius should therefore decrease, and a good indication of the overall effect of the angular momentum transfer should be the evolution of the ratio of the half-mass radius, $R_M$, to the half-angular-momentum radius, $R_J$. (Following Lin & Pringle 1987, $R_J$ is defined as the radius of the sphere containing half of the total angular momentum of the gas.) No clear systematic evolution of this ratio can be seen in any of our tests (the $R_M/R_J$ values all oscillate around unity), so we conclude that angular momentum transport, although undoubtedly present, does not substantially affect the structure of our gaseous discs. We note that this result is a consequence of the symmetry and relatively high rotation rate of our initial conditions. Models with less regular and more slowly rotating initial conditions often pass through highly non-axisymmetric phases, during which there can be substantial exchange of angular momentum within the gas, and between the gas and the dark matter (Katz & Gunn 1991; Navarro & Benz 1991). Such exchange is, of course, physical in origin, and might obscure artificial numerical effects of the kind we are searching for here.

The lack of significant gravitational or viscous torques in our models can be shown even more clearly by examining the angular momentum evolution of individual particles. In an axisymmetric inviscid system, each dark matter and fluid element should conserve its own angular momentum about the axis of symmetry. As a result the deviation of a particle's specific angular momentum ($J_i$) from its initial value provides an estimate of the effects of inhomogeneity (both large-scale effects and effects due to discreteness noise) and of viscosity in the system. The rms $J$-deviation

$$\delta J^2_{\rm rms} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{J_{i,\rm init}-J_{i,\rm final}}{J_{i,\rm init}}\right)^2$$

is shown in Fig. 17 as a function of the number of particles for collapse simulations both with and without cooling. Results are shown for the dark matter and for the gas separately. It can be seen that $\delta J_{\rm rms}$ is fairly small, and drops rapidly with increasing $N$. For $N\sim100$, each particle on average experiences a 10 per cent change in angular momentum, while this number drops to $\sim1$ per cent for $N\sim2000$. $J$-changes in the gas component are only slightly larger than in the dark matter, implying that viscous effects do not cause changes in angular momentum transport that are significantly larger than those associated with the noisy representation of the gravitational field.

In summary, tests with cooling show that reheating via shocks is unable to drive a collapsing gaseous cloud to a quasi-hydrostatic equilibrium state when $\tau_{\rm cool}$ is significantly shorter than $\tau_{\rm dyn}$. In this case, gaseous discs form and reach equilibrium very soon after collapse. In our experiments, the use of small $N_{\rm gas}$ does not seem to lead to any serious unwanted systematic effects. It does, however, lead to serious overestimates of the rate at which energy is radiated. Viscous angular momentum transport does not significantly affect the evolution of these discs.

## 6 COLLAPSE MODELS WITH STAR FORMATION

Although poor understanding of the process of star formation prevents us from devising conclusive tests of the performance of our star formation algorithms, we can gain considerable insight by comparing their results with observation. As described in the last section, the gas component, if allowed to cool, settles rapidly into a thin rotationally supported structure reminiscent of the discs of spiral galaxies. In real discs, there is evidence that birthrates have been constant or slowly declining in time and are quite strongly dependent on the depth of a galaxy's potential well, as measured by its rotation velocity (Kennicutt 1983, 1990). We can check whether our star formation algorithms produce similar trends. Other important questions that we can test are how our computed star formation rates (SFR) depend on the number of particles we use, on the particular star formation algorithm chosen, and on how star formation affects the surrounding gas. The present section presents simulations designed specifically to address these questions.

As discussed in Section 2.4, we model supernova feedback as a mechanism that can heat up the gas surrounding a newly born star, can perturb its local velocity field, or can do both. This introduces two additional input parameters, $f_v$, the fraction of the total available supernova energy that goes into velocity perturbations (the remainder being dumped as heat), and a flag that specifies whether the 'energy' or 'momentum' implementation of the feedback is to be used. Test cases without feedback were performed by forcing both the velocity *and* temperature perturbations to be zero. Since

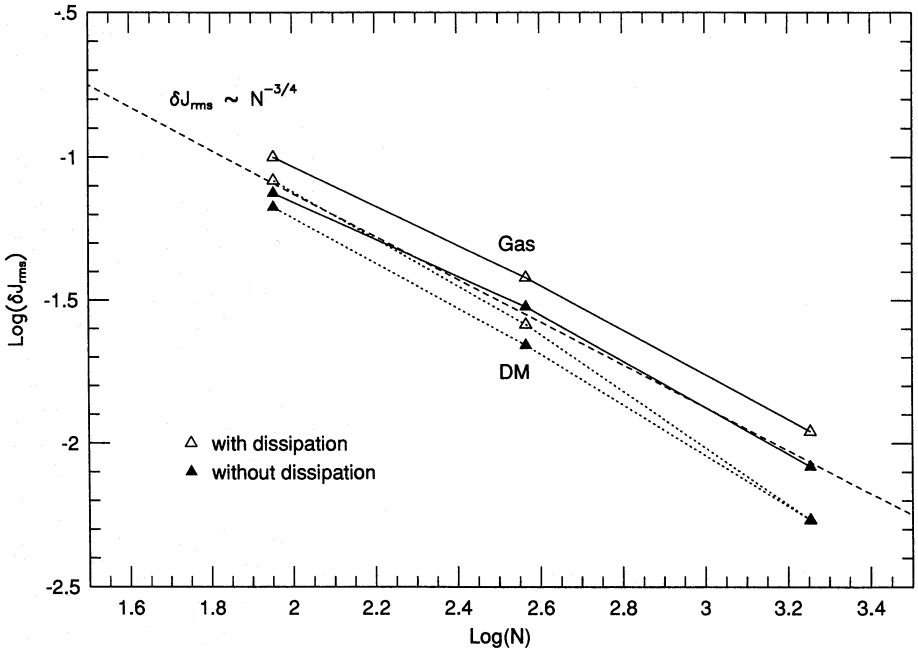


**Figure 17.** The $N$-dependence of the parameters $\delta J_{\rm rms}$, which measures the average fractional change in the angular momentum of a particle. $\delta J_{\rm rms}$ decreases rapidly with $N$ and is practically the same for the gas and the dark matter components. This suggests that pressure forces do not induce any spurious angular momentum changes in a particle in addition to those caused by small-$N$ fluctuations in the gravitational potential.

