Mon. Not. R. Astron. Soc. **343,** 75–106 (2003)

GALICS – I. A hybrid *N*-body/semi-analytic model of hierarchical galaxy formation

Steve Hatton,[1] Julien E. G. Devriendt,[1,2]★ Stéphane Ninin,[1] François R. Bouchet,[1] Bruno Guiderdoni[1] and Didier Vibert[1]

[1]*Institut d'Astrophysique de Paris, 98 bis Boulevard Arago, 75014 Paris, France*
[2]*Oxford University, Astrophysics, Keble Road, Oxford OX1 3RH*

Accepted 2003 March 17. Received 2002 January 18; in original form 2001 June 20

ABSTRACT

This is the first paper of a series that describes the methods and basic results of the GALICS model (*Galaxies In Cosmological Simulations*). GALICS is a hybrid model for hierarchical galaxy formation studies, combining the outputs of large cosmological *N*-body simulations with simple, semi-analytic recipes to describe the fate of the baryons within dark matter haloes. The simulations produce a detailed merging tree for the dark matter haloes, including complete knowledge of the statistical properties arising from the gravitational forces. We intend to predict the overall statistical properties of galaxies, with special emphasis on the panchromatic spectral energy distribution emitted by galaxies in the ultraviolet/optical and infrared/submillimetre wavelength ranges.

In this paper, we outline the physically motivated assumptions and key free parameters that go into the model, comparing and contrasting with other parallel efforts. We specifically illustrate the success of the model in comparison with several data sets, showing how it is able to predict the galaxy disc sizes, colours, luminosity functions from the ultraviolet to far infrared, the Tully–Fisher and Faber–Jackson relations, and the fundamental plane in the local Universe. We also identify certain areas where the model fails, or where the assumptions needed to succeed are at odds with observations, and pay special attention to understanding the effects of the finite resolution of the simulations on the predictions made. Other papers in this series will take advantage of different data sets available in the literature to extend the study of the limitations and predictive power of GALICS, with particular emphasis put on high-redshift galaxies.

Key words: galaxies: evolution – galaxies: formation.

1 INTRODUCTION

Over the last 5 years, the discovery of the cosmic infrared background (CIB) (Puget et al. 1996; Guiderdoni et al. 1997; Fixsen et al. 1998; Hauser et al. 1998) and the faint galaxy counts with *ISO* at 15 μm (Elbaz et al. 1999) and 175 μm (Puget et al. 1999), SCUBA at 850 μm (Smail, Ivison & Blain 1997) and MAMBO at 1.3 mm (Carilli et al. 2000) have shown that approximately two-thirds of the luminosity budget of galaxies is emitted by dust in the infrared (IR)/submillimetre range. Whilst the nature of the source that heats up the dust is still uncertain, it now seems increasingly plausible that the contribution of active galactic nuclei (AGN) to heating is not dominant, and that most of the energy is powered by starbursts due to gas inflows triggered by close encounters and merging. The IR/submillimetre wavelength range is actually tracking the star formation rate history of the Universe more accurately than the ultraviolet (UV) range, with strong sensitivity to the merging phenomenon that is the signpost of hierarchical galaxy formation. The luminous and ultraluminous infrared galaxies that contribute to the CIB are thought to be the progenitors of the bulges and elliptical galaxies in the local Universe. The goal of GALICS is to obtain a consistent panchromatic description of this process and of the luminosity budget of galaxies as it appears, for instance, through the faint galaxy counts at various wavelengths in the optical, IR and submillimetre ranges.

Various pieces of work have converged to build up a consistent description of galaxy formation within the paradigm of hierarchical clustering. Initial density perturbations are gravitationally amplified and collapse to form almost relaxed, virialized structures called dark matter (DM) haloes. In all variants of the cold dark matter (CDM) scenario (Peebles 1982; Blumenthal et al. 1984), smaller haloes form first, and bigger haloes form continuously from the

★E-mail: jeg@astro.ox.ac.uk

collapse of smaller haloes. Gas radiates and cools down in the potential wells of the haloes (White & Rees 1978). Haloes have little angular momentum, and dissipative collapse stops when the cold gas settles in rotationally supported discs (Fall & Efstathiou 1980; Dalcanton, Spergel & Summers 1997; Mo, Mao & White 1998). Star formation at galaxy scales can be reasonably described with Schmidt laws or more sophisticated recipes (Kennicutt 1989). Models of spectrophotometric evolution of the stellar populations produce luminosities, spectra and colours from the star formation rate histories and initial mass function (IMF) (Bruzual 1981; Guiderdoni & Rocca-Volmerange 1987; Bruzual & Charlot 1993), and are now extended to the modelling of dust thermal emission (Mazzei, Xu & de Zotti 1992; Silva et al. 1998; Devriendt, Guiderdoni & Sadat 1999, hereafter DGS). When they die, stars eject gas, heavy elements and energy into their environment. The energy feedback heats up the remaining gas and can produce galactic winds in the shallower potential wells that deplete the galaxies and quench subsequent star formation (Dekel & Silk 1986). Finally, spheroids form from major mergers (Toomre & Toomre 1972; Kent 1981). If the spheroid can still be a centre for gas cooling, a new disc forms around this bulge. As a result, the morphological and spectral types of galaxies are not fixed once and for all, but rather evolve as star formation, gas accretion and merging occur.

These ingredients can be put together with some success within a fully semi-analytic model (hereafter SAM) that starts from the power spectrum of linear fluctuations, and follows the various processes right up to spectral energy distributions of stellar populations. For instance, under the assumption that each newly collapsed peak produces a halo where a new galaxy forms, therefore neglecting the classical 'cloud-in-cloud' problem, it is possible to reproduce at least qualitatively the main statistical properties of galaxies (Lacey & Silk 1991; Lacey et al. 1993; Guiderdoni et al. 1998; Devriendt & Guiderdoni 2000). The extended Press–Schechter prescription (EPS, Press & Schechter 1974; Bond et al. 1991; Bower 1991; Lacey & Cole 1993) is a more efficient tool for describing gravitational collapse and estimate the merging history trees of dark matter haloes. The fate of the stars, gas and heavy elements can also be followed within the hierarchy of merging haloes with an implementation of these ingredients in the EPS formalism (White & Frenk 1991). However, it is only through Monte Carlo realizations of halo merging history trees (Kauffmann & White 1993; Lacey & Cole 1994; Somerville & Kolatt 1999) that galaxy merging can be followed and hierarchical galaxy formation can be addressed. Implementing a simple recipe for dynamical friction of satellite galaxies in the potential wells of haloes enabled Kauffmann, White & Guiderdoni (1993) and Kauffmann, Guiderdoni & White (1994) to follow the galaxy merging history. Though the 'block model' for hierarchical structure formation has been used for some time (Cole et al. 1994), most studies now involve this type of random realization based on the EPS for the merging history (Somerville & Primack 1999, hereafter SP99; Cole et al. 2000, hereafter CLBF and papers of these series). In addition to the cosmological parameters (H_0, Ω_0, Ω_Λ, Ω_B, the shape of the power spectrum and its normalization σ_8), the semi-analytic method introduces a limited set of free parameters because some of the processes have to be addressed phenomenologically: in the most general case they can be reduced to a star formation efficiency, a stellar feedback efficiency for the interstellar medium (ISM)/intergalactic medium (IGM), and a parameter that describes our ignorance on the complicated merging processes. These parameters are determined by requiring the results to fit certain data sets, commonly including the K-band luminosity density in the local Universe, the number of dwarf galaxies (which are the most sensitive objects to stellar feedback) and the number of elliptical galaxies (which only form from major mergers in the simplest hierarchical scenario). Once the free parameters are fixed, many predictions can be produced and compared with data.

A large number of papers have been devoted to various aspects of galaxy formation and evolution, ranging from the Butcher–Oemler effect (Kauffmann 1995), the formation of discs and bulges (Kauffmann 1996a), damped Lyman-α systems (Kauffmann 1996b), Lyman-break galaxies (Baugh et al. 1998; Somerville, Primack & Faber 2001) and the parallel evolutions of quasars and galaxies (Kauffmann & Haehnelt 2000).

However, this approach still suffers from a number of shortcomings. First, even if the EPS agrees with N-body simulations (Efstathiou et al. 1988; Kauffmann & White 1993; Lacey & Cole 1993, 1994; Somerville et al. 2000), it is clearly a limiting simplification of the complex dynamical processes that actually occur. The non-linear dynamics is computed with the top-hat model that assumes sphericity and homogeneity, and it overestimates the number of haloes on galactic and group scales (Gross et al. 1998). The only pieces of spatial and dynamical information that are stored in the halo merging history trees are the virial radii and circular velocities. There is no information on the spatial distribution and peculiar velocities of haloes. Consequently, the outputs of SAM cannot be used for synthesizing realistic mock catalogues and images that take into account spatial correlations, whereas there is an increasing need for these catalogues and images to analyse current and forthcoming observations and to test data processing techniques.

It is tantalizing to bypass some of these limits by using merging history trees produced from large cosmological N-body simulations. The basic idea is to obtain a description of the dark matter halo merging history trees, which can be computed accurately from simulations, and to keep the usual semi-analytic approach to model the more uncertain physics of baryons. This model rests on the assumption that baryons do not alter significantly the dynamics of the dark matter, except on the smallest scales. Hence the name 'hybrid model'. The result is a more realistic merging history for haloes, which necessarily reflects on galaxy formation and evolution. The drawback is a loss of flexibility since the values of the cosmological parameters H_0, Ω_0 and Ω_Λ, and the choice of the power spectrum $P(k)$ and normalization σ_8, are built in the N-body simulation. However, the value Ω_B, the physics of baryons and the associated free parameters can be changed *ad libitum*, very much as in classical semi-analytic models.

The first attempts at a hybrid model were proposed by White et al. (1987) and Roukema et al. (1997). Fewer than 10 snapshots of N-body simulations were considered at that time, and this crude time resolution had a significant impact on the results. Moreover, it is also obvious that any numerical simulation has a finite mass resolution, and that the unknown fate of baryons in systems below this threshold is going to propagate over the threshold in the picture of hierarchical clustering. A partial solution of this problem is to use the spatial information of a numerical simulation, but to build halo merging history trees from Monte Carlo realizations of the EPS (Kauffmann, Nusser & Steinmetz 1997; Benson et al. 2000; Governato et al. 2001). Unfortunately, it becomes impossible to follow the evolution of galaxies backwards in the structures, and the merging trees are not consistent with the merging histories of the individual haloes in the simulation. Only fully hybrid models keep this record and have fully consistent merging history trees (Kauffmann et al. 1999, hereafter KCDW, and following papers). Moreover, only the most recent hybrid models, taking advantage

of very high-resolution *N*-body simulations to dynamically follow substructure within dark matter haloes, are capable of accurately tracking properties of individual galaxies within clusters (Springel et al. 2001).

Here we propose the GALICS model of hierarchical galaxy formation, which is intended to provide a fully panchromatic description of the galaxy merging history, similar in spirit to that implemented by Granato et al. (2000) in a purely semi-analytic model. For that purpose, we will follow chemical and spectrophotometric evolution in a consistent way, estimate dust extinction and radiation transfer and compute spectral energy distributions of dust thermal emission, following the lines of the STARDUST model (DGS).

Our main goals are to:

(i) present an original hybrid model that is entirely independent of previous attempts, and to study its successes and failures compared with other models;

(ii) present an overall view of galaxy evolution by producing a host of predictions from a 'standard' reference model to compare with observed galaxy properties locally and at high redshift;

(iii) produce a panchromatic picture, which is closely linked to the hierarchical formation of structure, as galaxy mergers are bright in the infrared;

(iv) implement the effects of observational selection criteria that follow as closely as possible the actual observational processes;

(v) study the effect of mass and time resolution constraints on these predictions.

This paper (the first in a series) proposes an overall presentation of the model and the basic predictions for those statistical properties of galaxies in the local Universe that can be more easily compared with other works. In Section 2 we describe the procedure that has been developed to find the haloes and build the halo merging history trees from the series of output snapshots. Section 3 presents the 'recipes' for cooling of hot gas in the haloes. In Section 4, we introduce a simple, standard implementation of the dissipative physics of baryons, and the construction of the galaxy merging history trees. Section 5 describes the modelling of merging events, and how these events are assumed to drive galaxy morphologies. In Section 6 we present our methods for determining galaxy luminosities, which are largely based on the STARDUST model of DGS. Section 7 briefly summarizes the few free parameters that go into the models. In Section 8, we show results for the $z = 0$ properties of galaxies in our reference model, including sizes, colours, luminosity functions and structural relations [Tully–Fisher, Faber–Jackson, fundamental plane (FP)]. In Section 9 we investigate in detail the effect that the finite spatial, mass and time resolution of our simulations has on the predictions of galaxy properties. Section 10 presents a discussion of these first results. Appendix A describes the parallelized treecode that we use for large cosmological *N*-body simulations.

Four other papers will complete this description of the model. In Paper II (Devriendt et al., in preparation), we will explore the sensitivity of our reference model to changes in some of the modelling assumptions, and study the influence of variations in the choice of cosmology, recipes for the baryonic processes, and astrophysical free parameters. This paper will also discuss the evolution of galaxy properties. In Paper III (Blaizot et al., in preparation) we will focus more specifically on predicted properties for Lyman-break galaxies at a redshift of 3. In Paper IV (Devriendt et al., in preparation), we will give faint galaxy counts, source counts, and angular correlation functions in the UV, optical, IR and submillimetre ranges. Finally, in Paper V (Blaizot et al., in preparation), we will focus on the redshift distributions of different classes of galaxy, the redshift evolution of the spatial correlation function, $\xi(r)$, and statistical bias, b, and the influence of environment on galaxy properties. Forthcoming papers of the series will address several other issues with an improved modelling of the baryonic processes.

2 DARK MATTER HALOES

For the simulation we have used a parallel tree-code written and optimized for the Cray T3E. This code, based on hierarchical methods, is described in detail in Ninin (1999). We assume that, initially, the baryon density field traces that of the dark matter, and thus that galaxy formation occurs at local maxima of the underlying density distribution. For a given cosmology it can be shown that there is a certain turnaround density contrast, above which matter has collapsed into gravitationally bound systems, which have separated from the expansion of the Universe. The precise density contrast at which to define a 'virialized' halo is a matter of some debate (White 2001), and in general depends on cosmology. A physically sensible definition is that within the virial radius of an object the dark matter is virialized, and external to it material is infalling. This is found to occur at around 200 times the critical density, regardless of cosmology. It is this latter definition we will use in this work. We refer to regions attaining this overdensity as dark matter haloes, and it is assumed that all galaxy formation processes take place within these haloes, and furthermore that there is no communication between these haloes.

2.1 Simulation details

We outline the details of the treecode approach in Appendix A. We have initially simulated a spatially flat, low-density model (ΛCDM) with a cosmological constant, in a cube of comoving size $L_{\rm box} = 100\,h^{-1}$ Mpc. The amplitude of the power spectrum was set by demanding an approximate agreement with the present-day abundance of rich clusters (Eke, Cole & Frenk 1996). Initial conditions were obtained with the GRAFIC code (Bertschinger 1995). A logarithmic spacing in the expansion parameter, a, was used for the output file times: we finally had around 100 output files.

The simulation contains $N_{\rm part} = 256^3$ particles. We ran additional simulations at lower resolution ($N_{\rm part} = 128^3, 64^3$), which we will use to test for convergence in the properties of our dark matter haloes and galaxies (see Section 9). The full details of our standard simulation are summarized in Table 1.

Table 1. Simulation parameters used in our standard (ΛCDM, $N_{\rm part} = 256^3$) simulation.

Box size L (Mpc)	150.0
Particle mass (10^{11} M$_\odot$)	0.08272
Omega matter Ω_0	0.333
Omega lambda Ω_Λ	0.667
Omega curvature Ω_c	0.0
Hubble parameter h	0.667
Variance σ_8	0.88
$\Gamma = \Omega_0 h$	0.22
Initial redshift	35.59
Number of time-steps	19269
Number of outputs	100
Spatial resolution (kpc)	29.29
Work time (10^3 h)	56.

2.2 Detection of the haloes in the simulation

Many algorithms exist for detecting haloes of dark matter in N-body simulations. Examples include DENMAX, the spherical overdensity algorithm, and the 'hop' algorithm. None of them, in spite of their complexity, has proved to be clearly superior to the simplest of all, the 'friend-of-friend' (FOF) algorithm (Davis et al. 1985). This algorithm links two particles in the same group if their distance is less than a certain linking length. FOF is very simple, and has the advantage of depending on only one parameter, the linking length. This is usually expressed as a fraction of the mean interparticle separation. Groups of particles are thus identified which are bound by a density contrast

$$\delta_{\rm thresh} \approx 3/(2\pi b^3). \tag{2.1}$$

For the case where the agglomeration is well modelled by an isothermal sphere, the average density is three times the threshold density, so picking a link-length parameter $b = 0.2$ identifies groups that are at average densities approximately 180 times the mean density (Cole & Lacey 1996). So, as a first step, we detect haloes in every output time-step of the simulation with this algorithm and create a linked list of particles in each detected halo. In practice, we allow b to evolve with time such that the FOF picks up structures at an overdensity around 200 times that of the critical density. It has been shown (KCDW) that the minimum size of groups that are actually dynamically stable in numerical simulations is around 10 particles. We employ, in our fiducial model, a more conservative minimum particle number of 20. This results in FOF groups of minimum mass 1.65×10^{11} $M_\odot$. We find a total of 23 000 such groups in our final simulation output 'snapshot', of which 3000 are not dynamically stable (see Section 2.5).

2.3 Properties

For each halo, we compute and record:

(i) the virial mass of the halo, M_{200} which is assumed to be equal to the group mass, i.e. the number of particles multiplied by the particle mass;

(ii) the position of the centre of mass of the halo, and its net velocity;

(iii) a, b, c, the principal axis lengths of the mass distribution;

(iv) the thermal energy of the halo – this is simply the sum of the squares of the particle velocities relative to the net velocity of the halo, multiplied by the halo mass;

(v) the potential energy – this is computationally expensive to measure, since it involves a sum over particle pairs. In Appendix B we explain and justify our approximation for measuring this quantity;

(vi) the virial radius R_{200}, which is deduced from the measured mass and potential energy, assuming a singular isothermal sphere density profile for the halo;

(vii) the circular velocity of the halo, V_{200}:

$$V_{200}^2 = GM_{200}/R_{200}; \tag{2.2}$$

(viii) the halo spin parameter, λ, defined by

$$\lambda = \frac{J|E|^{1/2}}{GM_{\rm FOF}^{5/2}}, \tag{2.3}$$

where J, E and $M_{\rm FOF}$ are the total angular momentum, energy and FOF group mass, respectively.

2.3.1 Halo spin distribution

In the top left-hand panel of Fig. 1 we show the distribution of the spin parameter, for haloes in our ΛCDM, $N_{\rm part} = 256^3$ simulation. We also plot the empirical lognormal distribution used by Mo et al. (1998), which is found to be a good approximation to the distributions obtained from several other investigations (e.g. Barnes & Efstathiou 1987; Cole & Lacey 1996). Our distribution is slightly skewed relative to this model, with a tail down to low spin parameter, but it is peaked at $\lambda \approx 0.04$, close to the 'fiducial' median value of 0.05. Our median value of λ is 0.038 for bound haloes (see Section 2.5), and the dispersion in ln λ is 0.52. This is similar to results in these other investigations.

2.3.2 Halo mass function

In the top right-hand panel of Fig. 1 we show the halo mass function, i.e. the number of haloes as a function of virial mass. We compare this measurement with two analytic predictions. The Press–Schechter approach (Press & Schechter 1974) uses a model for spherical collapse of initial perturbations to predict the final distribution of halo masses in a given cosmology. This is plotted as a short-dashed line in Fig. 1.

The peaks formalism of Bardeen et al. (1986) uses the general properties of Gaussian random fields, assuming initial peaks will be associated with future haloes. Predictions for the number density of peaks, as a function of peak height (Lacey & Silk 1991; Guiderdoni et al. 1998), thus translate to a number density of haloes as a function of virial mass. The prediction from this model is shown by the long-dashed line in Fig. 1.

It can be seen that neither of these two models provides an accurate fit over the whole range of halo masses, with Press–Schechter overpredicting the number of massive haloes, and the peaks formalism overpredicting the mass function at the light end, below 10^{12} solar masses.

2.4 Building the merger tree

In hierarchical models of structure formation, small objects form first, and then merge together into more massive ones, or accrete matter from the background. These processes can be represented with a tree, the merging tree of the dark matter haloes. Our aim is to follow the evolution of haloes detected with the group finder in the simulation outputs, and to save in a tree structure all the processes of accretion and merging between those objects, before subsequently modelling semi-analytically the evolution of baryons within these haloes. This hybrid approach has the advantage that it allow us to use all the spatial information from the simulations as inputs to our models.

Considering a halo h_1 at one time-step, we will identify a halo h_2 at the previous time-step as a progenitor if at least one particle is common to both haloes.

Having compiled a progenitor list for each halo, we compute the mass fraction of the halo coming from each progenitor. This gives the basic structure of the tree. This is a non-binary tree, in the sense that each halo can have several 'sons'. In this general case, the sum of the masses of the progenitors of a halo can be greater than the mass of the halo itself (fragmentation, evaporation, tidal stripping of a progenitor) and differs from the implementation of KCDW as each halo can have more than one descendant.

It should be noted that an alternative method has been employed by other groups (SP99; CLBF). They have created a merger tree

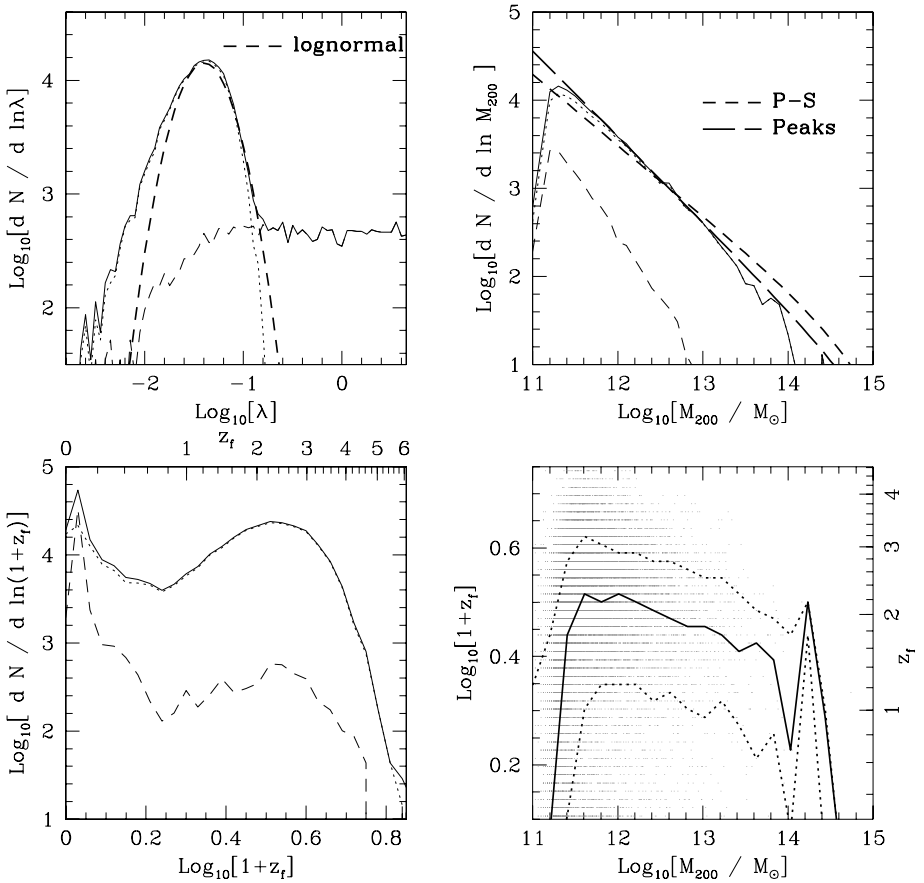

Figure 1. Distributions of several halo properties. In each plot, the solid black line is identified for all groups by the friends-of-friends method, the dotted line denotes haloes with negative binding energy (i.e. bound objects) and the dashed line denotes unbound systems. The top left-hand panel shows the dimensionless spin parameter, compared with the empirical (lognormal) function of Mo et al. (1998) (dashed line, with average $\bar{\lambda} = 0.04$, with arbitrary normalization). On the top right-hand side is the mass function (i.e. distribution of halo virial masses), compared with Press–Schechter and Peaks model predictions (short- and long-dashed lines). The bottom left-hand side shows the formation time, defined in (2.4). The bottom right-hand side shows the dependence of the formation time on the halo mass, showing the median and $\pm 1\sigma$ spread of the distribution.

using random realizations of the Press–Schechter formalism. Their approach has theoretically infinite resolution, as the merging history can always be followed back to arbitrarily small progenitors. Our method, although resolution limited, has the advantage that we have full spatial information concerning the haloes and their progenitors, and we can look at the 'genuine' history of a halo rather than a statistical realization of that history.

Having defined a list of progenitors for each halo, we can start to examine the history of how each halo acquired its present-day mass. One particularly simple statistic that has been employed to parametrize this history is the formation redshift, $z_{\rm f}$ of the halo, defined as the time at which half the mass of the halo was present in a single object (Lacey & Cole 1993). In practice this is calculated by recursively following the largest progenitor of the halo back through the tree, until it has half the present-day mass. If a halo has formed by fragmentation of another halo of mass greater than the current mass, we equate the fragmentation time with the formation time. In the bottom left-hand panel of Fig. 1 we plot the distribution of formation times, for haloes in the final ($z = 0$) output of our ΛCDM simulation.

In the bottom right-hand panel of this figure, we show the median and scatter ($\pm 1\sigma$) of the formation time as a function of halo mass, for only those haloes with negative binding energy. There is a clear trend that more massive haloes tend to have formed more recently, as expected given the results of Lacey & Cole (1993) and others. This relationship breaks down below $\approx 3 \times 10^{11}\,{\rm M}_{\odot}$, showing that, although we resolve haloes of this size, we cannot resolve their progenitors with half this mass, so the formation time becomes the time when the halo crossed the mass threshold to become detectable.

In Fig. 2 we present the merging history of a typical massive halo in the final simulation output. We show the locations and sizes of all the progenitors of the halo in the preceding time-steps. This clearly shows how virialized haloes start to form in filaments, and then fall down these structures towards the centre, feeding the central massive halo.

2.5 False halo identification

No group finder is totally accurate, as it is always possible to miss groups or to identify systems that are not in fact bound. Having determined the energy of our haloes, we find that many, especially the lighter ones, have positive total energy, i.e. they are not bound systems. Leaving these haloes in the sample can seriously skew the spin distribution away from its usual lognormal behaviour. This effect is illustrated in the top left-hand panel of Fig. 1, where we plot the distribution for all the haloes, and separately for those with positive and negative binding energies. It can be seen that excluding haloes

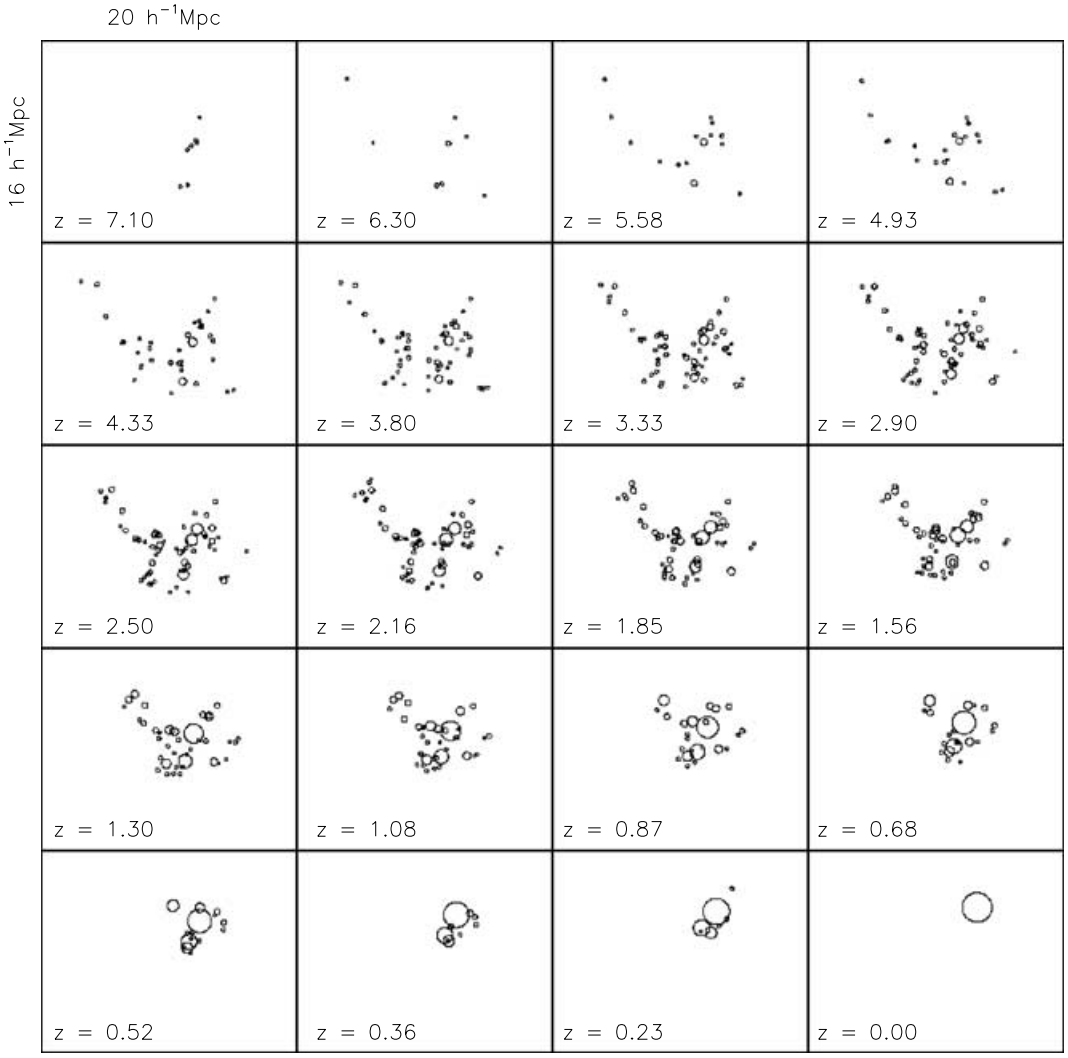

Figure 2. An example merging history for a dark matter halo with mass 1.9×10^{14} $M_{\odot}$ at a redshift of zero. We show all the progenitors of the halo as circles with radii representing their virial radii on the same scale as the axes.

with positive energy immediately leads to a huge improvement in the symmetry of the distribution for high spins.

From the other panels of this figure, we can see that the mass function of these haloes is steeper than that of 'good' haloes, and that the majority of them have extremely recent formation times. The recent formation time, in particular, suggests that the majority of haloes with positive binding energies are transient collections of particles that have been misidentified as bound systems by the group finder. This could possibly be remedied by using a different group finder (for instance SO, the spherical overdensity algorithm), or by rejecting the 'unbound' particles, which contribute most to the halo thermal energy.

However, it must be remembered that some of these haloes could be made from pairs of similarly sized objects that are currently going through a merger event and have not yet relaxed to virialization. In this case, we expect the kinetic energy to be rather higher than the virial theorem would predict, and the halo may have positive binding energy. We would not want to reject these haloes. Those objects with a longer formation time may be smaller haloes in high-density regions that appear disrupted due to strong tidal forces, probably shortly before they merge with a more massive companion. Clustering studies of these objects (Blaizot et al., in preparation) show that they are highly clustered and concentrated in the densest regions of the simulation.

2.6 Baryon tree

Although we preserve all the information concerning particles shared, and hence can reconstruct the halo merging tree in full detail, for practical purposes we will be interested in following only the bulk of the dark matter. Two-body (resolution) effects on small scales imply that the exchange of one or two particles is not a guarantee that two haloes are physically related – in a higher-resolution simulation, this exchange might not occur. For tracing the baryons and galaxies between time-steps, we will use a restricted version of the tree where we only consider merging between haloes (i.e. objects with more than 20 particles). Ideally, one would like all the mergers to be binary (as in Monte Carlo generated merging trees), because the order in which haloes merge might slightly change the properties of the final halo. In practice, due to finite time resolution, it is not possible to avoid multiple mergers between two outputs of the simulation, especially for the most massive haloes at low

redshifts. However, we find that most of our haloes do not undergo a merger between time-steps, and of the merger events that do take place, those where three or more progenitors can be identified in the preceding time-step are less numerous by a factor of 10 than the number of binary mergers (Ninin 1999). We will return to a more detailed study of time resolution effects in Section 9.2, but in the mean time we assume, for simplicity, that halo mergers always take place exactly in the middle of two time-steps.

When considering a halo, we sort the list of all its progenitors in decreasing order of contribution to its mass. For each of these haloes, we then see whether our halo is the main (i.e. most massive) son. If it is, it is assumed that the whole baryonic component (discussed in the next section) of the father is transferred to the son. In this simplified description, then, each halo has zero (it just formed) or several progenitors, and zero or one descendant. We note that this simple tree is very similar to KCDW's as their haloes are also allowed to have many progenitors but only one descendant.

3 BARYON COOLING IN HALOES

Having compiled a list of halo properties based on their dark matter content, and constructed a merger tree from the exchange of *N*-body particles, we come to the point where we use semi-analytic recipes to add a baryon component to these dark matter haloes. The key assumption is that the overall behaviour of the matter distribution is determined by the dark matter density field, and that the baryon field thus represents only a small perturbation to the total density. Secondly, it is assumed that gravity is the only force operating between haloes, so at a given time-step, each halo identified in the dark matter field can be considered completely independent from all other haloes in the simulation.

3.1 Keeping track of baryons

When a halo is first identified in the *N*-body simulation, we assign it a mass of hot gas consistent with the primordial baryon fraction of the Universe, i.e.

$$M_{\rm hot} = M_{200}\frac{\Omega_{\rm B}}{\Omega_0}, \tag{3.1}$$

where $\Omega_{\rm B}$ is a free parameter in our model, as the power spectrum used to construct the simulations has no dependence on the baryon fraction. For the results presented in this paper, we adopt a value of $\Omega_{\rm B} = 0.02\,h^{-2}$, suggested by observations of the deuterium abundance in quasi-stellar object (QSO) absorption lines (Tytler, Fan & Burles 1996).

As the halo subsequently accretes mass, we increase its stock of hot gas, assuming that the new mass accreted has the same primordial baryon fraction. If the halo loses mass, for example through evaporation of dark matter due to two-body effects, or from fragmentation (when the principal 'son' of the halo is less massive than its progenitor), we discard some of its hot gas, based on the *current* baryon fraction of the hot gas in the halo.

We will see in Section 4.2 that galaxy formation processes can also lead to the ejection of gas and metals from the halo. Whilst the physical processes are difficult to model, it seems reasonable to assume that a certain portion of this material will remain in the local environment of the halo, and may later be re-accreted. For this reason, we keep track of the mass of gas, metals and dark matter that have been lost from the halo through these processes, storing them in a halo 'reservoir'. It will be assumed that, when the halo subsequently accretes dark matter from the background, some of this dark matter will be 'primordial' (i.e. with baryon fraction $\Omega_{\rm B}$ and zero metallicity), and the rest will have the same baryon fraction and metal content as that of the halo reservoir, up to the point where the reservoir is fully used up. We parametrize this effect with the halo recycling efficiency, ζ, where $\zeta = 0$ implies permanent loss of ejected material, and $\zeta = 1$ is maximally efficient re-accretion, where all the reservoir is used up before the accreted dark matter is assumed to be primordial.

There have been previous attempts to deal with this issue. Benson et al. (2000) and KCDW both compare the standard model of CLBF in which gas is re-accreted after the parent halo has doubled in mass, with a model in which re-accretion is immediate (enriched gas can cool back on to the disc immediately after it has been ejected). Both of these models lack strong physical or observational support. SP99 adopt the extreme case where ejected material is forever lost from the halo. This corresponds to our model with $\zeta = 0$. For the rest of this paper, we will adopt a standard model of $\zeta = 0.3$.

3.2 Cooling the hot gas

At each time-step we allow the hot intrahalo medium to cool on to a disc at the centre of each halo. As described in Section 2.5, there are some haloes that have positive binding energy, either due to false identification by the group-finding algorithm, or extreme tidal disruption. We do not allow cooling in these haloes, or in haloes with spin parameter $\lambda > 0.5$, the threshold for complete rotational support.

The characteristic cooling time for the hot gas, as a function of radius, is given by

$$t_{\rm c}(r) = \frac{3}{2}\frac{\mu m_{\rm p} kT(r)}{\rho_{\rm hot}(r)\Lambda(T)}, \tag{3.2}$$

where $\mu m_{\rm p}$ is the molecular mass of the gas, $\rho_{\rm hot}$ is the hot gas mass density and $\Lambda(T)$ is the cooling function, which depends on the metallicity and temperature of the gas. We use the model of Sutherland & Dopita (1993) to find Λ for our haloes. We will make the assumption that the gas has the same temperature throughout, which is known to be a good approximation for clusters, at least outside the central cooling flow region (e.g. David, Jones & Forman 1996). This isothermal approximation, applied to the equation of hydrostatic equilibrium, results in a halo temperature (in kelvin)

$$T = \frac{\mu m_{\rm p}}{2k}V_{\rm c}^2 = 35.9\left(\frac{V_{\rm c}}{\rm kms^{-1}}\right)^2. \tag{3.3}$$

This latter result comes from fixing μ by assuming the gas is totally ionized, consisting of one part helium to three parts hydrogen (by mass). We will also assume that the dark matter distribution follows an isothermal profile, in which case $V_{\rm c}$, the halo circular velocity, is identical to V_{200}, defined in Section 2.3.

Given the cooling rate, we can calculate the mass of gas cooled during a time interval Δt:

$$m_{\rm cool} = \int_0^{r_{\rm max}} \rho_{\rm hot}(r)\,{\rm d}^3 r. \tag{3.4}$$

In equation (3.4), $r_{\rm max}$ is the minimum between the infall radius $r_{\rm inf}$, the cooling radius $r_{\rm cool}$ (defined, respectively, as the radii beyond which the gas has not had enough time to fall/cool on to the central galaxy in Δt), and the virial radius R_{200}. We now need a model for the distribution of the hot gas in the halo. Observationally, measurements of profiles of X-ray gas are often translated into a β-model (Cavaliere & Fusco-Femiano 1976), where the density follows

$$\rho_{\rm hot}(t) = \rho_0(t)\left(1+\frac{r^2}{r_{\rm c}^2}\right)^{-3\beta/2}. \quad (3.5)$$

Several numerical studies (Eke, Navarro & Frenk 1998; Navarro, Frenk & White 1995b) suggest that clusters are well fitted by a profile with $\beta = \frac{2}{3}$, i.e. a non-singular isothermal sphere with core radius $r_{\rm c}$. Here, we assume that the hot gas is in hydrostatic equilibrium within the virial radius of our isothermal dark matter haloes, which yields a hot gas distribution very similar to that of the β-model. The central density ρ_0 of the gas is computed at each time-step by integrating over the density profile given by the hydrostatic solution and equating the result to the current mass of hot gas in the halo (it is therefore time dependent). We can also define a 'time-integrated cooling radius', $r_{\rm cool}^{\rm int}$, as the radius that would contain the current cold mass of gas, if all the gas (hot and cold) were distributed according to this profile. This radius is close to that used in purely semi-analytic models to compute the quantity of gas which has cooled since a halo has formed.

3.3 Overcooling

A key problem with semi-analytic models is the overprediction of the bright end of the galaxy luminosity function. In the hierarchical framework, this can be interpreted as meaning that the central galaxies in massive dark matter haloes are generally too luminous, either because they have too much 'fuel' for star formation, or because the merger rate in the halo is too high. A number of methods have been employed or suggested to remedy this situation.

(i) The approach of KCDW has been to prevent gas from cooling altogether in haloes with circular velocity greater than a certain limiting value. Unfortunately, this very simple assumption, introduced to account for the observational fact that disc galaxies with rotational velocities in excess of 350 km s^{-1} are extremely rare, has no physical justification.

(ii) CLBF point out that large haloes are formed from small haloes that have already undergone cooling, and thus they will lack the low entropy gas that cools most easily, so one should expect the hot gas profile to be rather broader than that of small haloes, and cooling will be suppressed. Although this model has some theoretical basis, there seems to be no observational evidence that large clusters have flatter gas density profiles.