most stars form in high-density regions where thermal energy can be lost to radiation very quickly, the assumption that feedback just heats the ISM ($f_v = 0$) actually leads to supernovae having little influence on the evolution of the system. Dumping of the supernova energy as velocity perturbations ($f_v > 0$) does affect the star formation rate, because it can halt (or reverse) the local collapse of a region. However, we again stress that all our feedback implementations are very crude, and perhaps qualitatively incorrect representations of how supernovae interact with the multiphase ISM that surrounds real regions of star formation.

The evolution of a collapse with $f_v = 0$ is shown in Fig. 18. The initial parameters were identical to those of some of the simulations without star formation in the last section, i.e. a total mass of $10^{12}$ $M_\odot$ with a gas fraction of 10 per cent; the average rotation velocity of the final disc was again $\sim 270$ km s$^{-1}$. The evolution of the gas component is very similar to that found in the absence of star formation (Fig. 14), except that now some gas particles are reheated and are blown out of the disc. These are, however, a very small fraction of the total mass of the disc, and indeed the disc's structure is scarcely modified by the star formation process. This insensitivity is lost when feedback is assumed to perturb the kinetic energy of the ISM ($f_v > 0$). The gaseous disc is then much more strongly affected, and in some cases it can be almost completely disrupted.

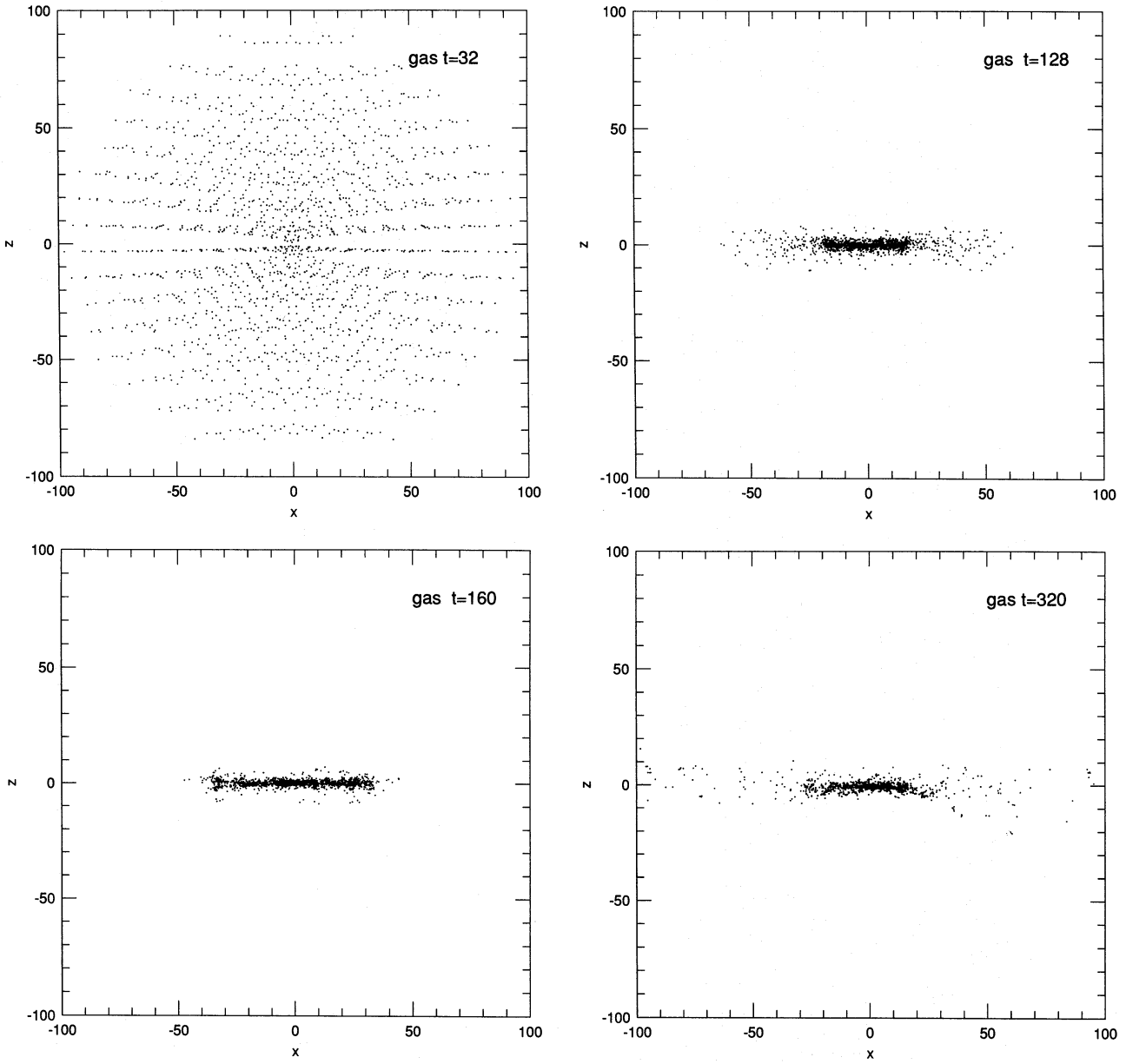


**Figure 18.** Particle plots of the gas and stellar $x$–$z$ distributions at different times during the evolution of a collapse including star formation. Note that the stars form in a thin disc, perhaps even thinner than the gaseous one. The gaseous disc is not disrupted by supernova feedback, which in this model is dumped in the surrounding gas purely as heat ($f_v = 0$). By $t = 320$, $\sim 50$ per cent of the gas mass has been transformed into stars.

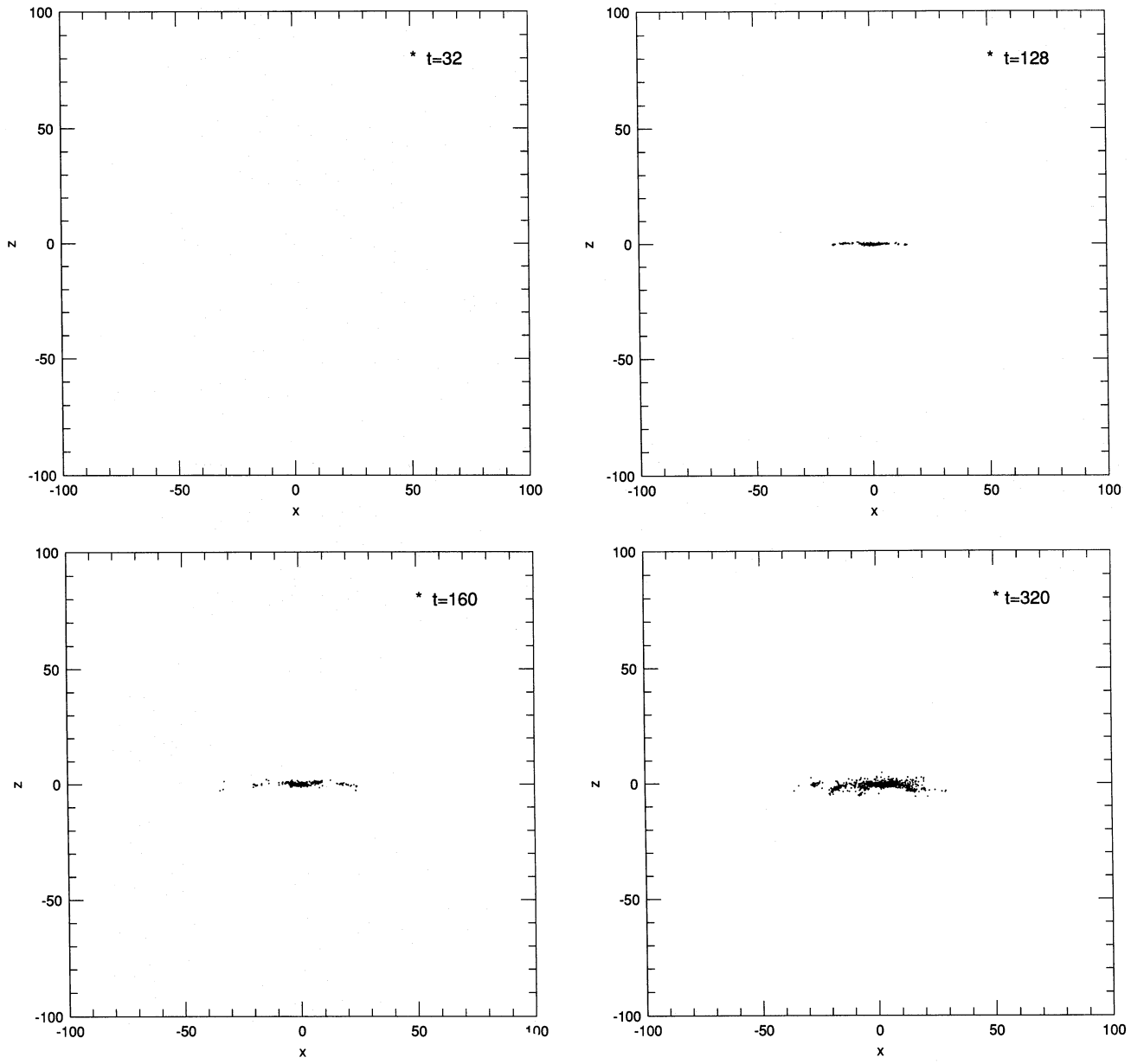


**Figure 18** – *continued*

Stars form in a disc structure similar to, but somewhat thinner than, that of the gas component (Fig. 18). As evolution proceeds, the $z$-structure of the stellar disc seems to evolve very little. This lack of evolution is, however, an artefact of the short time-span of the simulation. In fact, by $t \sim 300$ (code units) the disc has only been able to complete $\sim 3$–4 full rotations. Its thickness is actually increasing with time, and in simulations with smaller $N$ the stellar disc is much less well defined. An estimate of the 'coldness' of a stellar disc is given by the fraction of its kinetic energy that is in rotational motions, $K_{\rm rot}/K$. The evolution of this parameter is shown in Fig. 19 for simulations run with different numbers of particles. The stellar discs become less and less cold with time as fluctuations drive stars away from the nearly circular orbits on which they form. This spurious evolution is slower for larger $N$, and its effects seem to be fairly small for $N_{\rm gas} = 1796$. This is not a surprising result, and similar processes are thought to occur in real discs (e.g. Wielen 1977). It is, however, a significant limitation for simulations of the kind that interest us because it means that we cannot expect to form stable stellar discs with fewer than several thousand particles. Small-$N$ realizations of forming galaxies will then systematically tend to yield 'hot' spheroidal systems rather than cold flattened discs, an undesirable feature which should be borne in mind when analysing the structure and dynamics of 'galaxies' in such simulations.