(iii) Nulsen & Fabian (1995) point out that if the cooling time of some gas is longer than its freefall time to the centre of the halo, a cooling flow may occur, in which case the gas may remain in a compact cold cloud without forming stars, or it may form stars with a non-standard initial mass function, producing an excess of brown dwarfs that will not be visible.

(iv) Wu, Fabian & Nulsen (2000) show that the energy in the hot X-ray gas is such that there needs to be significant non-gravitational (feedback, AGN) heating of this gas. In other words, if we do not include these heat sources we may be significantly underestimating the temperature from which our hot gas must cool. However, it seems that the importance of this effect is likely to diminish with the mass of the halo, so this may not in itself solve the problem.

(v) Competitive cooling, whereby satellite galaxies can accrete cold gas and the central galaxy in a halo, is generally neglected in semi-analytic models, and would again help to prevent the assembly of overluminous galaxies in the centres of haloes. However, numerical simulations of satellites moving in a hot intracluster medium (ICM) strongly suggest that RAM pressure stripping is the dominant effect (e.g. Quilis, Moore & Bower 2000).

(vi) SP99 point out that mergers of haloes may shock heat the gas within, again resulting in the suppression of cooling on to the central galaxy.

The failure to reach a consensus between these different approaches, and the absence of one well-motivated, observationally and theoretically justified model, must regrettably be regarded as one of the failings of the semi-analytic treatment, even though it probably finds its roots in a poor theoretical understanding of cooling flows. Therefore, in what follows, we simply decide to take advantage of the observed correlation between AGN and bulge mass (Magorrian et al. 1998), and prevent gas from cooling in a halo when the bulge mass locked in its host galaxies becomes equal to 10^{11} M$_\odot$.

We defer a more detailed modelling and investigation of the observational consequences of some of the previously mentioned models to future work, as this is clearly beyond the scope of the present paper.

4 GALAXIES

As hot gas cools and falls to the centre of its dark matter halo, it settles in a rotationally supported disc. If the specific angular momentum of the accreted gas is conserved and starts off with the specific angular momentum of the dark matter halo (Mo et al. 1998), we assume it forms an exponential disc with scalelength $r_{\rm D}$ given by

$$r_{\rm D} = \frac{\lambda}{\sqrt{2}} R_{200}. \quad (4.1)$$

As gas builds up over a period of time to form a massive central disc we recompute its scalelength at each time-step, by taking the mass-weighted average gas profile of the disc and that of the new matter arriving, the scalelength of which is also given by (4.1). However, the half-mass radius of the disc (see the definition below) is never allowed to exceed the virial radius of the halo, or to become smaller than the half-mass radius of the bulge, provided a bulge is present. We do not recompute the disc scalelength of galaxies that become satellites. We also obtain the circular velocity of the disc, $V_{\rm c}$ by enforcing rotational equilibrium at the disc half-mass radius, including all the mass enclosed within, i.e. baryons and a dark matter core. We assume that the DM core is depleted and/or flattened by a rapidly rotating disc as we would otherwise overestimate the amount of dark matter present within Milky Way (MW) discs (Binney & Evans 2001). More specifically the fraction of dark matter within the half-mass radius of a pure disc $f_{\rm DM}$ is taken to be

$$f_{\rm DM} = 1 - \left(\frac{V_{\rm c}}{500\,{\rm km\,s^{-1}}}\right)^2, \quad (4.2)$$

so that a typical Milky Way disc only has 80 per cent of its DM core left. This correction should be considered as a first attempt to model the complex issue of interaction between dark matter and baryons, which we plan to address in more detail in future work. Finally, we point out that this derivation of disc velocity differs from that of CLBF as these authors use a NFW profile for the density of the dark matter halo and include fully self-consistent disc gravity.

We will be concerned with a number of properties of galaxies, including their sizes, and we will generally use the radius containing half the galaxy mass to define a size, since this is more physically relevant than the scalelength. For an exponential disc, the half-mass radius is given by

$$r_{1/2} = 1.68 r_{\rm D}. \quad (4.3)$$

Galaxies remain pure discs if their disc is globally stable (i.e. $V_c < 0.7\,V_{tot}$, where V_{tot} is the circular velocity of the disc–bulge–halo system; see, e.g., van den Bosch 1998), and they do not undergo a merger with another galaxy. In the case where the latter of these two events occurs, we employ a recipe which will be described in Section 5 to distribute the stars and gas in the galaxy between three components in the resulting, post-merger galaxy, i.e. the disc, the bulge and a starburst. In the case of a disc instability, we simply transfer the mass of gas and stars necessary to make the disc stable to the burst component, and compute the properties of the bulge/burst in a similar fashion to that described in Section 5 for galaxy mergers. Bulges are assumed to have a density profile given by the Hernquist (1990) model,

$$\rho(r) = \frac{M}{2\pi}\frac{r_B}{r(r+r_B)^3}. \tag{4.4}$$

The half-mass radius is given by

$$r_{1/2} = (1+\sqrt{2})r_B. \tag{4.5}$$

The bulges are assumed to be pressure supported with a characteristic velocity dispersion σ, computed at their half-mass radius. The material forming a starburst will be assumed to concentrate in a dense nucleus at the centre of the bulge. We assume that its geometry is also modelled by Hernquist's equation, but the scale radius, r_S, is given by

$$r_S = \kappa r_B, \tag{4.6}$$

where κ is a free parameter to which we give the fiducial value 0.1. Note that gas does not cool on to bulges or starbursts: after a merger, a new disc is formed around the bulge by the cooling gas.

Given a passively evolving galaxy, there are four processes that determine its make-up in terms of gas, stars, and metals, these being cooling, star formation, feedback and metallicity. Although the metallicity of a galaxy varies more slowly, as it takes time for stars to process hydrogen into metals and release it back into the gas content, the other three can vary on quite short time-scales, and are highly coupled. For this reason, we use an adaptive time resolution for calculating all quantities, breaking each simulation time-step (or each interval between mergers, whichever is shorter) into multiple substeps. Essentially, this fine resolution is a brute-force numerical way of solving the differential equations that relate cooling, star formation and feedback.

We now describe the interlinked processes responsible for making galaxies.

4.1 Star formation

At each time-step, a certain amount of the cold gas in a galaxy is allowed to form stars. The rate of this star formation is given by

$$\Psi_\star = \frac{M_{cold}}{\beta t_{dyn}}, \tag{4.7}$$

where t_{dyn} is the dynamical time-scale of the galaxy, for which we use the time taken for material at the half-mass radius to reach either the opposite side of the galaxy or its centre for a disc and a bulge, respectively, and is given by

$$t_{dyn} = r_{1/2} \times \begin{cases} \pi V_c^{-1} & \text{for discs} \\ \sigma^{-1} & \text{for bulges,} \end{cases} \tag{4.8}$$

where V_c is the circular velocity of the disc, the material is assumed to have purely circular orbits and σ is the bulge velocity dispersion, where we assume the matter in the bulge has only radial orbits. β^{-1} is the star formation efficiency. Based on data for a sample of bright galaxies (Kennicutt, Tamblyn & Congdon 1994) it has been shown by Guiderdoni et al. (1998) that a value $\beta \approx 50$ is capable of reproducing the mean Roberts times (Roberts 1963) for discs. We therefore use $\beta = 50$ as the fiducial value of this parameter. One can wonder why we define a star formation rate for the bulge as gas cannot cool on this component and stars form in the burst component during a merger. This is simply because stars eject gas and metals back into the ISM, which can be trapped in the bulge triggering star formation there.

As mentioned earlier, starbursts have the same geometry as bulges, so their star formation law is the same. The only difference is that the radius of the burst is smaller, the characteristic time-scale for star formation is much shorter, giving birth to an almost instantaneous 'burst' of star formation.

4.2 Feedback

For every mass of stars formed, a certain fraction will be contained in massive, fast-evolving stars that quickly become supernovae (SNe). When a supernova occurs, some of the energy released will be imparted to the ISM, 'blowing off' some of this gas. If this ejected gas has high enough kinetic energy, it may even be blown out of the host halo altogether. This process thus inhibits future star formation, and hence is termed feedback.

The feedback is given by (Silk 2003)

$$\dot{m} = 2\Psi_\star \frac{\epsilon_w \eta_{SN} E_{SN}}{v_{esc}^2}, \tag{4.9}$$

where ϵ_w is the efficiency of the supernova-triggered wind, which is proportional to v_{esc}^2 and depends both on the porosity of the ISM (see Silk 2001 for details) and the mass-loading factor. The latter accounts for entrainment of interstellar gas by the wind and can be considered as a free parameter where the value is around 10 (Martin, Kobulnicky & Heckman 2002). This is modelled by our efficiency parameter ϵ, and as a result, the outflow rate given by equation (4.9) for a starburst is of the order of the star formation rate. Note that in the previous equation, η_{SN} is the number of supernovae per unit star-forming mass, which is a prediction of the IMF chosen and E_{SN} is the energy of a supernova, assumed to be 10^{51} erg.

Equation (4.9) is applied to find the fraction of gas in the ISM that is lost by the galaxy and ejected into the intrahalo medium. We then equate the fraction of this gas that is completely ejected from the halo, to the galaxy/halo escape velocity ratio. The gas ejected from the halo is added to the halo reservoir where it may subsequently be accreted, as discussed in Section 3.1.

The escape velocity is given by

$$v_{esc}^2 = 2v^2 \times \begin{cases} 1+\ln 2 & \text{for discs} \\ 2 & \text{for bulges} \\ 1+\ln(R_{200}/r) & \text{for haloes,} \end{cases} \tag{4.10}$$

where v is the characteristic velocity (i.e. the circular velocity for discs and haloes, the dispersion velocity for bulges), and, in the case of halo ejection, r is either the orbital radius of the galaxy in the halo, or, for the central galaxy in the halo, the half-mass radius of the galaxy.

4.3 Metallicity

The baryonic gas in haloes and galaxies, initially composed solely of hydrogen and helium, acquires a metal content due to the processing

of these light elements by the stellar population, and the subsequent release of the heavy elements synthesized to the interstellar medium at the end of the life of a star. To track the metallicity, we need two components: a model for the rate at which metals are produced inside the stars and a model for the amount of material released by a given stellar population.

Using these models, we are able to use the tree to track the metals ejected by stars at each time-step, over the whole merging history of the galaxy, and let new stars form out of the enriched gas, assuming instantaneous mixing in the ISM. This is in contrast to the assumption of instantaneous recycling, the approach generally used in semi-analytic modelling (CLBF; KCDW; SP99). Correct modelling of the ejection is likely to be especially important for elements such as iron, nitrogen and possibly carbon, for which the production delay can be significant (e.g. 1 Gyr for Fe produced by SN Ia).

Given a mass M of stars formed at some time t_0, we can calculate the current contribution to the stellar ejection during a time-step from t_1 to t_2 as

$$\Delta M_\star = -M \int_{m(t_1)}^{m(t_2)} [m - w(m)]\phi(m)\,\mathrm{d}m, \tag{4.11}$$

where $m(t)$ is the mass of a star having lifetime t, $w(m)$ is the mass of the remnant left after the star has died and $\phi(m)$ is the IMF. Our fiducial model uses a Kennicutt IMF, and in Paper II we will explore the effects of using a Scalo or Salpeter IMF (see Salpeter 1955; Kennicutt 1983; Scalo 1986). For all IMFs considered, we form no stars lighter than 0.1 $\mathrm{M}_\odot$ or heavier than 120 $\mathrm{M}_\odot$.

This formula can be extended to include the entire star formation history, producing a rate of increase in the gas mass due to ejection from the stellar population,

$$\mathcal{E}(t) = \int_{m(t)}^{\infty} \Psi_\star(t - t_m)[m - w(m)]\phi(m)\,\mathrm{d}m, \tag{4.12}$$

where t_m is the lifetime of a star of mass m [i.e. the inverse of $m(t)$].

Equation (4.12) can be adapted to predict the amount of metals produced, by making the replacement:

$$[m - w(m)] \longrightarrow [m - w(m)]Z_{\mathrm{cold}}(t - t_m) + mY_Z(m), \tag{4.13}$$

where the first term on the right-hand side represents the reintroduction of the metals that were originally in the stars when they formed, and $Y_Z(m)$ is the fraction of the initial stellar mass transformed via stellar nucleosynthesis into metals, known as the stellar yield. Our models for the remnant mass and the yield come from the STARDUST code of DGS, and are calculated self-consistently with the models for spectral evolution we will discuss in Section 6.

Throughout this work, we assume chemical homogeneity (instantaneous mixing), such that outflows caused by feedback processes are assumed to have the same metallicity as the interstellar medium, though in reality the material in the outflow could be metal-enhanced (Pagel 1998).

5 MERGING

In the hierarchical picture of structure formation, mergers are clearly crucial in understanding the properties of galaxies observed at the present day. In our models, mergers are responsible for the formation of massive spheroids, and are assumed to trigger starburst activity. The detailed modelling of the physics of individual mergers between galaxies is currently beyond the scope of what can be achieved in a cosmological simulation. Only the highest-resolution N-body simulations of Springel et al. (2001) are able to dynamically follow the merger of $0.2L_\star$ dark matter subhalos within a single cluster. Therefore, the best we can hope for is to approximately model the rate and effects of merging in a global sense. To deal with the rate of merging, we model two effects, the gradual tendency of satellite galaxies to lose orbital energy and sink towards the centre of the cluster potential well, and the likelihood of occasional encounters between satellite galaxies in a cluster, based on probabilistic, cross-section arguments.

5.1 Halo mergers

We identify mergers between dark matter haloes using the halo tree, as described in Section 2.4. When a merger occurs, the properties of the dark matter itself are obtained directly from the properties of the new halo in the N-body simulation. We apply recipes, however, to deal with the baryonic component. First, the ICM, i.e. the hot cluster gas, of the progenitors are added together proportionally to the dark matter mass that ends up in the descendant halo (the constant of proportionality being the current baryon fraction of each progenitor), and given the virial temperature of the new halo. Then, the properties of the two halo reservoirs are simply added. Note that the new halo contains all the galaxies that were present in its progenitors (even if fragmentation occurs and the halo is less massive than all/some of its progenitors) and that the fraction of the ICM of the progenitors, which does not end up in the new halo is put in its reservoir, ensuring conservation of metals.

When two haloes merge (sometime in between two output times), there is a discontinuity between the centre of the new halo and the centres of mass of the progenitors (linearly extrapolated from their positions and velocities in the previous time-step). We measure this 'jumping' distance, R_j, for each of the progenitors directly from the N-body simulations and use it to assign for each galaxy an orbital radius in the new halo, using the cosine law:

$$r_{\mathrm{new}} = \sqrt{r_{\mathrm{old}}^2 + R_\mathrm{j}^2 - 2r_{\mathrm{old}}R_\mathrm{j}\cos\theta}, \tag{5.1}$$

where r_{old} is the orbital radius of the galaxy in the progenitor halo and r_{new} is its orbital radius in the new halo. This assumes that the galaxy was initially at an angle θ to the vector joining the centres of the two haloes. We select $\cos\theta$ randomly from a flat probability distribution between -1 and 1, taking account of the spherical symmetry.

To illustrate how this works in practice, we consider two cases. The first is a merger between a small halo and a much larger halo. It is clear that the centre of mass of the large halo is perturbed only slightly by the encounter. R_j is thus small, and the galaxies that were previously in the large halo have $r_{\mathrm{new}} \approx r_{\mathrm{old}}$. For the small halo, R_j is close to the virial radius of the large halo. Thus $r_{\mathrm{new}} \approx R_\mathrm{j}$ for the galaxies that were in this halo, in other words they are placed close to the virial radius of the new halo. For a collision between equal-mass haloes, R_j is approximately the progenitor virial radius, so the central galaxies are placed at this distance, while non-central ones are placed randomly throughout the new halo.

Any galaxy where the orbital radius after the merger is less than its own half-mass radius becomes the new central galaxy of the halo. If there is more than one of these objects, they are merged with each other in order of descending mass. It may well happen that there is no central galaxy after a merger, in which case we cool gas on to the galaxy closest to the halo centre. Although this scheme is really only a naive geometrical model, we use it to approximately replicate the sort of scatter in galaxy positions we expect to see when haloes collide. It has the advantage that it reproduces the desired behaviour for the extreme cases of equal or very unequal mass mergers in a natural, continuous way without free parameters.

5.2 Dynamical friction

Dynamical friction causes satellite galaxies to sink gradually to the centre of their host halo, resulting in a merger if there is already a galaxy at the centre. After a merger between haloes, the galaxies are given orbits as prescribed in Section 5.1. In subsequent time-steps, the radii of the orbits are decreased by an amount obtained from the differential equation (Binney & Tremaine 1987):

$$r\frac{\mathrm{d}r}{\mathrm{d}t} = -0.428\frac{Gm_{\rm sat}}{V_{\rm c}}\ln\Lambda, \tag{5.2}$$

where r is the orbital radius, $m_{\rm sat}$ is the mass of the satellite (including its tidally stripped dark matter core, see the next section), $V_{\rm c}$ is the circular velocity of the halo and ln Λ is the Coulomb logarithm, approximated by

$$\Lambda = 1 + \left(\frac{m_{\rm halo}}{m_{\rm sat}}\right) \tag{5.3}$$

(see, e.g., Mamon 1995). Calculating the amount of orbital decay at each time-step seems far more natural than the approach favoured by other authors (CLBF, KCDW) who assign each galaxy a dynamical friction time-scale, and assume the merger occurs after that fixed time, unless a major merger occurs, at which point this time-scale is recomputed. By modelling the true radial coordinate, albeit in a naive way, we automatically take into account possible evolution of the halo properties, and avoid having to differentiate between 'major' and 'minor' mergers.

As soon as the orbital distance of a galaxy becomes lower than the sum of its half-mass radius and the half-mass radius of the central galaxy in the halo, it is assumed to merge with this central galaxy.

Other workers (CLBF; KCDW; SP99) have taken into account the fact that galaxy orbits are not in general circular. This fact alters the extent of dynamical friction, and Navarro, Frenk & White (1995a) have shown that applying the formula for circular orbits results in an average overprediction by a factor of 2 of the time taken for an infalling galaxy to reach the centre. A proper consideration of this effect requires a modified form of the dynamical friction as a function of orbit eccentricity (Lacey & Cole 1993), and an assumption for the distribution of ellipticities. However, careful examination of Navarro et al. (1995a) (especially their fig. 8), shows that one can obtain the true dynamical friction time more accurately by simply halving the time for circular orbits, than by using the actual orbital eccentricity and applying the modified form of Lacey & Cole (1993). We thus take into account the effect of non-circular orbits by multiplying the right-hand side of equation (5.2) by a factor of 2.

5.3 Tidal stripping

When a merger between haloes occurs, the new satellite galaxies will, to some extent, retain the dark matter cores of their previous host haloes, and this dark matter will continue to dominate their dynamics. However, as the galaxy slowly descends to the centre of the new host this subhalo will be stripped by tidal interactions. Eventually, as the mass of the remnant becomes comparable to the baryonic mass of the galaxy itself, it will no longer dominate over the effects of the self-gravity of the galaxy, and the dynamics may undergo a serious change. Following SP99, we assume that the subhalo is stripped down to a radius $r_{\rm t}$ such that $\rho_{\rm c}(r_{\rm t}) = \rho_{\rm h}(r_0)$, where $\rho_{\rm c}$ is the core density and $\rho_{\rm h}(r_0)$ is the halo density at the orbital radius of the remnant (for isothermal spheres this density is simply proportional to the mean density inside the same radius). Note that we do not attempt to model the stripping of baryons by tidal effects, only that of the dark matter.

5.4 Satellite–satellite mergers

Satellite mergers are assumed to occur through direct collisions between galaxies in clusters. In the absence of a detailed numerical simulation of the precise positions and velocities of the galaxies within a halo, it is appropriate that some sort of cross-section argument be used to determine the merging rate. There is no detailed study of how such cross-sections behave in a realistic halo environment; a first analytical estimate by Mamon (1992) has been followed numerically by Makino & Hut (1997), with Mamon (2000) showing that these results are in good agreement. We assume the probability of a galaxy having a merger in time Δt is given by

$$P = \frac{\Delta t}{\tau}, \tag{5.4}$$

where τ is the merger time-scale, the dependence of which on galaxy and halo characteristics is parametrized by Makino & Hut as

$$\tau^{-1} = \psi(N_{\rm g} - 1)\left(\frac{r_{1/2}}{R_{200}}\right)^3\left(\frac{v_{\rm gal}}{V_{\rm c}}\right)^3\left(\frac{v_{\rm gal}}{r_{1/2}}\right). \tag{5.5}$$

In this formula $N_{\rm g}$ is the number of galaxies in the halo and $v_{\rm gal}$ is taken to be a mass-weighted average of the disc rotation speed and the bulge dispersion velocity. A direct transformation of the results of Makino & Hut (1997) leads to a value of $\psi = 0.017$ s Gyr^{-1} km^{-1} Mpc, which is the value we shall take for our fiducial model. However, their work assumes all galaxies are spheroids of the same size and mass, which is not the case in our models, and in Paper II we will investigate the effect of altering ψ from this standard value. We are aware that our extrapolation of these authors' results is quite crude and should be, in the best of cases, taken as a rough estimate of the importance of satellite–satellite mergers. Finally, we note that the results obtained by Springel et al. (2001) for subhalo merging in dark matter clusters support the view that this kind of merger is negligible in such environments.

5.5 Post-merger morphology

In our model, mergers and disc instabilities are both responsible for creating galaxy bulges, and as such are drivers of morphological evolution. In the case of a disc instability, the mass transfer between galaxy components has already been discussed in Section 4, and therefore we will only be concerned with mergers in this section. The standard method for modelling a merger-driven morphology (SP99; CLBF; KCDW) has been to take the ratio of the masses of the two objects merging, and add the stars of the lighter galaxy to the disc of the heavier one if the mass ratio is less than $f_{\rm bulge} \approx 0.3$, or destroy the disc and form a bulge if the ratio is higher.

It has been shown (Walker, Mihos & Hernquist 1996) that a galactic disc can be completely disrupted by an encounter even when the interloper is rather less massive than the disc itself, and this simple method reproduces this behaviour. However, it does not allow for any intermediate behaviour – there is a sharp cut-off at $f_{\rm bulge}$ between totally disrupting, or not disrupting, the morphology of the galaxy, which seems quite false.

Our model is constructed in terms of the function X. In general, this can be any function of the ratio of progenitor masses that maps the range one to infinity on to the space 0 to 1. The step function used be other workers is one example of such a function, but we

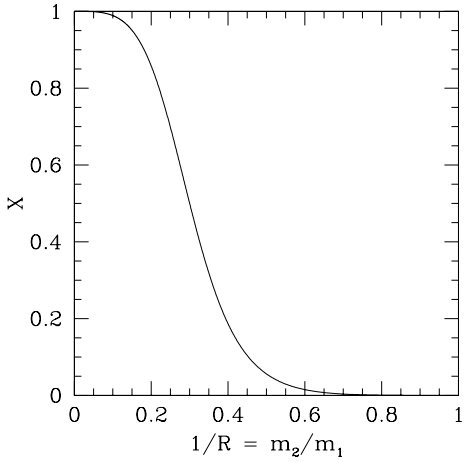

Figure 3. Our algorithm for calculating the fraction of disc material remaining in the disc after a merger, as a function of the mass ratio of the two progenitors.

will consider a smoother function, drawing inspiration from Fermi–Dirac statistics, and defining

$$X(R) = \left[1 + \left(\frac{\chi - 1}{R - 1}\right)^{\chi}\right]^{-1}. \tag{5.6}$$

This function is shown in Fig. 3, for the value $\chi = 3.333$. In equation (5.6), R represents the mass ratio of the heavier to the lighter progenitor and χ is the critical value of that ratio, i.e. the value that gives X = 0.5. Since step functions with $f_{\rm bulge} \approx 0.3$ have been found to give good results in the past, we set $\chi = 1/0.3$ as our fiducial value. In Paper II we will explore the effects of varying χ.

We then define two matrices,

$$\mathbf{A}_{\rm star} = \begin{pmatrix} X & 0 & 0 \\ 0 & 1 & 0 \\ 1 - X & 0 & 1 \end{pmatrix} \tag{5.7}$$

$$\mathbf{A}_{\rm gas} = \begin{pmatrix} X & 0 & 0 \\ 0 & 0 & 0 \\ 1 - X & 1 & 1 \end{pmatrix}. \tag{5.8}$$

Considering the vector $\boldsymbol{V} = (D, B, S)$, where D, B, S refer to the mass in the disc, bulge and starburst, respectively, we have:

$$\boldsymbol{V}^{\rm new}_{\rm gas} = \mathbf{A}_{\rm gas}\boldsymbol{V}^{1}_{\rm gas} + \mathbf{A}_{\rm gas}\boldsymbol{V}^{2}_{\rm gas} \tag{5.9}$$

and similarly for the stars.

In other words, during a merger a fraction X of gas and stars originally sitting in the disc remain in the disc. The rest of the gas from the disc goes into the starburst, and its stars. Any stars that were already in the bulge stay in the bulge, but all of its gas is put in the starburst. All the material (gas and stars) that was originally in a starburst remains in that starburst. Note that gas is never added to the bulge in this process, but that a small amount of gas is generally present in the bulges, coming from the evolution of its stellar content.

5.6 Properties of merger remnant

Having ascertained how much mass goes where, we must decide on the properties of the resulting galaxy after the collision, i.e. its rotation/dispersion speed and size. For this, we adopt a similar model to that of CLBF.

Under the virial theorem, the total internal energy is given by $E = -T$, where T is the internal kinetic energy. Applying conservation of energy,

$$T_{\rm new} = T_1 + T_2 - E_{\rm int}, \tag{5.10}$$

where $E_{\rm int}$ is the interaction energy of the collision. For a dynamical friction encounter as well for satellite–satellite mergers, we use the total orbital energy,

$$E_{\rm int} = -\frac{Gm_1m_2}{r_1 + r_2} \tag{5.11}$$

(r_1 and r_2 are the half-mass radii of the two progenitors). Note that m_1 and m_2 denote all the mass (dark matter core and baryons) contained in each progenitor, respectively.

These energy considerations, coupled to mass conservation enable us to obtain $r_{\rm new}$, the half-mass radius of the merger product:

$$r_{\rm new} = \frac{(m_1 + m_2)^2}{-E_{\rm int}/0.4G + m_1^2/r_1 + m_2^2/r_2}, \tag{5.12}$$

where the factor of 0.4 depends (not sensitively) on the exact density profile (Hernquist in our case). As for the pure disc case, we do not allow $r_{\rm new}$ to become bigger than R_{200}. We then assume that $r_{\rm new}$ is the half-mass radius of the bulge and hypothesize that the largest disc is the most likely one to survive the merger. Following the method presented in Section 4, we can calculate the characteristic velocities of our galaxy components. In particular, we use equation (4.6) to calculate the radius of the starburst after merging and assume that the starburst has the same velocity dispersion as the bulge.

In the case of a satellite–satellite merger, the orbital radius of the merger remnant is obtained through a simple mass weighting of the orbital radii of the two progenitors.

6 GALAXY LUMINOSITIES

Having decided where the baryons are, and what their physical properties are, we apply a set of models to calculate the amount of light they produce. Luminosities in visual and near-infrared (rest-frame) wavelengths are calculated from stellar synthesis models, which take into account the metallicity and age of the stellar population. We then add geometry- and metallicity-dependent models for the absorption and re-emission of this light by the dust and gas in the interstellar medium.

6.1 Stellar spectra

For each time output of the simulation, we compute the stellar contribution $F^{\star}_{\lambda}(t)$ to the galactic flux at time t, which can be written as

$$F^{\star}_{\lambda}(t) = \int_0^t \int_{m=0}^{\infty} \Psi_{\star}(t - \tau)\phi(m) f_{\lambda}(m, \tau, Z_{\star})\, {\rm d}m\, {\rm d}\tau, \tag{6.1}$$

where $f_{\lambda}(m, \tau, Z_{\star})$ is the flux at wavelength λ of a star with initial mass m, initial metallicity $Z_{\star}$ and age τ [i.e. $\tau = 0$ corresponds to the zero-age main sequence, and $f_{\lambda}(m, \tau, Z_{\star}) = 0$ if $\tau > t(m)$]. $\phi(m)$ is once again the IMF. We neglect the nebular component. For $f_{\lambda}(m, \tau, Z_{\star})$ we use the STARDUST model of DGS, and full details can be found in that paper.

We use the full merging history of the galaxy to compute the stellar spectrum, summing up the contribution to the present-day spectrum from all the stars formed in the previous time-steps in all the progenitors of the galaxy.

6.2 Dust absorption

To estimate the stellar flux absorbed by the interstellar medium in a galaxy, one first needs to compute its optical depth. As in Guiderdoni & Rocca-Volmerange (1987), we assume that the mean perpendicular optical depth of a galaxy at wavelength λ is

$$\tau_\lambda^z = \left(\frac{A_\lambda}{A_V}\right)_{Z_\odot} \left(\frac{Z_g}{Z_\odot}\right)^s \left(\frac{\langle N_H\rangle}{2.1\times10^{21}\,\mathrm{atom\,cm^{-2}}}\right), \tag{6.2}$$

where the mean H column density (accounting for the presence of helium) is given by

$$\langle N_H\rangle = \frac{M_{\rm cold}}{1.4 m_p \pi (a r_{1/2})^2}\,\mathrm{atom\,cm^{-2}} \tag{6.3}$$

and a is calculated such that the column density represents the average (mass-weighted) column density of the component, and is 1.68 for discs, 1.02 for bulges and starbursts. The extinction curve depends on the gas metallicity Z_g according to power-law interpolations based on the solar neighbourhood and the Large and Small Magellanic Clouds, with $s = 1.35$ for $\lambda < 2000$ Å and $s = 1.6$ for $\lambda > 2000$ Å (see Guiderdoni & Rocca-Volmerange 1987 for details). The extinction curve for solar metallicity $(A_\lambda/A_V)_{Z_\odot}$ is taken from Mathis, Mezger & Panagia (1983).

For the spherical components we use the generalization given by Lucy et al. (1989) for the analytic formula giving obscuration as a function of optical depth $\tau_\lambda^{\rm sph}$ (Osterbrock 1989) to the case where scattering is taken into account via the dust albedo, ω_λ (we use the model of Draine & Lee 1984 for the albedo):

$$A_\lambda(\tau) = -2.5\log_{10}\left(\frac{a_\lambda}{1-\omega_\lambda+\omega_\lambda a_\lambda}\right), \tag{6.4}$$

where

$$a_\lambda(\tau) = \frac{3}{4\tau_\lambda}\left[1-\frac{1}{2\tau_\lambda^2}+\left(\frac{1}{\tau_\lambda}+\frac{1}{2\tau_\lambda^2}\right)\exp(-2\tau_\lambda)\right]. \tag{6.5}$$

We assume dust and stars are distributed together uniformly and thus there is no need to average over inclination angle, since the bulge is spherically symmetric.

For discs, the situation is more involved, due to inclination effects. We use a uniform slab model (Guiderdoni & Rocca-Volmerange 1987) for the extinction as a function of the inclination angle, θ:

$$A_\lambda(\tau,\theta) = -2.5\log_{10}\left[\frac{1-\exp(-a_\lambda\sec\theta)}{a_\lambda\sec\theta}\right], \tag{6.6}$$

where $a_\lambda = \sqrt{1-\omega_\lambda}\,\tau_\lambda^z$. To calculate this quantity, we pick a random inclination of the galactic disc to the line of sight of the observer. For comparisons of the same galaxy at different epochs, and for examining inclination-corrected statistics (e.g. the Tully–Fisher relation), we also compute and store the face-on magnitudes of the discs ($\sec\theta = 1$).

Our final result is the extinguished stellar spectrum:

$$F_\lambda = F_\lambda^\star \mathrm{dex}(-0.4A_\lambda). \tag{6.7}$$

6.3 Dust emission

Dust absorption in bulges is spherically symmetric, so the emission is trivially calculated, but for discs it is anisotropic, so to calculate the total energy budget available for infrared emission by dust one must include contributions from all directions. We use the angle-averaged version of equation (6.6),

$$\bar{A}_\lambda(\tau) = -2.5\log_{10}\left[\int_0^1 \frac{1-\exp(-a_\lambda/x)}{a_\lambda/x}\,\mathrm{d}x\right]. \tag{6.8}$$

The value of the integral on the right-hand side of this equation is

$$\frac{1}{2a_\lambda}\left[1+(a_\lambda-1)\exp(-a_\lambda)-a_\lambda^2 \mathrm{E}_1(a_\lambda)\right], \tag{6.9}$$

where E_1 is the first-order exponential integral.

Given an unextinguished stellar spectrum $F^\star{}_\lambda$, the total bolometric infrared luminosity of the galaxy is

$$L_{\rm IR} = \int F_\lambda^\star[1-\mathrm{dex}(-0.4\bar{A}_\lambda)]\,\mathrm{d}\lambda. \tag{6.10}$$

We assume the emission is isotropic, since dust itself is generally optically thin to infrared light (except in heavily obscured starbursts). To compute the IR emission spectrum, one needs to model both the size distribution and the chemical composition of dust grains in the ISM. The model we use is based upon the MW model of Désert, Boulanger & Puget (1990), which includes contributions from polycyclic aromatic hydrocarbons, very small grains and big grains. Our chief refinement to this model is to allow a second population of big grains, closer to the star-forming region, to be in thermal equilibrium at a higher temperature for galaxies undergoing massive starbursts. Our emission model makes use of the colour–luminosity correlations observed by *IRAS*, and is detailed in DGS. We stress that a key weakness of this model is that it is based on a local sample and therefore assumes that dust properties do not evolve with time.

7 FREE PARAMETERS

The semi-analytic recipes described here contain a number of free parameters that will affect our results. We detail them in this section.

(i) Ω_B is the baryon fraction of the Universe. Whilst a high baryon fraction can have a serious effect on both the shape of the initial power spectrum of density perturbations (e.g. Sugiyama 1995) and the internal dynamics of haloes, providing it remains small it can be treated as a free parameter of the galaxy formation recipes. Increasing Ω_B results in more fuel for the star formation process, and should thus produce brighter galaxies.

(ii) β^{-1} is the star formation efficiency. This is the parameter that appears in equation (4.7). Increasing β will clearly result in less cold gas being turned into stars, moving the peak in the star formation history closer to the present day.

(iii) ϵ^{-1} represents the efficiency of mass loading during the triggering of a galactic wind by SN explosions (Section 4.2). Decreasing ϵ produces more feedback, heating more cold gas, ejecting more hot gas from haloes and thus reducing the amount of gas that can potentially form stars.

(iv) ψ is the normalization for the satellite–satellite merging law, equation (5.5). From Makino & Hut (1997) we expect its value to be around 0.017, but as stated in Section 5.4, we will treat it as a free parameter. Satellite–satellite merging competes with dynamical friction merging (especially for massive galaxies, see Mamon 2000), and the energetics of these collisions are different, so increasing ψ produces more diffuse galaxies, slowing the global star formation rate.

(v) χ parametrizes our merging law in equation (5.6), it is the critical mass ratio, i.e. that at which half the disc is disrupted. Our

Table 2. Summary of free parameters in the models.

Name	Meaning	Fiducial value	Range
Ω_B	Baryon fraction	0.02 h^{-2}	0.01–0.03 h^{-2}
β	Inverse star formation efficiency	50	10–100
ϵ	Inverse of mass loading for feedback	0.1	0.01–1.0
ψ	S–S merging normalization	0.017	0.01–0.03
χ	Galaxy merger power-law	3.333	2.0–5.0
ζ	Recycling efficiency	0.3	0.0–1.0
κ	Ratio of burst-to-bulge radius	0.1	0.05–0.2

fiducial value is 1/0.3, raising χ means that a larger disc will be disrupted by a smaller interloper. It therefore affects the morphological mix of our galaxy population.

(vi) ζ is the efficiency with which gas stored in the halo reservoir is admitted back to the halo during accretion, as explained in Section 3.1. Turning ζ right down will mean that metals and gas ejected from a halo are only returned very slowly, restricting the supply of metals to the cooling gas.

(vii) κ is used to determine the starburst radius from the bulge radius, as shown in equation (4.6). Smaller κ produces smaller starburst regions, where star formation is therefore faster, brighter and more heavily extinguished.

These parameters are summarized in Table 2. In addition, there are various other inputs to the models, including cosmology, initial stellar mass function (our fiducial model uses the Kennicutt 1983 IMF) and dust recipe, which will have an impact. We will consider the impact of changing some of these inputs, and varying the above free parameters within the range quoted in Table 2 in Paper II.

8 RESULTS

In order to test our models against observations, we will use a number of constraints in the literature. These fall into two main categories; intrinsic properties, i.e. relationships that apply for individual galaxies such as the Tully–Fisher relation or the dependence of metallicity on galaxy mass; and global properties, such as luminosity functions and clustering statistics. In this paper we will just present work concerning the local (i.e. low-redshift) Universe, and we will not use any spatial information. Paper II will include results from higher redshift, and the effect on the results of changes in the parameters described in Section 7. Papers III and IV will be concerned with the evolution of many of the statistics presented here, including the luminosity function and predictions of faint galaxy counts. We will also employ the spatial information from the dark matter simulations to investigate the clustering properties of our galaxies in Paper V.

In Fig. 4 we present a slice through the simulation at its final output time, with the dark matter density shown as a grey-scale. The slice is of side length 100 h^{-1} Mpc, the whole simulation box, but is 10 h^{-1} Mpc in thickness. We overplot the locations of bright ($M_B < -19.5$) galaxies, with the sizes of the points representing the galaxy B-band magnitudes. One can clearly see the way that the galaxies trace the high-density regions of the simulation, with most of the light concentrated in filaments and clusters. The power of the hybrid approach is immediately apparent when one considers that state-of-the-art attempts to model the subgrid physics with smoothed particle hydrodynamics produce around 2000 galaxies in a cube with one-third the volume of ours (Pearce et al. 2001), when we have 30 000 galaxies in total in our final time-step, thus giving us access to a much broader range of galaxy mass and merging history.

8.1 Effect of resolution

The identification of haloes with a given cut-off mass produces an effective resolution limit on our results. We define the galaxy resolution limit as the smallest halo mass times the baryon fraction of the Universe. Galaxies of baryonic mass less than this exist in our models, but no galaxy of mass greater than this could exist in unresolved haloes. This is thus a 'safe' limit for completeness in our sample, and produces a formal completeness limit of galaxy baryonic mass $m_{\rm gal} = 2.2 \times 10^{10}$ M$_\odot$ for the ΛCDM simulation.

In Fig. 5 we show the dependence of galaxy luminosity on total baryonic mass for galaxies in our fiducial model, defined in Section 7. The tight relationship between these two quantities allows us to estimate the magnitude completeness limit corresponding to the mass completeness limit we just defined. For each waveband, we perform a linear fit of magnitude on luminosity, and take the intercept of this fit with the line $m_{\rm gal} = 2.2 \times 10^{10}$ M$_\odot$. The results of this treatment for a number of optical and near-infrared filters are presented in Table 3. It will be noted that in the B band the resolution limit is approximately 1.5 mag fainter than the value of $M_\star = -20.5$ (assuming our value of H_0) from the Stromlo–APM survey (Loveday et al. 1992). We are therefore complete to galaxies with around one-quarter the luminosity of an $M_\star$ galaxy.

Although we are effectively complete down to these magnitude limits, we do not fully resolve the merging history of the faintest objects. This is an advantage of the method employed by Benson et al. (2000), where haloes identified in a dark matter simulation have their merging histories computed with the Press–Schechter formalism with theoretically infinite resolution, so even the smallest objects have a fully resolved history. The smallest galaxies in our models are in haloes that have only recently been identified in the simulations, and hence have a high gas content and star formation rate, but low metal and stellar content, and an inevitable spiral morphology. We expect the effects of this lack of information to propagate over the formal mass completeness limit. We consider resolution effects in rather more detail in Section 9.

8.2 The halo ICM

Before examining the galaxies themselves, we will look at the overall behaviour of the baryons in the dark matter haloes. In the left-hand panel of Fig. 6, we show the ratio of hot gas to total baryonic mass for dark matter haloes as a function of their circular velocity. Note that for isothermal spheres there is a one-to-one mapping from circular velocity to halo dark matter mass, $V_{\rm c} \propto M^{1/3}$.