The growth of the mass in stars in several of these models is shown in Fig. 20(a). As before, tests with 368 and 1796 gas particles (squares and filled triangles, respectively) agree well, but large deviations occur when only 90 gas particles are used (open triangles). Models with $f_v = 0$ and 0.1 (the 'momentum' implementation) are shown by solid and dotted

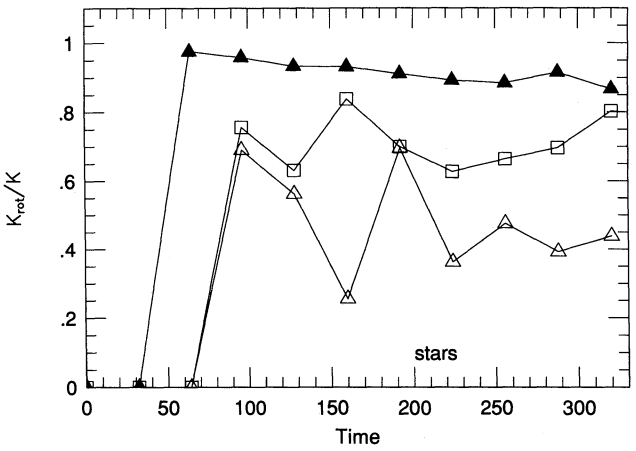


**Figure 19.** The 'coldness' of the stellar discs, as measured by the fraction of the total kinetic energy that is in coherent rotational motions, $K_{rot}/K$. The evolution of this parameter indicates that stars diffuse rapidly away from the quasi-circular orbits on which they are formed. This process is mainly due to small-$N$ fluctuations, as indicated by the fact that the disc formed in the simulation with 1796 gas particles is much 'colder' and evolves more slowly than the others. Only cases with $f_v = 0$ are shown here. Symbols are as in Fig. 3.

lines, respectively. The substantial difference between them clearly indicates the enhanced importance of the feedback when it is allowed to affect the local velocity field of the ISM.

When $f_v = 0$, the star formation rate (SFR) evolution closely resembles that in the 'no feedback' case (shown as a dot–dashed line without symbols). The SFR peaks at collapse ($\sim 50\ M_\odot\ yr^{-1}$), and consumes almost half the gas in just $\sim 8 \times 10^8$ yr. Thereafter it declines to about $\sim 20\ M_\odot\ yr^{-1}$ (upper curve in Fig. 21). At this rate all the remaining gas would be used up in $\sim 2 \times 10^9$ yr, only a small fraction of a Hubble time. When about 70 per cent of the gas has been transformed into stars the SFR declines further, dropping well below $\sim 10\ M_\odot\ yr^{-1}$, the typical value for a galaxy with this circular velocity (White & Frenk 1991). This SFR evolution is similar to that reported by Katz (1992), who also found that the SFR in such systems peaks at collapse and declines rapidly thereafter. His implementation of feedback resembles ours with $f_v = 0$.

The good agreement between 'no feedback' tests and those with $f_v = 0$ indicates that if supernova energy is dumped as heat then feedback does not play a significant role in regulating star formation rates. The transformation of gas into stars proceeds so rapidly in this case that most disc galaxies would have depleted their gas content substantially