We note that:

(i) the most massive objects have around 70 per cent of their baryonic mass in the form of a hot intracluster medium, with a weak dependence on circular velocity;

(ii) for objects with lower circular velocity, $V_{\rm c} < 300$ km s^{-1}, the gas fraction starts to fall off rapidly as the halo mass decreases, with this scale defining the difference between a galaxy and a cluster;

(iii) finally, around 100 km s^{-1}, the scatter becomes very large. This is due to the finite resolution of our approach, in that the smallest objects do not have well-resolved merging histories, so cooling does not start until the time-step on which they are discovered, resulting in an overestimate of the hot gas fraction. The cooling in these objects

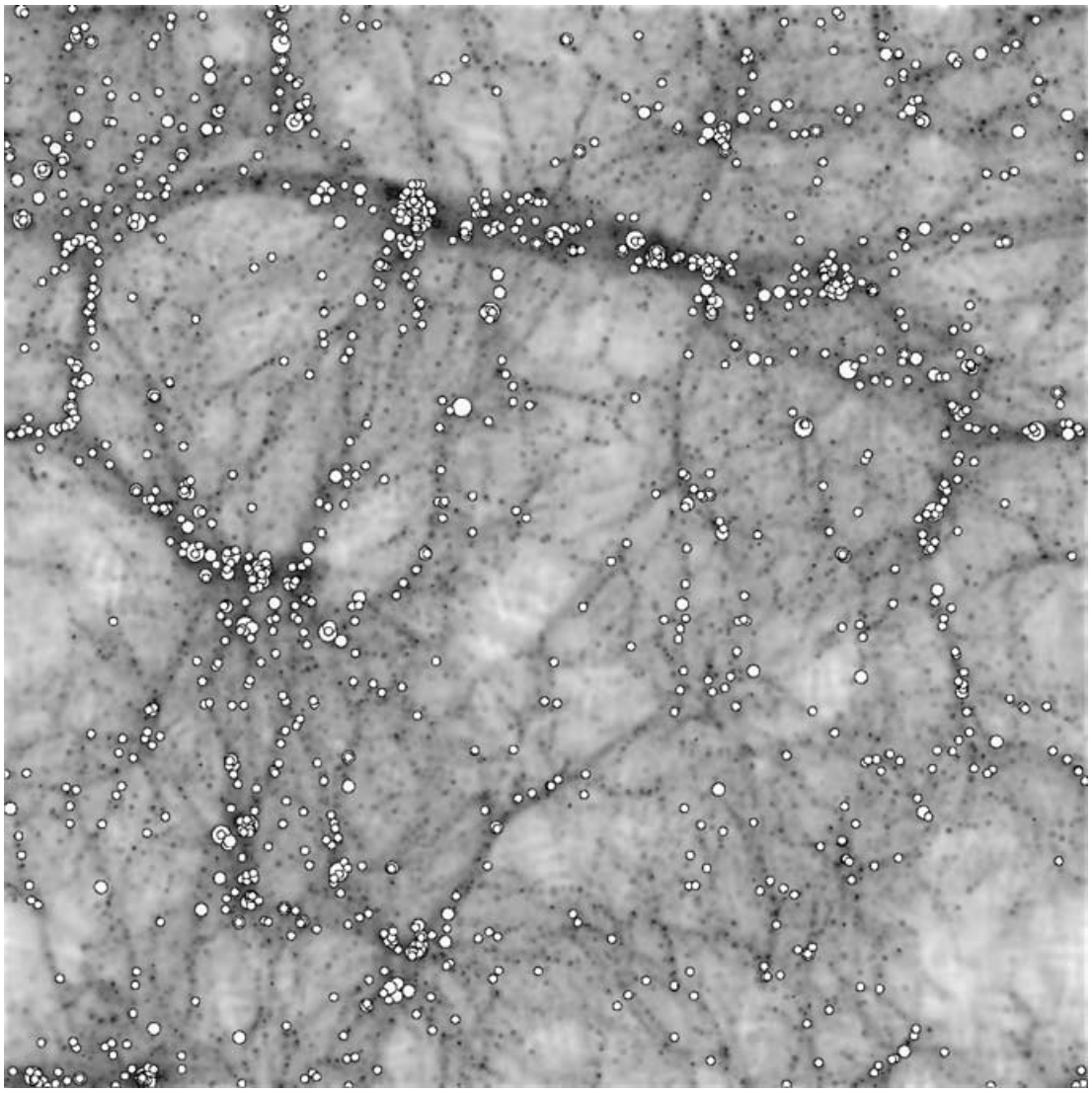

Figure 4. A slice through our simulation at a redshift of zero. The dark matter density is represented by a grey-scale, while the white circles mark galaxies, with their sizes depending on the *B*-band luminosity. Only bright galaxies ($M_B < -19.5$) are shown. The slice is 100 h^{-1} Mpc on a side (the entire simulation cube) and 10 h^{-1} Mpc in thickness.

is rapid enough, however, by the time the mass has increased slightly, they have joined the main locus on the plot.

In the right-hand panel of Fig. 6 we show the behaviour of the metallicity (defined as the mass in metals relative to the mass in gas) of the hot ICM. The resolution effects are clearly obvious again at low halo mass, with many objects of extremely low metallicity, since their stars are very young and have not had time to contribute via nucleosynthesis and ejection. Above 100 km s^{-1}, the metallicity is a decreasing function of halo mass, suggesting that larger haloes have a higher proportion of primordial gas, and galactic winds are less important in clusters than in galaxies. This result is in contrast to the observational results of Renzini (1999), who finds for the largest clusters ($V_c > 1000$ km s^{-1}) that the metallicity, as measured from the iron abundance, has a constant value of around one-third solar. This is between five and 10 times more metals than the amount we find in our haloes.

In Section 8.4 we will show that our galaxy metallicities appear to be quite accurate. Could the missing metals be hiding in the galaxies, or possibly in the halo reservoir? We have performed the same analysis as in Fig. 6, but added to the ICM all the metals contained in the halo reservoir, and all the metals locked up in the cold gas inside galaxies. This thus represents the maximum metallicity we can attain without somehow lowering the mass of non-metals in the ICM. Although this has a large effect for small haloes, there is little change in the metallicity of the largest clusters. We conclude that we are not simply 'misplacing' the heavy elements.

There are several potential explanations for this discrepancy.

(i) Observations measure only the abundance at the centre of the cluster, assuming the hot gas is chemically homogeneous. In fact, metallicity gradients have been observed in many clusters (e.g. Koyama, Takano & Tawara 1991 for the Virgo cluster, White et al. 1994 for several cooling flow clusters) and the observations thus yield an upper limit on the total metallicity.

(ii) The metallicity is measured using the iron abundance, whereas the metallicity we quote is the total mass of metals in the cluster. In fact, our models pay no attention to the energy injected

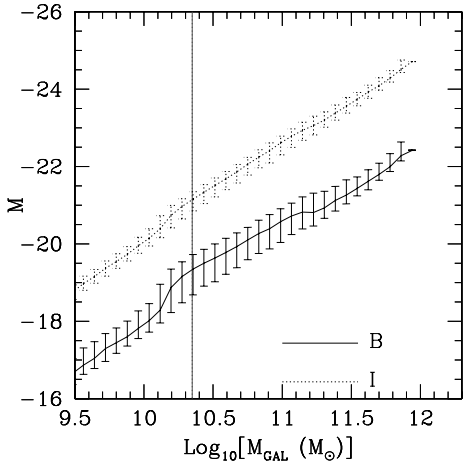

Figure 5. Galaxy absolute magnitudes as a function of the total baryonic mass in the *I* and the *B* bands. The solid vertical line shows the effective completeness limit as defined in Section 8.1. Results for the completeness limit in a number of wavebands are presented in Table 3.

Table 3. Resolution limits in a variety of optical and infrared bands (magnitudes for Johnson–Mould filters, λL_λ in solar luminosities for the *IRAS* bands). Magnitudes are absolute, in the Vega system, and calculated using the process described in Section 8.1.

Waveband	Completeness limit
U	−18.8 mag
B	−18.9 mag
V	−19.7 mag
R	−20.4 mag
I	−20.9 mag
J	−21.6 mag
H	−22.2 mag
K	−22.7 mag
12 μm	0.41×10^9 $L_\odot$
25 μm	0.25×10^9 $L_\odot$
60 μm	0.79×10^9 $L_\odot$
100 μm	1.34×10^9 $L_\odot$

into the ISM from type Ia supernovae (though these supernovae do contribute to the metallicity). The overall contribution of these objects is likely to be small, but, since these stars contain an iron core, the relative abundance of iron can be strongly enhanced, especially from evolved, gas-poor, early-type galaxies, such as are found in clusters. Thus the simple assumption used to convert Fe abundance to total metallicity, $Z/Z_\odot = Z^{Fe}/Z^{Fe}{}_\odot$, may be invalid in clusters, and will produce an upper limit on the total metallicity.

(iii) One should also consider the existence of Population III stars in the early Universe, which may have reionized the Universe homogeneously and likewise introduced metals into the primordial gas. This could be simulated by allowing a non-zero primordial metal abundance in our models.

(iv) The yields we use could be too low, especially if the IMF is biased towards massive stars in the early stages of cluster formation.

(v) Although these effects could all play a role in the metallicity underestimate, it seems more likely that the major contribution we are unable to model comes from objects below our resolution limit. Indeed, it is sensible to argue that the majority of metals seen in the ICM today came from old dwarf galaxies (Renzini 1999, 2000), in shallow potential wells, and we are a long way from achieving the resolution needed to see these galaxies.

8.3 Mass-to-light ratios

The mass-to-light ratio of dark matter haloes tells us important information concerning measuring the dark mass from observations, and concerning which environments are most efficient at producing energy. In Fig. 7 we show mass-to-light ratios, in solar units, for our haloes. We only include contributions to the luminosity from galaxies above our resolution limit. We compare results in two wavebands, the *B* band, where the energy is dominated by young stars, and the *K* band, which has been shown to reliably estimate the total stellar content (Kauffmann & Charlot 1998). One can see that the distributions are qualitatively similar in shape and that low-mass, galactic haloes are brighter per unit mass than the larger, cluster-sized objects. In the *B* band, MW-type haloes have a mass-to-light ratio of around 60, rising to 300 for the biggest clusters in the sample, whereas in the *K*-band the mass-to-light ratio is rather constant around 30 for galaxy size haloes ($V_c < 300$ km s^{-1}) and rises to 60 for the biggest objects.

Our failure to fully resolve the merging history of haloes with a circular velocity of less than 100 km s^{-1} makes it difficult to draw any firm conclusions concerning the shape of the mass-to-light ratio at the low-mass end.

8.4 Assorted galaxy properties

In Fig. 8 we show scatter plots for some of the intrinsic properties of our galaxies as a function of the galaxy characteristic velocity. This is the circular velocity for discs, three-dimensional (3D) dispersion velocity for bulges and a mass-weighted average of the two for objects of mixed morphology.

First, we show the total mass of baryons, i.e. cold gas plus stars, for our galaxies. The main locus of these points is seen to approximately follow the relationship $M_{bar} \propto V_{gal}^{5/2}$, although there is a substantial scatter to low baryon mass for galaxies with low velocity. This is another example of resolution effects entering our results, in that these are galaxies in small haloes that have not yet had time to cool. Compared with haloes, the virial masses of which go as the cube of the circular velocity, we already see a difference for galaxies: the baryonic mass depends more weakly on the velocity. We can already see the effects of the cooling and merging, in that the galaxy mass and velocity are obviously not scaling linearly with those of its parent halo.

Secondly, we present the metallicity of the cold gas in the galaxy. This is a slowly varying function of circular velocity. One can contrast this with the results for the halo metallicity presented in Fig. 6: although the metallicity of the cluster ICM decreases with increasing cluster mass, the metallicity of the central object in the cluster increases. There is also a clear upper cut-off in metallicity around twice the solar value, which corresponds to galaxies having consumed almost all of their gas and therefore having a very low, if any star formation rate at all, and the very few metal-rich young stars produced fail to enrich the ISM much more. It is worth noting that the main locus of our points is around a flat line with $Z \approx Z_\odot$, indicating that for our mass resolution the metallicity of the Milky Way is quite typical.

In the bottom left-hand panel, we show the star formation rate as a function of the characteristic velocity. We note that the majority of galaxies lie close to the line defined by the star formation rate

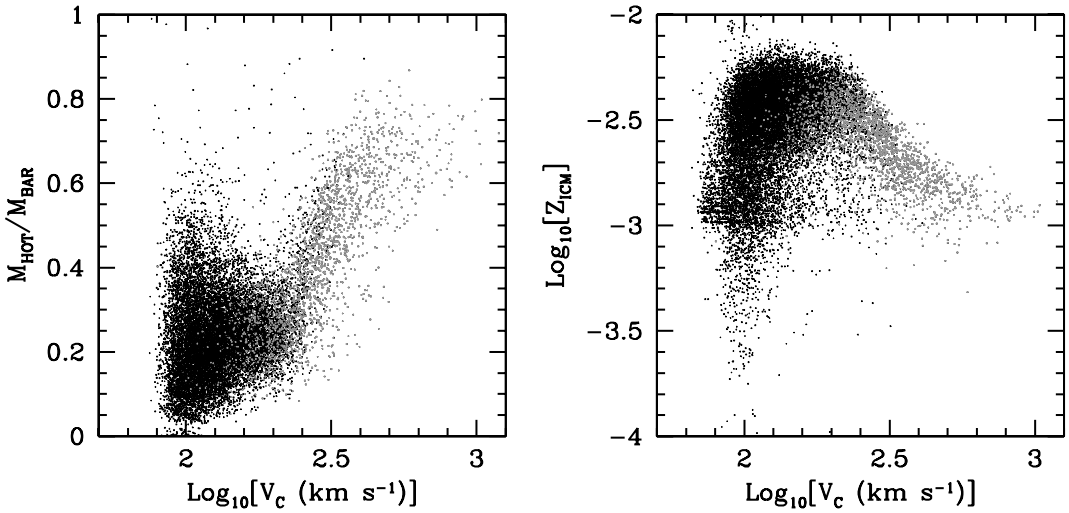

Figure 6. Baryonic properties of haloes, as a function of the halo circular velocity. On the left we show the ratio of hot gas in the halo to the total mass of baryons. On the right we show the metallicity of this hot phase.

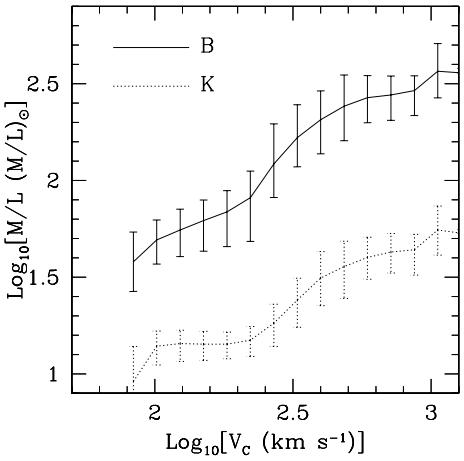

Figure 7. Mass-to-light ratios of our haloes, in units of the solar mass-to-light ratio. The upper line is for the *B* band and the lower line is for the *K* band. The lines are the medians, with error bars showing the $\pm 1\sigma$ scatter. We only include contributions from galaxies with baryonic mass greater than the formal completeness limit (Section 8.1).

being proportional to the velocity squared, but that there is a small population with very high star formation rates of up to 100 solar masses per year, presumably representing galaxies that are undergoing merger-driven starbursts. There is also considerable scatter at the low and high circular velocity end, due to galaxies that have exhausted their gas supply.

We leave a detailed consideration of galaxy sizes (bottom right-hand panel) to the next section.

8.4.1 Predictions for the Milky Way

To provide a consistency check for our models, we will compare MW-like galaxies in our simulations with the known properties of our galaxy. On each panel of Fig. 8 we show a grey hexagon for the approximate location of the Milky Way itself. The values adopted for the Milky Way are presented in Tables 4 and 5. We also select MW candidates from our galaxies by applying the selection criteria shown in Table 4, and requiring that the galaxy has spiral (or Sb) morphology (defined in Section 8.6). A selection of this kind produces around 600 MW galaxies in the simulation. The properties of these galaxies are shown in Table 5, where we quote the median and one standard deviation of the distributions, and compare with data for the Milky Way itself. We find a generally good agreement, in that applying the selection of Table 4 produce galaxies that do look a lot like the Milky Way. For most of the properties considered, the Milky Way would fall within 1- or 2σ from the median.

8.5 Galaxy sizes

In the final panel of Fig. 8, we show the galaxy half-mass radius. This is obtained from a mass-weighted average of the disc and bulge half-mass radii. The straight line is the radius of a galactic disc with the same velocity as its parent halo, if the radius is given by $1.68\bar{\lambda}R_{200}/\sqrt{2}$, i.e. equation (4.1) with $\bar{\lambda} = 0.04$.

In Fig. 9 we examine the sizes of field disc galaxies (i.e. galaxies that are alone in their haloes) emerging from our model. We compare our predictions with the observed bivariate space density distribution of de Jong & Lacey (2000) derived from a sample containing more than 1000 field spiral galaxies. As stressed by these authors, the main advantage of using this data is that it should be quite insensitive to surface brightness incompleteness effects. Note that in the model we included all the spirals with a bulge-to-total-light ratio of less than 0.33, i.e. morphological types later than Sab, whereas the data does not contain types later than Sd. Indeed, these very late-type galaxies cannot be excluded from the model sample as the only morphological information we are able to compute is a bulge-to-disc ratio. We would therefore expect an overprediction of the total number of galaxies, especially in the faint magnitude bins and for small disc sizes. However, we do not see this excess of small faint galaxies but rather an excess of bright small galaxies with respect to the data. Moreover, as pointed out by de Jong & Lacey (2000) the model distribution is (only slightly in our case) broader than the observed one. As we assume specific angular momentum conservation during disc formation, the obvious culprit is the too broad initial spin distribution of the dark matter haloes. However, we would like to draw the attention of the reader to the fact that

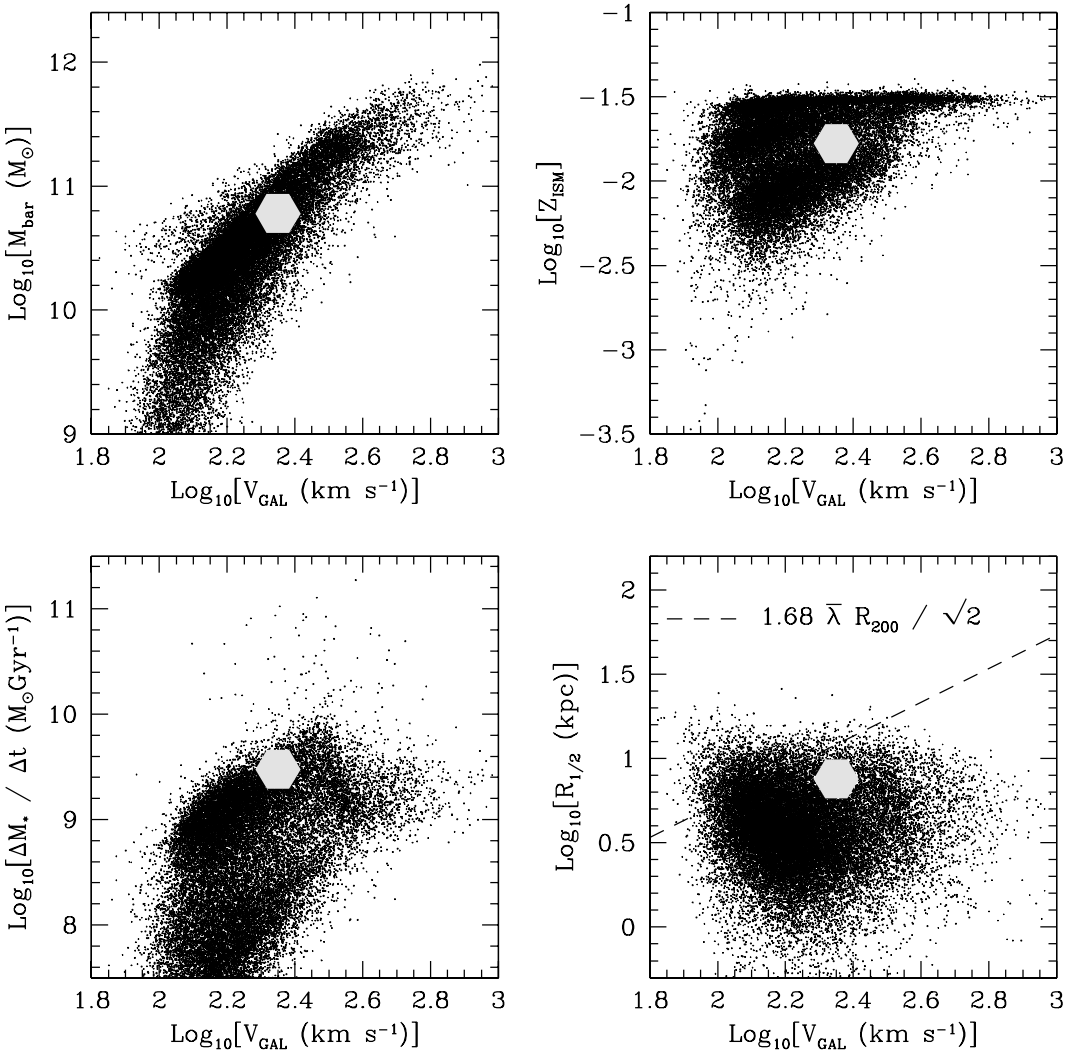

Figure 8. Key galaxy properties, as a function of the characteristic velocity, which is a mass-weighted average of the disc circular velocity and the bulge dispersion velocity. The top left-hand panel shows the total baryonic mass of the galaxy, i.e. stars plus cold gas. The top right-hand panel shows the metallicity of the interstellar medium, the ratio of the mass in heavy elements to the total gas mass. In the lower left-hand panel we show the star formation rate, and in the lower right-hand panel the galaxy radius, defined as the mass-weighted average of the disc, bulge and burst radii. For clarity, in this last figure we also plot the half-mass radius of a disc forming in a halo with velocity $V_{\rm gal}$ if its spin parameter were equal to 0.04 (equation 4.1). In each panel, the grey hexagon shows the approximate location of the Milky Way, from data given in Tables 4 and 5.

unlike the CLBF model, which overpredicts the density of small discs by quite a lot, we seem to predict approximately the right number of small-sized bright galaxies. Although our mass resolution biases us against small-sized galaxies, we estimate (see Table 3) that our sample is almost complete for galaxies with absolute magnitudes brighter than -20.9 in the I band. Therefore, the results displayed in the first three panels of Fig. 9 should not be too sensitive to an increase in resolution.