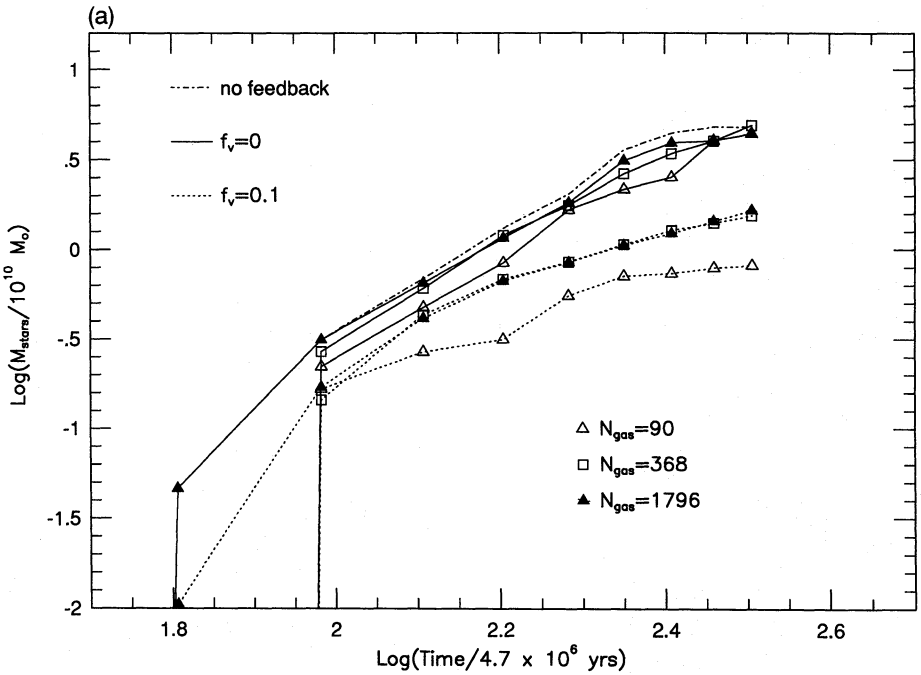


**Figure 20.** (a) The cumulative mass in stars as a function of time in a gaseous disc whose equilibrium circular velocity is $V_c \sim 270$ km s$^{-1}$. The results of equivalent tests with different numbers of particles are shown. Dotted lines indicate the cases where part of the supernova energy is fed back into the ISM in the form of velocities ($f_v = 0.1$). Tests with $f_v = 0$ are shown with solid lines. The typical evolution when the feedback is completely neglected is shown by the dot–dashed line without symbols. The total duration of the simulations spans only 3–4 full rotations of the disc. The evolution with $N_{gas} = 90$ is clearly discrepant, but the tests with 368 and 1796 gas particles agree quite well. (b) as (a), and again using our 'momentum' implementation of feedback, but for different values of $f_v$, as indicated next to each curve. Note the importance of $f_v$ in regulating the evolution of the star formation rate. Small values of $f_v$ ($\ll 0.1$) result in high star formation rates that would consume the available gas in much less than a Hubble time. Large values of $f_v$ ($> 0.2$) lead to large-scale winds which can disrupt the gaseous disc and suppress further star formation. $f_v \sim 0.1$ seems to be a reasonable choice, preserving the gaseous disc almost undisrupted and giving a star formation rate of $\sim 10\ M_\odot\ yr^{-1}$, in good agreement with observational estimates for a spiral galaxy with $V_c \sim 270$ km s$^{-1}$. (c) as (b), but using our 'energy' implementation of feedback.

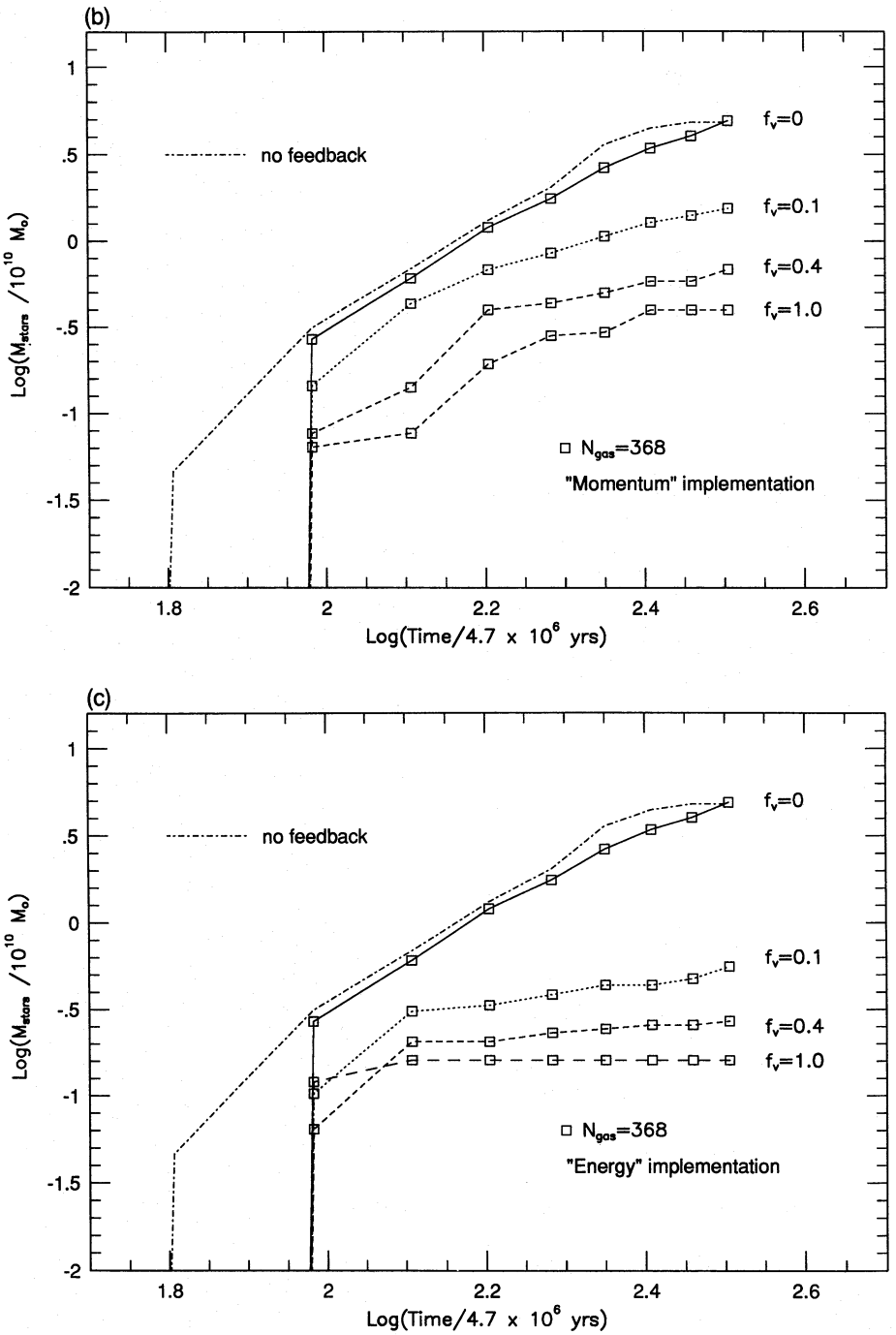


**Figure 20** – *continued*

by the present time. This might be a viable description for early-type spirals, but observed star formation rates in Scs and later types are clearly higher than is allowed by this kind of SFR history (Kennicutt 1983, 1990). Furthermore, some process such as supernova feedback must limit star formation rates at early times if hierarchical clustering models for galaxy formation in a flat universe are to be viable (White & Rees 1978; Cole 1991; White & Frenk 1991). Thus, although $f_v = 0$ might be appropriate for some systems, it seems that this scheme does not match the slow SFR evolution observed in many spirals, and it may lead to overproduction of stars in current cosmogonical theories unless additional processes are invoked to suppress star formation.