Once again we emphasize that our recipe for disc sizes depends crucially on the assumption of specific angular momentum conservation, and it has been shown (e.g. Navarro et al. 1995a) that dissipative effects can play a part in destroying some of the angular momentum of the gas, producing smaller discs. Domínguez-Tenreiro, Tissera & Sáiz (1998) show that this loss is less important for gas cooling on to 'evolved' galaxies with stellar bulges, so these effects are likely to preferentially reduce the size of the smaller, less resolved galaxies in our model.

8.6 Morphologies

We predict galaxy morphologies based on the ratio of B-band luminosities of the disc and the bulge components. It has been found (Simien & de Vaucouleurs 1986) that this ratio correlates well with the Hubble type. We define a morphology index, $I = \exp(-L_{\rm B}/L_{\rm D})$, such that a pure bulge has a value of 0 and a pure disc has a value of 1. Following Baugh, Cole & Frenk (1996), we translate this index into a morphology by assuming ellipticals have $I < 0.219$, S0s have $0.219 < I < 0.507$ and spirals have $I > 0.507$.

We first consider the morphological mix of galaxies in our simulation. Using the APM Bright Galaxy Catalogue (BGC, Loveday 1996), Baugh et al. (1996) point out that in the local Universe, the ratio E:S0:SP + Irr = 13:20:67. In fact, one cannot directly compare this to the ratio arising in our simulations, since the catalogue is a flux-limited sample of the Universe, whereas our samples are effectively volume-limited, containing an entire simulation cube at

Table 4. Observed parameters of the Milky Way, used to define a MW-type galaxy in the simulation.

Property	Actual MW	Reference	Range adopted
$m_{\rm gas}/m_{\rm bar}$	0.10	1	$\pm$0.05
M_K	$-$23.7 mag	2	$\pm$0.3 mag
$V_{\rm c}$	220 km s^{-1}	3	$\pm$20 km s^{-1}

The references are: 1. Prantzos & Aubert (1995); 2. Kent, Dame & Fazio (1991); 3. Binney & Tremaine (1987).

Table 5. Observed properties of the Milky Way compared with our predictions from a set of galaxies derived by applying the conditions presented in Table 4. The references are as that table except: 4. Binney & Evans (2001) and 5. Pagel (1998). The uncertainties in the fourth column are the $\pm1\sigma$ limits of the distributions.

Property	Actual MW	Reference	Simulation
$m_{\rm bar}$	0.6×10^{11} M$_\odot$	4	$(0.5 \pm 0.1) \times 10^{11}$ M$_\odot$
$\Psi_\star$	2–4 M$_\odot$ yr^{-1}	1	(1.5 ± 0.4) M$_\odot$ yr^{-1}
$r_{\rm D}$	4–5 kpc	2	(2.8 ± 0.6) kpc
$Z_{\rm ISM}$	0.016	5	(0.012 ± 0.002)

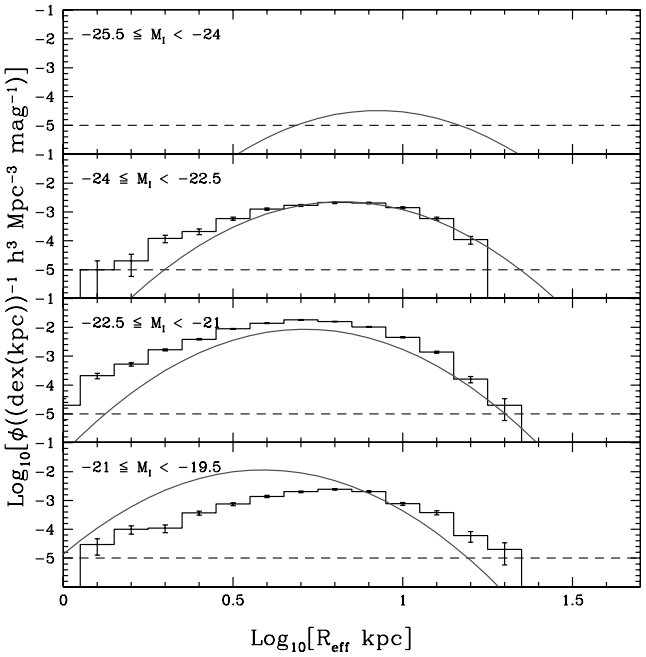

Figure 9. The bivariate density distribution of field spiral galaxies as a function of effective disc radii in different *I*-band absolute magnitude bins. Magnitude cuts are indicated in the top corner of each panel. The solid histogram represents results from the hybrid model and the solid line represents the best fit to the data of de Jong & Lacey (2000) (their equation 7). The horizontal dashed lines in each panel represent our volume resolution.

$z = 0$. If there is any correlation between B-band luminosity and morphological type, some types may be seen over a larger volume in the BGC than others.

For a better comparison, we take the Stromlo–APM redshift survey (Loveday et al. 1996) and construct from it a volume-limited sample of galaxies containing all of those with a well-measured morphology, down to the absolute magnitude limit of our models, $M_B < -18.8$. We obtain E:S0:SP + Irr = 13:11:76 (the sample contains just 120 galaxies, so the Poisson errors are $\pm$3:3:8). Thus it appears that, since S0s are intrinsically brighter objects than spirals, the ratio from the BGC overestimates their abundance when this selection is made. We compare with the ratio found for our models with a similar magnitude limit. We find 17:16:67 for a sample of 20 000 galaxies, with a total number density around 20 per cent higher than that measured in the volume-limited Stromlo–APM sample. Thus it seems that for these galaxies we slightly overrepresent Es and S0s and underrepresent spirals, although it must be remembered that the statistics from the real data are poor.

If we apply a more conservative limit, going 0.75 mag brighter (i.e. a factor of 2 in luminosity), we find the Stromlo–APM data give 14:13:73 (250 galaxies in the sample, since the volume probed has increased, which implies Poisson errors 2:2:5) and 21:17:61 from our models. Again our number density is around 20 per cent higher than in the real data. In other words, as we increase the absolute magnitude of the selected objects, the quantitative comparison with the data remains quite stable. This effect is unsurprising, since in Section 8.1 we stated that our magnitude for completion in the B band is around -18.9.

It is worth noting that, with more complete surveys such as 2dF and SDSS, volume-limited samples of the local Universe with much better statistics will be available, and these numbers can be checked more accurately. As it is, we are thus broadly confident that our methods produce approximately the correct morphological mix at low redshift. However, it must be appreciated that we are comparing a simple recipe for determining morphologies with what is observationally a sophisticated and subjective process (i.e. assuming a one-to-one mapping between bulge-to-disc ratios and Hubble type), and one should not place too much emphasis on these results.

8.7 Colours

Next, we consider the dependence of colour on galaxy morphology. Fig. 10 presents $B - V$ colours as a function of morphology for our galaxies. The solid line represents data taken from Buta et al. (1994). Our spiral galaxies closely follow the data both in average value and scatter, but more important is the dependence of colour on morphological type for early-type galaxies, which is not as strong as observed, leading to elliptical/SO galaxies that are too blue with too strongly scattered colours, independently of the exact criterion used to define these morphological types. We will come back to this important issue in Paper II, but we mention that this is due to recent overcooling leading to late star formation in early-type galaxies. Indeed, if we remove ellipticals and S0s with more than 1 per cent of gas in mass still present at $z = 0$, our colours (and dispersions) are in good agreement with the observations, although more than half of our early-type galaxies drop out of the sample. Finally, we note that even though this discrepancy is problematic, our median colours are still within 1–2σ of the observations, whatever the morphological type considered. Furthermore, RAM pressure stripping could help redden the colours of S0s and cluster ellipticals.

8.8 Luminosity functions

The luminosity function, i.e. the number density of objects as a function of magnitude, is a commonly measured observational statistic. In Fig. 11 we show the results for the luminosity functions arising from our models in different wavebands. We compare these with data coming from several different sources. For the optical and near-infrared (B and K band) we use recent data from the 2dF galaxy redshift survey (Cross et al. 2001 for the B band, Cole et al. 2001 for

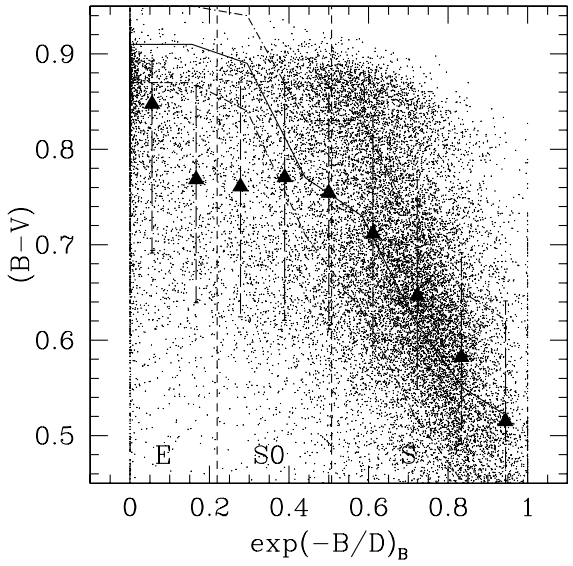

Figure 10. $B - V$ colours as a function of morphology for our galaxies brighter (than -18.8 in the B band) seen face-on. The solid triangles show the median of the distribution, with the dashed error bars giving the 1σ range. The solid line shows the measurements of Buta et al. (1994) (converted to a bulge-to-disc ratio index according to Simien & de Vaucouleurs 1986) with the observed 1σ dispersion (dot-dashed lines). The short-dashed vertical lines show where the divisions are drawn between elliptical, lenticular and spiral morphologies.

the K band). For the *IRAS* wavebands (25 and 100 μm), we compare with results from Soifer & Neugebauer (1991).

As noted in Section 8.1, the finite mass resolution of our simulations produces a lower bound to the galaxy luminosity to which we are complete in any given waveband. Although we have galaxies fainter than this limit, there will be galaxies of similar luminosity in haloes that we do not resolve, and hence we expect to see some depletion of the counts lower than this limit. We plot the magnitude limit corresponding to the mass limit with a vertical dashed line in each of the four plots in Fig. 11. The resolution effect, i.e. the depletion of counts faintwards of this line, is apparent in all four wavebands. For the optical and near-infrared, this depletion seems to occur quite close to the resolution limit, starting 0.5–1.0 mag brighter. This suggests that this is an accurate estimate of the resolution limit in these bands. In the *IRAS* bands, the turnover in the counts occurs at a luminosity of around 0.5 dex (roughly 1 mag) brighter than the formal limit.

On the whole, the plots show quite good agreement with observed galaxy number densities over two orders of magnitude in wavelength. In the B band, the counts agree reasonably well with the shape and normalization of the 2dF data. This is a significant improvement on the hybrid model results of KCDW who found a fairly flat, power-law slope for the B-band luminosity function in a ΛCDM universe. Our results are similar to those of CLBF, although they are able to obtain a slightly better amplitude of the luminosity function because they have a free parameter to control the fraction of stars forming brown dwarfs (which of course do not contribute to the luminosity function), which they fix using the B-band number density at $M_{b_J} - 5\log_{10} h = -19.8$.

In the K band, we obtain a good (although on the high side) overall number density, and fit the data quite well at the 'knee' of the luminosity function, representing most of the old stellar mass in the Universe. Moreover, we note that our overprediction of intermediate luminosity galaxies with respect to the 2dF data might not be so much of a problem since a more recent survey with similar number statistics (Huang et al. 2003) seems to also find a higher number density of these galaxies. Our luminosity function has an overall appearance that is similar to the Schechter function, which describes the data well, i.e. we reproduce the exponential cut-off at the bright end that is present in the data. This behaviour is due both to the prescription we employ to solve the overcooling problem (see Section 3.3) and the more efficient feedback we implemented (see Section 4.2). KCDW, with their different prescriptions for both of these effects, obtain a power-law shaped K-band luminosity function that overestimates the number of bright objects. CLBF obtain a better fit to the data, which is insensitive to dust emission and absorption but once again relies on the free parameter to control the number of brown dwarfs formed. The conclusion is that overpredicting the bright end of the K-band luminosity function seems a rather natural feature of semi-analytic models and one has to invoke at least a rather strong feedback (from supernovae and/or AGN) to remove stars in the brightest galaxies.

For the *IRAS* number densities, we again find fairly good agreement with available data except at the faint end where resolution effects dominate. In both wavebands the agreement is of excellent quality over the whole range of luminosities above the formal resolution limit.

On the whole, we find broad agreement with observations over several orders of magnitude in galaxy luminosity and two orders of magnitude in wavelength, a good confirmation of our models both for the semi-analytic recipes of the physics of galaxy formation, the determination of spectra and the modelling of extinction effects.

8.9 Optical–IR colour–magnitude relation

In Fig. 12 we examine the behaviour of the optical–IR 'colour', i.e. the difference in luminosity between these wavelength ranges, as a function of galaxy luminosity. We include all galaxies with baryonic masses greater than the formal mass limit, as defined in Section 8.1. Soifer et al. (1987) find that the colours of bright galaxies in the *IRAS* sample are closely related to their total IR luminosities, but have no correlation with the optical luminosity, implying that galaxies have similar optical properties no matter what their IR emission is. It can be seen from Fig. 12 that our galaxies follow a similar trend. On the left-hand side we show the dependence of the ratio of bolometric IR luminosity to B-band luminosity against the IR emission. For low to moderate luminosities, this relationship is close to linear. A linear regression reveals that the slope of this correlation is 0.70. Fitting a linear relationship 'by eye' to the data of Soifer et al. (their fig. 5a) yields a slope of 0.75, so the results are quite comparable. Our distribution is somewhat skewed, with a population of galaxies of quite high IR flux densities and extremely red colours, departing from the slope that fits the rest of the galaxies quite well. Since galaxies in Soifer's plot need both IR and B-band luminosities to be measurable, galaxies that are intrinsically faint in the B band are discriminated against in this sample. It is thus likely that many of the galaxies in the upper part of our distribution would not be included when this sort of selection is applied. This is especially clear in the right-hand panel of Fig. 12 where we plot the colour against the B-band luminosity. No correlation is visible, except for the handful of extremely red galaxies, which all have B-band luminosities less than 3×10^9 $L_\odot$. Note that in Soifer et al. (1987) (their fig. 5b) there are very few galaxies indeed with B-band luminosities this low. Furthermore, the *IRAS* sample is flux-limited, whereas ours is volume-limited, and therefore the observations should be biased towards IR bright and distant objects, which partly explains why

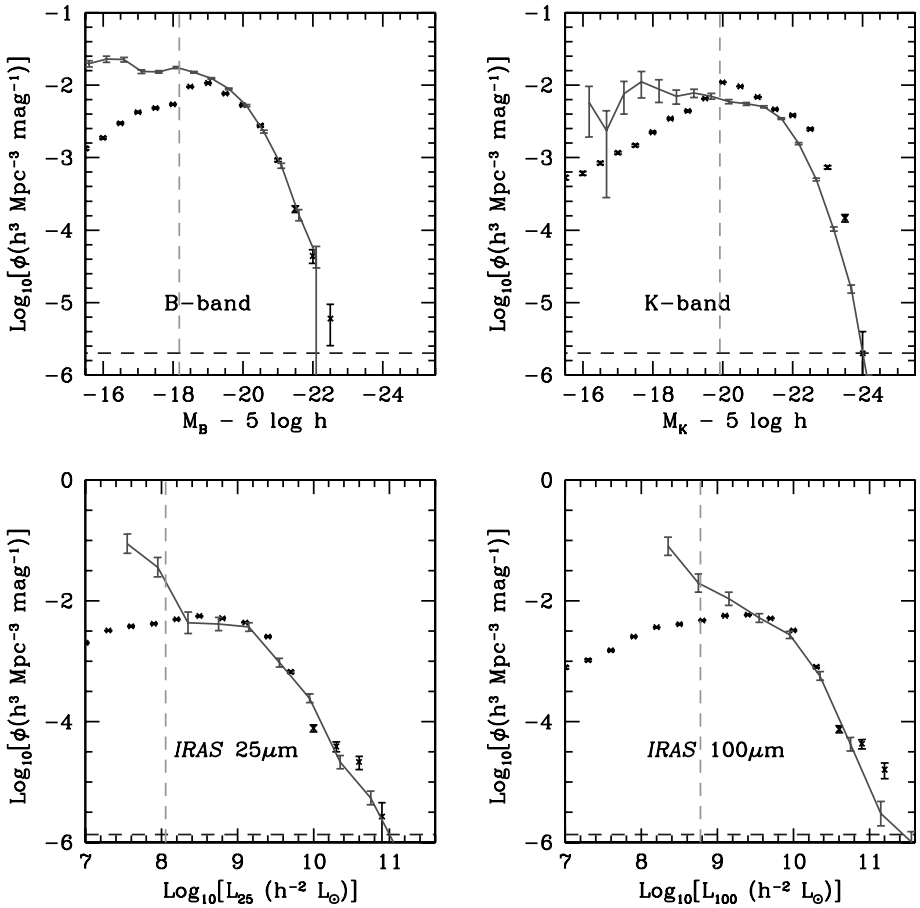

Figure 11. Galaxy luminosity functions. The top left-hand side: the *B*-band luminosity function (points with error bars), with data from the 2dF galaxy redshift survey (Cross et al. 2001) (connected, lighter curve with error bars). The top right-hand side: the *K*-band luminosity function, with data from 2dF (Cole et al. 2001). The error bars are Poisson errors, and the magnitudes are given in the AB system. The bottom left-hand side: *IRAS* 25-μm luminosity function. The bottom right-hand side: *IRAS* 100-μm luminosity function. *IRAS* data comes from Soifer & Neugebauer (1991). In each panel, the dashed vertical line is the formal resolution limit defined in Section 8.1, and the dashed horizontal line is the volume limit (i.e. one galaxy per magnitude bin).