Stars form significantly more slowly when $f_v > 0$ (dotted lines in Fig. 20a). Even if only 10 per cent of the available energy is allowed to perturb the velocity field of the ISM ($f_v = 0.1$), large modifications of the SFR are produced. In this case, star formation begins at collapse and quickly reaches a roughly constant rate of about 11 $M_\odot$ $yr^{-1}$ which is maintained for the rest of the simulation. The gas consumption time in this case is $\sim 10^{10}$ yr, comparable to a Hubble time, and the implied SFR evolution is similar to that

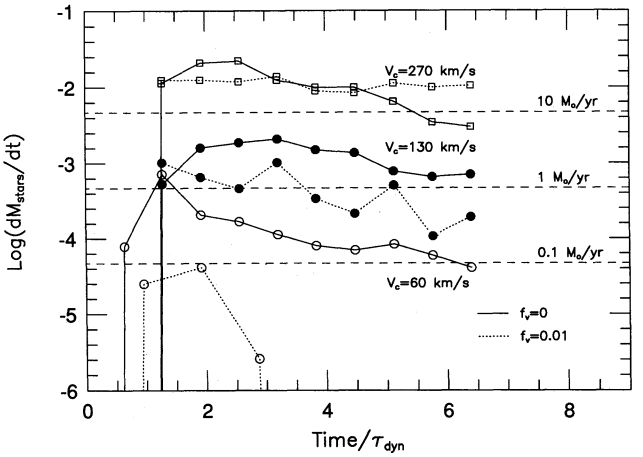


**Figure 21.** Long-term evolution of the star formation rate in discs of different masses and circular velocities. Solid and dotted lines are used to indicate tests with $f_v = 0$ and 0.01, respectively, while the symbols denote systems with different circular velocities. Only 368 gas particles were used in these long runs. If $f_v = 0.01$, the average star formation rates for different values of the circular velocity agree quite well with observational estimates of those in spiral and irregular galaxies (see Table 2). As discussed in the text, $f_v = 0$ seems to lead to rapidly evolving SFRs and short gas consumption times, particularly in the largest systems.

inferred for Scs. Very similar evolution is seen when either 368 or 1976 gas particles are used, but the test with $N_{gas} = 90$ again gives seriously discrepant results.

Figs 20(b) and (c) show the evolution of the mass in stars for various strengths of our 'momentum' and 'energy' implementations of feedback. The dramatic effect of dumping supernova energy into bulk motion is very clear. Such feedback can not only regulate the SFR evolution very effectively, but may even suppress further star formation altogether. For example, when *all* their energy is assumed to affect the kinetic energy of the ISM (the curves labelled $f_v = 1.0$ in Figs 20b and c), supernovae are able to drive a wind that completely disrupts the gaseous disc, thereby preventing further star formation. When the 'energy' implementation is used, such disruption occurs even if $f_v$ is as small as $\sim 0.1$ due to the reversal of convergent gas flows that is forced by this algorithm. Star formation rates obtained with $f_v = 0.1$ and this implementation ($< 1\ M_\odot\ yr^{-1}$) are too low compared with those typically observed in disc galaxies of comparable circular velocity.

Much less dramatic results are obtained when our 'momentum' implementation is adopted, and for $f_v = 0.1$ the gaseous disc survives largely unaffected, forming stars at a rate that seems to be roughly balanced by the feedback. A value of $f_v = 0.2$ (not shown in Fig. 20) gives a star formation rate of $\leq 5\ M_\odot\ yr^{-1}$, perhaps a little low for a system with a rotation velocity of 270 km s$^{-1}$. Larger values of $f_v$ give even lower SFRs. We therefore conclude that for this algorithm $f_v \leq 0.1$ seems to be a good choice, since such values can prevent the sudden formation of most of the stars in a disc and can give star formation histories that span the range inferred to give rise to observed galactic discs.

In order to investigate the long-term SFR evolution in more detail, we ran a few models for a substantially longer period after collapse. Running such long tests with 1796 gas particles is quite expensive, so only 368 were used in these models. We also used these tests to explore the dependence of star formation rates on disc circular velocity (or mass). This was accomplished by running models with the same gas fraction $f_b = 0.1$, the same $\lambda = 0.1$, the same $N_{gas} = 368$ and the same initial projected density $M_{tot}/R_{tot}^2$, but with different mass and so different final disc circular velocity. We tried three different combinations of $M_{tot}$ and $R_{tot}$, which resulted in the following average equilibrium rotation velocities: 270, 130 and 60 km s$^{-1}$, respectively. We ran these tests for 6 $\tau_{dyn}$ following the collapse of each system.

The resulting star formation rates, shown in Fig. 21, are significantly higher for larger circular velocities, even when expressed as a fraction of the total gas mass. Tests with $f_v = 0$ and 0.01 (the 'momentum' implementation) are shown by solid and dotted lines, respectively. As discussed above, when $f_v = 0$, or in the absence of feedback, most of the gas is transformed into stars in a rapid burst of star formation just after collapse. The star formation rate in this case declines steadily at later times. Inclusion of as little as 1 per cent of the feedback energy in velocity perturbations has a significant effect on the SFRs, especially for low circular velocities. In fact, for $V_c = 60$ km s$^{-1}$ star formation is almost halted once about 10 per cent of the initial gas has been transformed into stars. Supernovae are in this case able to expel about half of the initial gas from the system; the rest is pushed out of the disc into the halo. The bound material will eventually cool and fall back into the disc, leading to recurrent bursts of star formation. A similar 'bursting' behaviour can be seen in the model with $V_c = 130$ km s$^{-1}$, but it is not evident in the high $V_c$ case, where the energy from supernovae is insufficient to expel much gas.