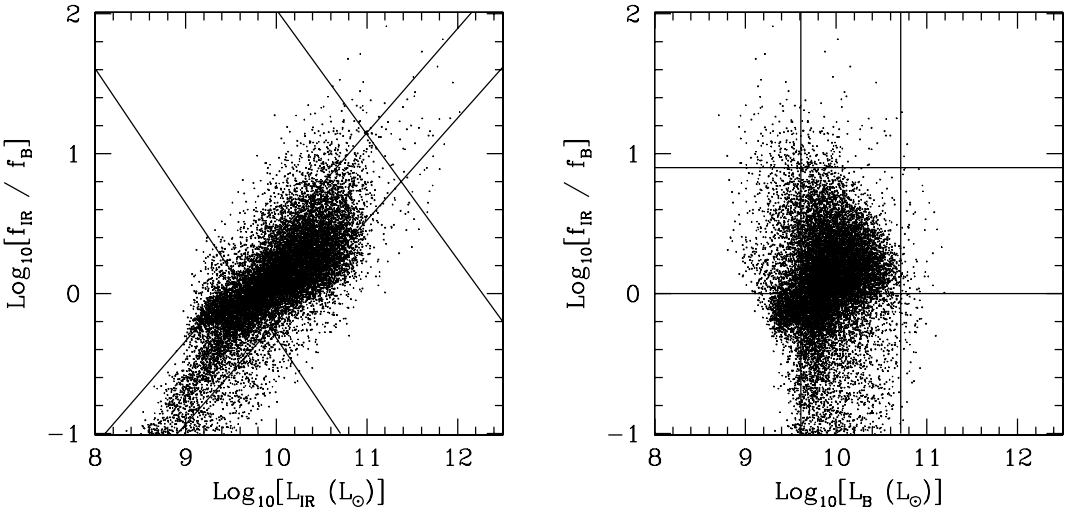

Figure 12. The relationship between the optical–IR colour, defined as the ratio of infrared to blue luminosity, and the absolute bolometric luminosity in the infrared (left) and *B* band (right). All galaxies with baryonic masses greater than the formal mass limit are plotted (Section 8.1). Solid lines represent lines excluding approximately 1σ of the data plotted in figs 5(a) and (b) of Soifer et al. (1987).

our population of galaxies seem to populate the low-IR part of both diagrams. The other, and more fundamental reason is that although we only plot galaxies above the formal mass resolution limit, we do not satisfactorily resolve the star formation history of galaxies just above this threshold, which obviously leads to bluer colours and lower IR emission because according to the simulation, most of them have formed very recently.

Bearing these limitations in mind, we conclude that our model is fairly successful at reproducing the observed correlations between luminosity and colour in the infrared and blue wavebands.

8.10 Tully–Fisher relation

There is a well-established empirical relationship between the luminosity and rotation velocity of galactic discs. This effect is known as the Tully–Fisher relation (Tully & Fisher 1977). In the top left-hand corner of Fig. 13 we present a scatter plot for the *I*-band magnitude against circular velocity for all spiral galaxies in our simulation. The dashed line on the plot represents the best fit to the *I*-band Tully–Fisher relation found by Giovanelli et al. (1997), where we have assumed that the 21-cm linewidth is simply twice the galaxy circular velocity. We plot the relation only for the range of velocities over which Giovanelli et al. have data. We note that the value of the Hubble constant derived from this data set is $h = 0.69 \pm 0.05$, which is consistent with our value of $\frac{2}{3}$. It is clear from this plot that the majority of galaxies in our sample do indeed fall in a single locus on these axes, although there is significant scatter in the relation and a population of galaxies with rather lower luminosities for a given circular velocity. The main locus itself is close to the relation of Giovanelli et al. (1997), although our galaxies appear to be somewhat fainter, and the slope of the correlation is shallower.

We proceed to perform a more detailed analysis by splitting our galaxy sample into three subsamples. These are:

(i) 'field' galaxies, that are the sole occupants of their dark matter haloes;

(ii) 'cD' galaxies, which are defined as central galaxies in haloes with more than one occupant; and

(iii) 'satellites', all non-central inhabitants of multimember haloes.

In the other three panels of Fig. 13 we show the Tully–Fisher relation for each of these subsamples. In each case, we plot the line

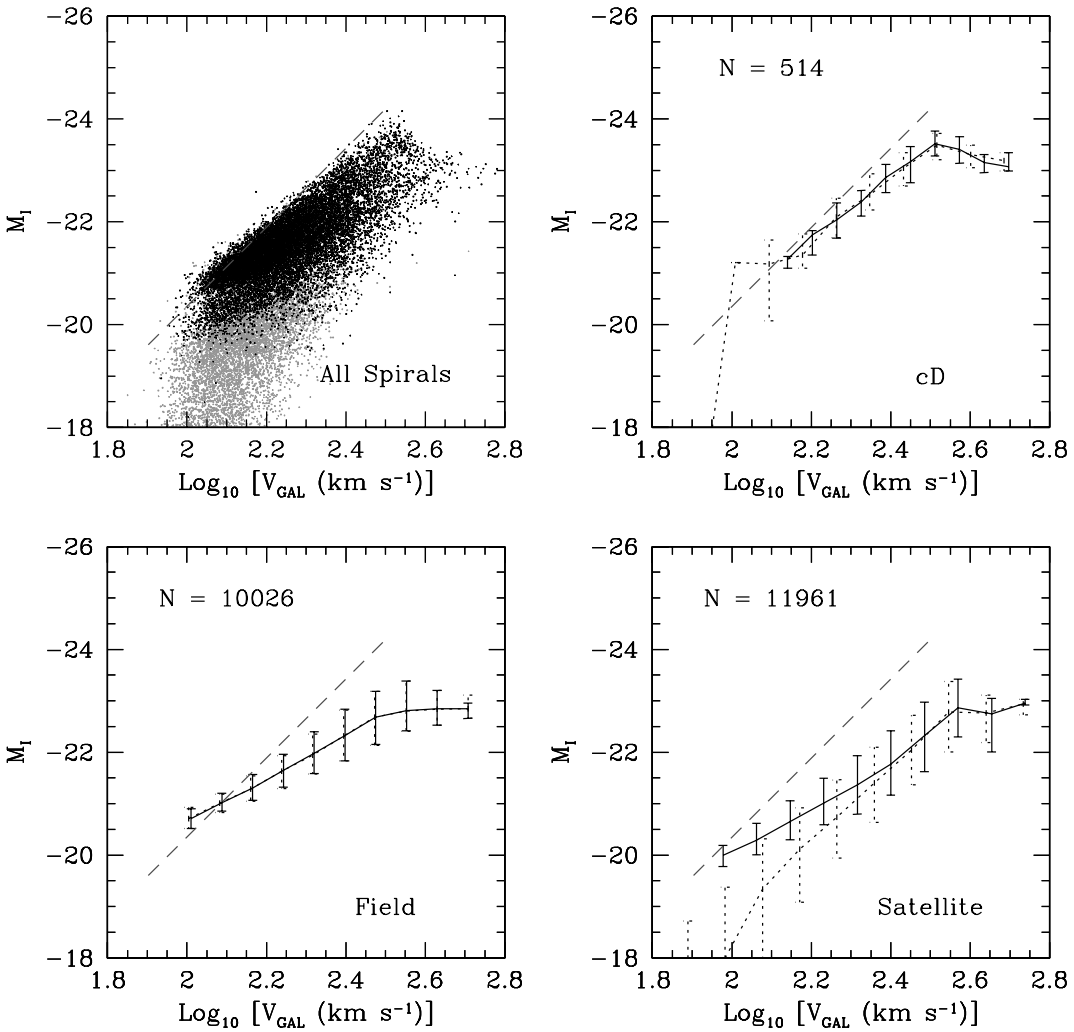

Figure 13. The Tully–Fisher relation in the *I* band for all our spiral galaxies above the resolution limit (top left). The other three panels show the same data split into subsamples by galaxy environment, with the median and $\pm 1\sigma$ scatter. In each panel, the dashed diagonal line is the observational Tully–Fisher relation derived by Giovanelli et al. (1997). We also show the distribution both before (dashed line) and after (grey solid line) making the gas column density selection of equation (8.1).

connecting the median of the binned distribution, and add error bars to show the $\pm 1\sigma$ range. We again compare with the line from Giovanelli et al. (1997). It can be seen from the plots that our cD galaxies follow the observed Tully–Fisher relation quite well, although these galaxies are slightly too faint (or rotate too fast).

It is clear that the observational Tully–Fisher relation is not followed nearly so well by our field and satellite galaxies (the solid lines in the panels of Fig. 13). The median magnitude as a function of circular velocity is well below the observed trend, and the 1σ scatter is very large compared with that for the cD galaxies. Also the satellite galaxies are largely responsible for the large scatter in the plot in the top left. This is especially interesting because the sample used by Giovanelli et al. (1997) to measure this relation is actually a sample of spiral galaxies in 24 galaxy clusters, i.e. corresponding to our satellites. There are two key reasons for this discrepancy: the first is that satellites or galaxies with massive bulges are no longer being 'fed' with hot gas cooling on to the discs, with the result that they use up their store of cold gas, and consequently have low star formation rates and luminosities. In other words, galaxies that have been satellites for different lengths of time can therefore have quite different gas fractions, and this is responsible for the large scatter in the relation. Secondly, and more fundamental reason is that the circular velocity of a galaxy results from the complex interaction between baryons and dark matter, and as a result our simple treatment of the disc dynamics is probably too crude, especially for the fast rotating (massive) systems, which are also likely to have a complex merging history.

As pointed out by CLBF, the data in fact are subject to a strong selection effect in that galaxy rotation velocities are measured using H I or optical emission lines, and a significant column density of interstellar gas must be present in order to make those measurements. In fact, many of our satellite spirals are gas poor, and so would not be included in the data set of Giovanelli et al. To briefly demonstrate this selection effect, we applying a cut-off mimicking that which must exist in the data,

$$M_{\rm cold}/\pi r_{1/2}^2 > 10^7\,{\rm M}_\odot\,{\rm kpc}^{-2}, \tag{8.1}$$

which corresponds to a column density of approximately $N_{\rm H} = 1.25 \times 10^{21}$ atom cm^{-2}. This selects only the satellites that still have a significant amount of gas, and the effect is clearly visible on the Tully–Fisher relation (solid line in Fig. 13). The median now has similar behaviour to that of the field galaxies, although the satellites are still between 0.25 and 0.5 mag fainter than field galaxies depending on the circular velocity. The scatter has been vastly diminished, although it is still larger than that of the field galaxies. This also reduces the number of satellite spirals by approximately a factor of 2.5, in keeping with a number of observational results (e.g. Solanes et al. 2001) demonstrating that spirals in cluster environments are often gas-depleted.

We can also compare the scatter in the Tully–Fisher relation with the intrinsic scatter estimated from the observations of Giovanelli et al. We compare the scatter for a MW-type galaxy with circular velocity 220 km s^{-1}, using their equation (11). Their prediction is a scatter of $\epsilon_{\rm int} = 0.22$ mag. We find a slightly higher scatter for our field galaxies, by interpolating a spline fit to the error bars shown in Fig. 13, $\epsilon_{\rm int} = 0.33$. The 'cD' galaxies have a similar scatter, $\epsilon_{\rm int} = 0.24$, and the cluster spirals have much larger scatter to the observations predict, $\epsilon_{\rm int} = 0.97$ for the whole sample and $\epsilon_{\rm int} = 0.49$ after applying the column density selection of equation (8.1). Whilst it is encouraging that our cD galaxies have similar scatter than the data, it seems that even with selection, our model predicts a wider variety of spirals in cluster environments than are observed. These results imply that, in reality, either gas cooling plays the dominant role in determining the properties of spirals in both cluster and field environments, and our suppression of cooling on to satellite and massive field galaxies is responsible for their variety, or our estimates of galaxy velocities are systematically offset towards higher values and more dispersed than in reality.

8.11 Faber–Jackson relation

Analogous to the Tully–Fisher relation for spiral galaxies, observations show a correlation between the luminosities and dispersion velocities for elliptical and S0 galaxies, known as the Faber–Jackson relation (Faber & Jackson 1976). In Fig. 14 we show the Faber–Jackson relation for our galaxies, and compare with that found by Forbes & Ponman (1999) for cluster ellipticals (shown by the solid line). We mimic the observational selection by plotting only elliptical and S0 cluster galaxies (i.e. haloes containing more than 10 galaxies), but note that the match is qualitatively similar if we use all the galaxies we classify as elliptical or S0s.

It will be seen that the galaxies sit around a line that is slightly fainter (or at a slightly higher velocity dispersion) than that of Forbes & Ponman, once again consistent with too high velocity dispersion estimates. However, we consider this a fairly good imitation of the observed Faber–Jackson relation for our galaxies because of the scatter in the relation, which is shown as an inset in this figure. This has been obtained from the differences between the straight line and the individual velocities as a function of B-band magnitude. We have shaded the central area containing 68.3 per cent of the data, and we can thus read off the 1σ limit as $+0.12$ and -0.15 dex on each side of the observed relation, corresponding to approximately ± 0.13 dex of scatter if we were to define a Faber–Jackson relation directly from our model galaxies. Estimates of the scatter by Forbes & Ponman yields a 1σ limit of ± 0.10 dex, so we are confident that even though our relation looks slightly offset from the data our models still are in fairly good agreement with observations.

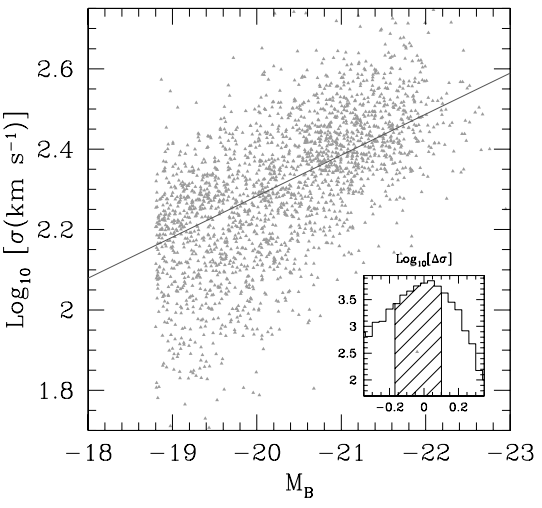

Figure 14. The Faber–Jackson relation for our elliptical and S0 cluster galaxies. We only plot galaxies above our resolution limit $M_B = -18.8$. The solid line shows the fit found by Forbes & Ponman (1999) for local ellipticals. The inset shows the scatter of our galaxies from the observed relation.

8.12 Fundamental plane

It has been found observationally that the scatter in the Faber–Jackson relation is rather larger than what would be expected simply from measurement errors, leading to the idea (Djorgovski & Davis 1987; Dressler et al. 1987) that there is a further variable involved. The so-called fundamental plane relates the galaxy luminosity, velocity dispersion and radius, and is found to apply to elliptical galaxies, and, to some extent, the bulges of spirals as well.

Following a similar procedure to that employed by observers, we assume that the plane is of the form:

$$\log R_{\rm e} = a \log \sigma + b\mu_{\rm e} + c. \tag{8.2}$$

In this equation, $R_{\rm e}$ is the effective radius of the spheroid, defined as the radius that contains half the light of the galaxies when viewed in projection (for our bulges this is $R_{\rm e} = 1.815\, r_{\rm B}$, see Hernquist 1990). σ is the one-dimensional velocity dispersion and $\mu_{\rm e}$ is the average surface brightness within $R_{\rm e}$.

We determine the coefficients (a, b, c) in equation (8.2) by minimizing the total absolute deviation of all our elliptical and S0 galaxies above the formal resolution limit from this plane. In order to do this, the variables are first transformed into a 'normalized' coordinate system by subtracting off their mean value and dividing by the standard deviation, i.e.

$$\sigma_1 = (\sigma - \bar{\sigma})/\sqrt{\langle \sigma^2 \rangle}. \tag{8.3}$$

We show our fundamental plane relation in Fig. 15, in both the B and the I bands. There is clear evidence for a fairly tight correlation of these three bulge properties, in both wavebands, with the relation given by

$$(a, b) = \begin{cases} (1.47, 0.38) & B \text{ band} \\ (1.48, 0.36) & I \text{ band.} \end{cases} \tag{8.4}$$

For a sample of 74 early-type galaxies towards the Coma cluster, Scodeggio et al. (1998) find that the fundamental plane that minimizes the absolute deviation is given by (their table 6):

$$(a, b) = \begin{cases} (1.40, 0.35) & B \text{ band} \\ (1.70, 0.33) & I \text{ band.} \end{cases} \tag{8.5}$$

These results for the coefficients (a, b) are in good agreement with those we find (and reasonable agreement with other observations, e.g. Pahre & Djorgovski 1997). It will be noted that we also succeed in replicating a trend that existed in the data as well, namely that the parameter b is quite independent of the waveband. The steepening of the slope a of the dependence on the velocity dispersion as a function of increasing wavelength, is not so marked in our simulated galaxies.

In the insets to the two panels in Fig. 15 we show the distributions of the residuals about the plane, i.e. the differences between the actual log $R_{\rm e}$ values and the prediction from the derived FP relation. The width of these distributions are approximately ± 0.15 and ± 0.12 for the B and the I bands, respectively. An estimate from Scodeggio et al. (Fig. 7) shows the 1σ scatter in the data to be approximately ± 0.10 dex in both the B and the I bands. Thus we not only produce fundamental planes with a similar tilt to that observed, but also obtain a good match for the residuals. This also suggests that much of the scatter in the observed FP relation is intrinsic, due to the different merging histories of the objects examined, rather than being caused by measurement errors.

This is the first time that semi-analytic models have been pushed as far as looking at the fundamental plane, and we find the results to be an encouraging vindication of our methodology.

9 RESOLUTION TESTS

The issue of resolution enters our models in several ways, and it is important to check that we understand how resolution affects the results presented here. In this section, we will endeavour to show that our results are quite stable to an increase in resolution. This will generally be performed by comparing with a simulation of slightly *worse* resolution, and showing that the results have already converged.

In the following sections, we consider the effects of finite force, time and mass resolution. We will use the abbreviations HTR and LTR to denote high and low time resolution, and HMR and LMR for high and low mass resolution. Thus the HTR (or HMR) is generally the original, high-resolution simulation.

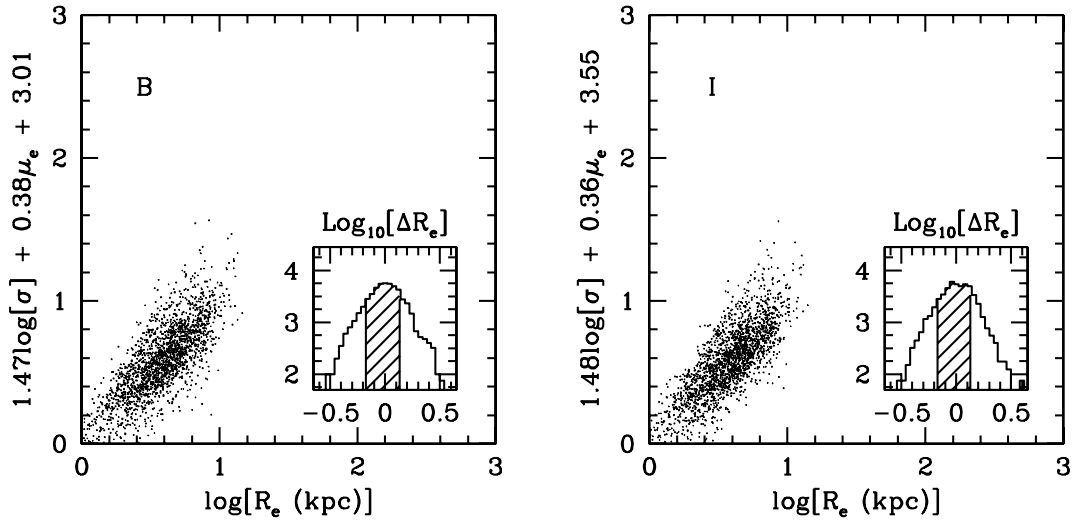

Figure 15. The fundamental plane for our elliptical and lenticular galaxies, in the B (left) and I bands (right), above our resolution limit. The coefficients on the y-axis are found by minimizing the absolute deviation from the line $y = x$, as explained in the text. The insets show the distribution of the residuals about the plane.

9.1 Spatial resolution

Initially, in the *N*-body simulations, there is a spatial resolution set by the softening length used in computing the smoothed gravitational potential. In our code, this softening length is set at 1/20 of the mean interparticle separation. For our ΛCDM simulation, at a redshift of zero, this is around 30 kpc. We identify haloes with a density contrast of 200 relative of the critical density, which thus have an average particle–particle separation length of the mean separation divided by $(200/\Omega_t)^{1/3}$, where Ω_t is the ratio of the matter density to the critical density as a function of time. The mean distance between particles inside haloes at $z = 0$ is thus twice the softening length. Improved resolution may then have an effect on the central, dense regions of haloes, where the separation length is smaller than the softening length, but is unlikely to influence the virial mass itself. Thus, attempts to fit density profiles to these inner regions would be affected by the softening, but we do not do this, measuring only virial masses and energies, and assuming an isothermal profile for all our haloes. The spatial resolution limit set by gravitational softening is thus not a relevant concern for this work.