Table 2 indicates the fraction of the initial gas that is expelled by supernovae-driven winds ($f_{loss}$). This fraction depends strongly on $V_c$ and $f_v$, and can be very substantial in weakly bound systems (98 per cent for $V_c = 60$ km s$^{-1}$ and $f_v = 0.1$). The effect is less dramatic in systems with higher $V_c$, where the velocity perturbations seem to allow the system to regulate its star-forming activity. It is clear from Table 2 and Fig. 21 that we need $f_v \leq 0.01$ for a significant fraction of stars to form in haloes with circular velocities below 100 km s$^{-1}$. Quantitatively, the $f_v = 0.01$ SFRs average $\sim 20\ M_\odot\ yr^{-1}$ for $V_c \sim 270$ km s$^{-1}$; $\sim 1\ M_\odot\ yr^{-1}$ for $V_c \sim 130$ km s$^{-1}$; and $\sim 0.02\ M_\odot\ yr^{-1}$ for $V_c \sim 60$ km s$^{-1}$, in rough agreement with estimates of the current star formation rates of some spirals

**Table 2.** Main parameters of the star formation runs.

| $V_c$ [km/s] | $f_v$ | $f_{loss}$ | $< \dot{M}_* >$ [$M_\odot$/yr] | Peak $\dot{M}_*$ [$M_\odot$/yr] |
|---|---|---|---|---|
| 270 | 0.00 | 0.00 | 21.2 | 47.1 |
| | 0.01 | 0.00 | 20.0 | 29.3 |
| | 0.10 | 0.09 | 12.4 | 17.2 |
| 130 | 0.00 | 0.00 | 2.35 | 4.46 |
| | 0.01 | 0.10 | 0.96 | 2.17 |
| | 0.10 | 0.43 | 0.20 | 0.63 |
| 60 | 0.00 | 0.00 | 0.37 | 1.53 |
| | 0.01 | 0.54 | 0.02 | 0.09 |
| | 0.10 | 0.98 | 7e-3 | 0.03 |

and irregulars (these averages were computed over $\sim 6\ \tau_{\rm dyn}$, the time-span of the simulations).

Interestingly, if we represent the dependence of the average star formation rate on the disc rotation velocity as a power law, $\langle \dot{M}_* \rangle \propto V_c^x$, the exponent $x$ is found to be a function of the feedback parameter $f_v$. In particular, $x \approx 3$, 4 and 5 correspond to choices of $f_v = 0$, 0.01 and 0.1, respectively. The existence of correlations between the total mass in stars and the circular velocities of disc galaxies can then be used to constrain the values of the feedback parameter one should use in cosmological simulation, perhaps reducing the arbitrariness of the choice.

To summarize the results of this subsection, the star formation algorithm implemented in our code seems to give similar results for simulations with $N_{\rm gas} = 368$ and 1796. As in previous tests, large discrepancies are found if as few as 90 gas particles are used. Stars form in a thin disc on nearly circular orbits, but diffuse away from them as a result of small-$N$ fluctuations in the gravitational potential. Star formation peaks at collapse, when the bulk of the disc is formed, and tends to a constant value afterwards. Feedback processes seem crucial to prevent runaway star formation, but in our code they only have a significant effect when implemented as perturbations to the kinetic energy of the ISM. The average SFR in a disc depends strongly on its circular velocity, and agrees quantitatively with observational estimates for spirals if $\sim 1$–$10$ per cent of the available energy is dumped in bulk motions. Feedback from supernovae may induce a bursting behaviour in low-mass systems, where it may also lead to substantial gas loss.

Although encouraging, results based on these (or any other) star formation algorithms should be considered preliminary and highly uncertain. In particular, several effects that are regarded as essential in successful models of disc formation have not been considered here. These include the continuous replenishment of the gas supply to the disc and the role of mergers and accretions of satellite galaxies. It is possible that the preferred values of $f_v$ found in this work turn out to be highly dependent on these effects. A final answer will probably have to wait until we have a much better understanding of this complicated subject.

## 7 CONCLUSIONS

We have described tests of a code designed to follow the evolution of a mixture of collisional and collisionless fluids in three dimensions. The code combines a novel approach to collisional fluid dynamics (SPH) with the speed and flexibility of an $N$-body code based on a tree data structure. The tests are relevant to our final aim of simulating the formation of galaxies in a hierarchically clustering universe, and in particular address the shock-heating of cold gas during the collapse, virialization and merging of rotating spherical systems of gas and dark matter. We especially focus on the behaviour of simulations with small numbers of particles, in order to understand the nature of the systematic effects that can arise in small-$N$ clumps, and to identify the lowest well-resolved level of the hierarchy.

Other effects analysed in these tests include radiative cooling, star formation, and the feedback effects of supernovae, all of which are expected to play a major role in the formation of galaxies in a hierarchical scenario. Our main conclusions are as follows.

(1) The code seems to handle the shock-heating of initially cold gas quite well. Accretion shock tests show that, in post-shock regions, the code gives results that compare very well with analytic similarity solutions. The deviations that are present when only a small number of particles lie within the shock do not seem to compromise the further evolution of the system, which converges to the correct self-similar solution as soon as $N \geq 250$ particles are inside the shock. The inclusion of a collisionless component does not substantially alter the performance of the code, although the noise induced in the gravitational potential somewhat increases oscillations in the post-shock regime.

(2) More than $\sim 100$ gas particles are needed in order to follow the shock-heating process in a quasi-spherical collapse. Although the tests we have carried out do not allow us to determine exactly how many particles are needed, it seems that $N_{\rm gas} \geq 250$–$350$ is safe. Adiabatic tests show that dark matter and gas settle into similar equilibrium structures. The gas is heated up to the virial temperature of the system at collapse, and reaches hydrostatic equilibrium soon thereafter. Small-$N$ systems suffer from spurious energy transfer between the gas and dark matter components. The gas always seems to gain energy from the dark matter.

(3) During mergers between systems containing a mixture of dark matter and adiabatic gas, there is a transfer of energy from the dark matter to the gas. The gas thus ends up less tightly bound. This effect increases with the orbital energy of the collision, and can lead to substantial differences in the spatial distributions of gas and dark matter in the final merged object.

(4) Errors that occur in a hierarchical clustering simulation at early times, when non-linear clumps are poorly resolved, do not significantly affect the final structure of the gas component provided that later evolution entails a fairly large number of mergers. This is because, in situations where the specific binding energy of matter in clumps increases with time, the spurious energy perturbations that occur in small-$N$ collapses are insignificant in the final energy budget. Structure growing at the rate expected in a universe with a power-law perturbation spectrum index, $n \approx -1.5$ (appropriate on a galactic scale in a CDM universe), satisfies this condition. Our techniques thus seem adequate for simulating the evolution of structure in such a universe. We note, however, that spurious effects are most apparent in the inner regions of clumps, and that considerable caution is therefore needed when analysing the structure of these regions for comparison, for example, with the cooling and X-ray emitting regions of galaxy clusters.

(5) If cooling is as important as expected in the inner regions of protogalaxies ($\tau_{\rm cool} \ll \tau_{\rm dyn}$), the gas can radiate very efficiently and shocks are unable to drive a significant fraction of it into the hot phase. As a result, a gaseous disc forms rather quickly after collapse and evolves little thereafter. Viscous effects are more important in small-$N$ systems, and can lead to substantial overestimates of the amount of energy radiated by the equilibrium disc. The discs formed in our tests appear to be stable, a result probably related to the high initial rotation we adopt ($\lambda = 0.1$).

(6) The star formation algorithms we test can yield star formation rates that are in rough agreement with observa-

tions of spirals when supernova feedback is allowed to influence the surrounding gas with a suitably chosen efficiency, and so to regulate the local star formation rate. However, small-$N$ fluctuations prevent the formation of fully stable stellar discs unless several thousand particles are used. This is likely to be a serious problem when attempting to identify the structure of the 'galaxies' formed in large-scale cosmological simulations of galaxy formation.

(7) An even more critical problem may turn out to be the fact that the qualitative outcome of a simulation can vary dramatically as a function of the scheme adopted to treat hydrodynamic feedback from supernovae. Since a variety of arguments suggest that such feedback played a major role in the formation of real galaxies, our inability to treat it other than in a highly schematic manner may prove a fundamental limitation to the reliability of all presently conceivable galaxy formation simulations. This possible difficulty clearly needs further investigation.

## ACKNOWLEDGMENTS

JFN would like to acknowledge enlightening discussions on SPH with W. Benz, M. Herant and M. Davies. The code described in this paper is largely based on a copy of the SPH code kindly made available by Willy Benz. JFN gratefully acknowledges the hospitality of the Institute of Astronomy, Cambridge, where most of the work reported in this paper was performed.

## REFERENCES

Benz W., 1990, in Buchler J. R., ed., The Numerical Modelling of Nonlinear Stellar Pulsations. Kluwer, Dordrecht, p. 180
Benz W., Bowers R. L., Cameron A. G. W., Press W. H., 1990, ApJ, 348, 647
Bertschinger E., 1985, ApJS, 58, 39
Cen R. Y., Jameson A., Liu F., Ostriker J. P., 1990, ApJ, 326, L41
Cole S., 1991, ApJ, 367, 45
David L. P., Forman W., Jones C., 1990, ApJ, 359, 29
Efstathiou G., Frenk C., White S. D. M., Davis M., 1988, MNRAS, 235, 715
Evrard A. E., 1988, MNRAS, 235, 911
Evrard A. E., 1990, ApJ, 363, 349
Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189
Fall S. M., Rees M., 1985, ApJ, 298, 18
Hernquist L., 1989, Nat, 340, 687
Hernquist L., Katz N., 1989, ApJS, 70, 419
Katz N., 1992, ApJ, 391, 502
Katz N., Gunn J., 1991, ApJ, 377, 365
Kennicutt R. C., 1983, ApJ, 272, 54
Kennicutt R. C., 1990, in Schull J. M., Thronson H. A., eds, The Interstellar Medium in External Galaxies. Kluwer, Dordrecht, p. 405
Lin D. C., Pringle J., 1987, ApJ, 320, L87
Monaghan J. J., 1985, Comput. Phys. Rep., 3, 71
Navarro J. F., Benz W., 1991, ApJ, 380, 320
Toomre A., 1964, ApJ, 139, 1217
White S. D. M., 1991, in Combes F., Casoli F., eds, Dynamics of Galaxies and Their Molecular Cloud Distributions. Kluwer, Dordrecht, p. 383
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Wielen R., 1977, A&A, 60, 263