9.2 Time resolution

To calculate the particle trajectories in the *N*-body simulations, we use around 20 000 time-steps (Table 1), equally spaced in expansion factor from $z \approx 35$ to the present day. To construct the halo merging tree, a subsample of these steps is used, spaced logarithmically in expansion factor. This produces 70 time-steps between the present day and the first identification of a virialized object with 20 particles, at a redshift of $z \approx 10$. The spacing between the penultimate and the final time-steps corresponds to 60 of the simulation time-steps. It is thus the choice of output times for constructing the merging tree that really dictates time resolution. This especially has an effect on the baryon cooling tree described in Section 2.4, since we will only be interested in binary mergers in this section. If a massive halo has accreted two smaller haloes in the course of one time-step, we only follow the galaxies in the larger of the two haloes. The occurrence of more-than-binary mergers obviously increases with the length of the time-step, so more galaxies are lost with worse resolution.

To check our sensitivity to this time-step resolution, we compare results with those from a tree made using only half the number of simulation outputs (i.e. with a factor of 2 worse time resolution). There is a one-to-one correspondence between haloes (but not galaxies) in the final time-step of the two trees, so we can look to see how resolution has had an effect on the properties of individual haloes.

Fig. 16 shows a halo-by-halo comparison for several key halo properties. The solid line shows the number of objects with a given fractional difference between their cold gas masses in the full simulation and in the low time resolution version. It can be seen that there is a very small systematic offset between haloes in the two simulations, with haloes in the HTR having slightly more cold gas, and there is some scatter (the curve has a FWHM of the order of ±20 per cent).

For hot gas, the picture is rather different. Although there is a sharp peak at zero difference, this is due to haloes that have just been discovered in the final time-step, and so have not cooled at all in either version. For the rest of the haloes, the peak in the distribution occurs when haloes in the HTR run have around 40 per cent more hot gas than those in the LTR. However, the total amounts of hot gas in the two simulations are similar, with the HTR run having around 8 per cent more gas than the LTR. This implies that the haloes containing large amounts of the hot gas are the same

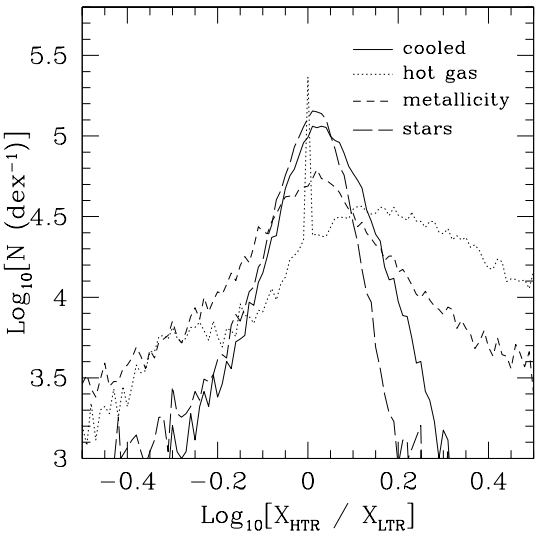

Figure 16. The difference between several halo properties (computed on an individual basis) between haloes in the full resolution (HTR) run and haloes in the run only using half the number of time-steps (LTR), as explained in Section 9.2.

in the two simulations, but that a large number of haloes containing small amounts of gas are discrepant, and their precise hot gas mass is sensitive to the time resolution chosen. It is interesting to note that this effect does not have a noticeable impact on the metallicity of the hot gas, shown by the short-dashed line in Fig. 16: the metallicity difference is peaked close to zero, showing that whatever mechanism is responsible for losing gas in the LTR run also loses a similar fraction of metals. With the long-dashed line, we show the difference in total stellar mass for each halo. Happily, this is peaked close to zero, so it seems that time resolution has little effect on the total stellar masses.

In Fig. 17 we present three curves comparing luminosities between the two runs. The upper curve is the *B*-band magnitude difference between the brightest galaxy in the halo (BCG) in the two different runs, for haloes of virial mass less than $3 \times 10^{12}\ M_\odot$, although the majority of haloes in this mass range in fact only contain a single galaxy. It can be seen that this is peaked to the left of zero, implying brighter BCGs (by around 0.2 mag) in the HTR run. This is despite the fact that we know from above that there is no offset in the total mass of stars between the two runs. Since it is the *B*-band magnitudes we are comparing, which are dominated by contributions from the youngest stars, this implies that the long time-step in the LTR smooths out the star formation over too long a period. The second (dotted) curve compares the BCGs in the other, more massive haloes. The offset is reduced (≈0.1 mag) so this effect is less important for massive haloes. In the lower curve, we show the difference in total halo magnitude when summed over all galaxies in the halo except the brightest one. Again, there is more luminosity in the HTR run. In fact, the final time-step in the HTR run contains a total of 6355 satellite galaxies, as opposed to 5096 in the LTR. This is because galaxies are lost due to more-than-binary merging in the LTR run, as explained at the beginning of this section. This 20 per cent difference in galaxy numbers is largely responsible for the 10 per cent difference in satellite luminosity in this plot.

In summary, there would have been little difference had we run our simulation with worse time resolution, except for the depletion of low-mass satellites and slightly fainter field galaxies. Since these effects are all of the order of 0.1–0.2 mag, which is quite

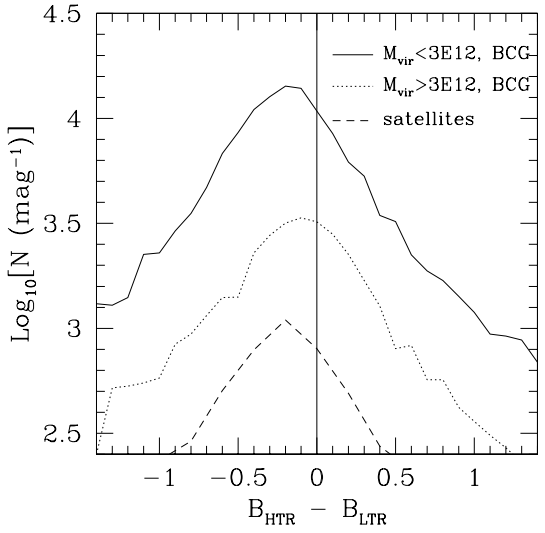

Figure 17. Differences in *B*-band magnitudes for individual haloes in the full and half time resolution runs. The upper (solid) line is the magnitude difference of the brightest halo member (BCG) for haloes with mass less than 3×10^{12} $M_\odot$. The middle (dotted) line is the same statistic for the other, more massive haloes. The lower (dashed) line is the difference in satellite magnitude (i.e. summed over all galaxies in the halo excluding the brightest one) for each halo having more than one member galaxy. A vertical line has been plotted at zero difference to guide the eye.

small compared with the systematic errors due to uncertainties in the modelling techniques, we conclude that our time resolution is close to convergence. Full convergence could be achieved with better time resolution, though we have chosen in the previous parts of this paper to gain improvement by allowing the code to cope with multiple merging of haloes.

9.3 Mass resolution

There are two ways in which mass resolution enters our results. Initially, in the *N*-body simulations, there is a resolution limit set by the finite mass of each particle. Subsequently, when identifying dark matter haloes, there is a minimum number of particles needed. We will consider the impact of both these effects.

9.3.1 N-body resolution

To investigate the effect of *N*-body resolution, one would ideally run two simulations of different resolution and compare the results. Although we have runs of 64^3 and 128^3 particles for the same box size and initial conditions as the standard (256^3) simulation, there is very little overlap in the halo mass function between them, making a direct comparison rather unproductive. Instead of comparing the simulations, we degrade the resolution of the HMR run by selecting, at random, half of the particles, assigning them double their original mass. We then run the halo-finding, tree-building and galaxy formation codes on this LMR (degraded) simulation. In Fig. 18 we show how the luminosity functions (in the *B* band and *IRAS* 60-μm filter) of these simulations differ. For faint galaxies, the LMR run underestimates the number density by a factor of ≈2.5. For bright galaxies, the estimates converge. It is the behaviour between these two regimes that is especially interesting. The vertical line in the plot is the formal resolution limit ($M_B - 5\log_{10}h < -18.8$ or $M_{60\,\mu m} - 5\log_{10}h < 22.9$, see Section 8.1) of the LMR run, and it will be seen that it is right in the middle of the transition between these two regimes, and that the LMR simulation does not fully converge with the HMR until approximately three-quarters of a magnitude brighter than this limit. As stated in Section 8.1, we expect to have problems for slightly brighter galaxies than this limit, since we cannot accurately follow the merging histories of the lightest objects.

Assuming that this behaviour is similar whatever the resolution limit, we have therefore shown that our galaxy 'selection' is complete for galaxies around three-quarters of a magnitude brighter than the formal resolution limit. Hence, our results are reliable for galaxies down to 1 mag fainter than $L_\star$.

This is not a direct test of the effects of *N*-body resolution, since we have compared with a degraded version of the same simulation rather than a simulation actually run with fewer particles. In order to check the validity of this degrading approach, we sample the 256^3 simulation at a much lower rate, selecting one particle in 4^3, thus giving it the same resolution as the 64^3 simulation. Making a similar comparison, we find no significant differences in the luminosity functions. Having verified this approach with a factor of 64, it is reasonable to expect that sampling with a factor of 2 will produce a good imitation of a simulation run with half the mass resolution.

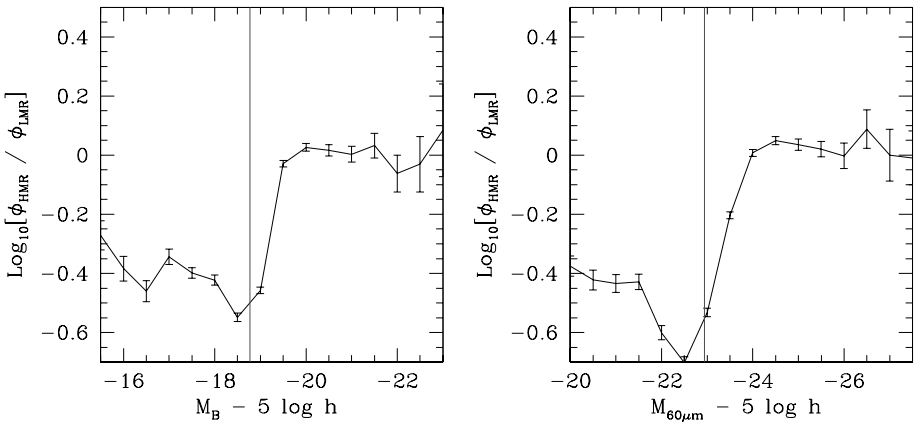

Figure 18. Differences in the *B*-band (left-hand panel) and *IRAS* 60-μm (right-hand panel) luminosity functions for the full- and half-resolution runs. The vertical line is plotted at the estimated resolution limit of the LMR simulation (approximately 0.75 mag brighter than values for the HMR simulation quoted in Table 3).

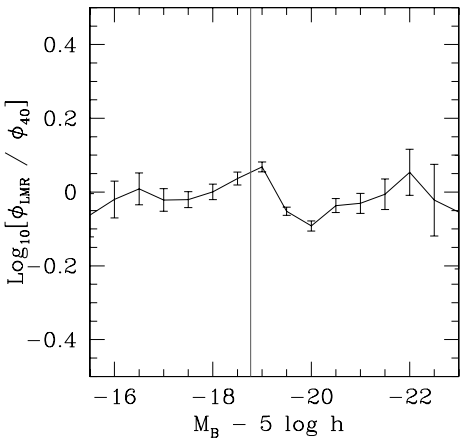

Figure 19. As in Fig. 8, but this time we compare the luminosity function from a run with a minimum particle number per halo of 40, with that of the half-resolution sampled run.

9.3.2 Minimum particle number

Increasing the minimum particle number for haloes will mean that the building blocks of our galaxy formation process are more massive, so the effect should be similar to that of degrading the mass resolution of the simulation. However, the haloes will be better resolved in this new case, so there may be some differences. To investigate this, we compare the LMR (degraded) simulation of Section 9.3.1 with the HMR simulation when the minimum particle number is increased from 20 to 40. In Fig. 19 we compare the *B*-band luminosity functions of galaxies in these two runs. As can be seen, there is no significant difference between the two.

The overall conclusion is that the limit of our resolution, in terms of noticeable effects in the local Universe, depends only on the mass of the smallest haloes identified, and is not individually sensitive to the actual numerical resolution of the simulation or the number of particles per halo.

This also suggests that it would be valid to improve the resolution of these runs by extending the treatment to haloes with less than 20 particles in them, for example going down to as few as seven, the limit quoted by KCDW for dynamical stability. This alone should result in roughly a 1 mag improvement in resolution.

10 DISCUSSION

10.1 Summary

In this paper, we have presented our *N*-body/semi-analytic hybrid model for studying galaxy formation, and demonstrated its application to a reference model within a ΛCDM cosmology. Here, we summarize the assumptions and weaknesses of the model, emphasize some of its results, and comment on several aspects of the programme.

10.1.1 Implementing new recipes

At its root, our approach is entirely independent of pre-existing work as we use different *N*-body codes and different methods for tracing the history of dark matter haloes. However, most of the recipes used to model physical processes (e.g. gas cooling, star formation, feedback, dynamical friction) are similar to those that have appeared elsewhere. We include a number of innovations that have not previously been used in hybrid models.

(i) We have measured the halo properties directly from groups identified in the *N*-body simulation, rather than assuming distributions, for example, for the spin parameter.

(ii) We have noticed that the friend-of-friend group finder identifies a number of (mainly low-mass) haloes that are not bound systems. We have dealt with these objects by suppressing cooling in them. If they are simply random collections of particles, misidentified, they will thus not affect our results.

(iii) In general, we have endeavoured to use 'natural', continuous models as opposed to the all-or-nothing methods applied by other workers. Examples include the model for halo reservoirs, where metals and hot gas ejected from the halo are retained in the local environment and gradually re-accreted, and the recipe for merging, where the effects are a smooth function of the mass ratio of the progenitors, rather than the discontinuous step function.

(iv) We have modelled disc instability as another possible mechanism for the formation of bulges.

(v) We allow for mergers between satellite galaxies in clusters, and gradual tidal stripping of dark matter cores of subhalos. These follow schemes that have been applied to purely semi-analytic models by SP99.

(vi) We use the detailed chemical and stellar evolution models described in DGS, giving us self-consistent, metallicity-dependent spectra and cooling. We do not make the common assumption of instantaneous recycling of metals.

A few key weaknesses of our recipes present themselves as follows.

(i) The process used to subdue cooling in massive haloes, and in small haloes after recent mergers, is critical to avoid overcooling on to the central object, and hence overly bright central galaxies. Our approach has been to treat this problem phenomenologically but to point out in detail where it affects our final results. The main reason behind this choice being that, whilst an approach such as that of CLBF can be qualitatively justified, and is perhaps more reasonable than the simple approximation of forgetting about cooling in haloes above a certain circular velocity (KCDW), there is no quantitative or observational justification for it.

(ii) The treatment of the interactions between dark matter and baryons is simplistic. Clearly, this problem needs to be tackled by high-resolution hydrodynamical plus *N*-body simulations. We plan to undertake a detailed study of the impact of various astrophysical processes on this interaction in the future. In particular, this will affect our predictions for correlations involving the dynamics of galaxies (Tully–Fisher, Faber–Jackson and fundamental plane).

(iii) The modelling of dynamical friction relies on Chandrasekhar's formula, which is correct on average but does not reflect the diversity of cases that appear in numerical simulations (Springel et al. 2001).

(iv) Whilst we estimate the radial positions of galaxies inside haloes, the angular coordinates are not determined in this model.

In particular, these points draw attention to the need for more detailed/accurate modelling of what actually occurs to both gas and galaxies within cluster environments.

10.1.2 Addressing the optical/IR luminosity budget of galaxies

We have introduced an efficient technique for modelling the distribution, composition and effects of dust in order to provide a full,

panchromatic view of galaxy formation and evolution processes. For this purpose, we use the ingredients of the STARDUST model of chemical and spectrophotometric evolution of stellar populations (DGS). This model was specifically designed to be implemented within a semi-analytic model of galaxy formation and evolution (Devriendt & Guiderdoni 2000). The star formation rate and cold gas metallicity histories are followed along the merging history trees of galaxies, and, once an IMF is chosen, the model gives theoretical spectra of the stellar populations. Given the galaxy radius and column density, the extinction and the IR/submillimetre spectral energy distribution are computed without additional free parameters. Computing the IR properties of galaxies is particularly important in the picture of hierarchical galaxy formation since observations show that mergers radiate mostly in the IR.

10.1.3 Studying resolution effects

We provide some simple tests of the sensitivity of our results to time and mass resolution. This enables us to put confident bounds on the regime where the results should be trusted. This is a very important step in the development of hybrid modelling, and resolution effects should be studied in more detail. The actual limit due to the baryonic mass of the lightest haloes makes our data incomplete fainter than $M_B = -18.8$ at $z = 0$. This limitation is well understood and we simply have to appreciate that our predictive power is restricted to volume-limited samples above this magnitude. However, the inability to resolve merging histories of small galaxies within this limit is a subtler problem, and requires more work to deal with. For instance, we attribute the underestimate of the metallicity of the intracluster medium to the absence of dwarf galaxies. Moreover, the results of Section 9.3 suggest that we are only completely accurate down to galaxies around three-quarters of a magnitude brighter than this $M_B = -18.8$ limit.

10.1.4 Mimicking selection effects

Our models produce a wide diversity of galaxies, but observational data only probe the emerging part of this iceberg. We have stressed several times in this paper the need to understand observational selection effects when comparing models to data. Examples include the Malmquist bias in using the APM catalogue to measure the relative abundance of different galaxy types, and the surface brightness selection that is generally present in all astrophysical data sets. We have roughly tested the influence of these selection effects on the comparison of the theoretical and observational samples in these two cases. As a result, we manage to imitate the data sets in a plausible way. This deserves more work, and will be addressed in forthcoming papers.

10.2 Main results

While we have the flexibility to alter the free parameters governing the star formation rate, feedback efficiency, etc, we find that our observationally motivated initial choice of reference parameters is well able to match a number of observational constraints.

(i) We find a wide variety of galaxy properties, among which the Milky Way appears to have 'normal' properties for its circular velocity, in terms of baryonic mass, ISM metallicity, star formation rate and size.

(ii) The luminosity functions in the optical and IR are well reproduced by the model, above the resolution limit.

(iii) The optical/IR luminosity budget in $z = 0$ galaxies is well reproduced by the model, as it appears from the fit of the luminosity functions around $L_\star$ at various optical/IR wavelengths, and from the existence of an increasing IR/optical luminosity ratio with IR luminosity as observed in the *IRAS* data. This is particularly interesting as the optical/IR properties of galaxies are closely linked to the process of hierarchical galaxy formation, since mergers radiate most of their luminosity in the IR. These first results are thus very encouraging.

(iv) Because the definition of morphological types in usual catalogues of bright galaxies relies on a mix of 'objective' and somewhat 'subjective' classification criteria, it is not easy to compare the predictions of this kind of model with data. With our simple choice of the bulge-to-disc luminosity ratio as a tracer of the type, we fairly reproduce the type fractions in a volume-limited sample. We also reproduce the overall behaviour of the colour-type correlation, and find that the $B - V$ colour monotonically reddens from the late to the early types. However, we predict too slow an evolution when going from a morphological type to another, and therefore $B - V$ colours, which are too blue for the S0s/ellipticals. We are also unable to model the transition of the low-luminosity spirals to the 'dwarfs', which is expected to occur around $M_B = -17$ (well below our resolution limit).

(v) The sizes of high surface-brightness discs are in the same range as the observational data, but we predict slightly too many large bright discs compared with the number observed. We recall that the sizes are computed under the assumption of specific angular momentum conservation during the dissipative collapse of baryons, and in the absence of dynamical effects of the feedback mechanism on the structure of the galaxies. Given the crudeness of the recipes which are used, we consider that this fit is still quite satisfying.

(vi) We obtain an arguably decent match of fundamental structural properties: the Tully–Fisher relation for the spiral galaxies and the Faber–Jackson relation for the ellipticals. We compute the dispersion in these relations, and find good agreement for the TF of cD galaxies, but the scatter is too high and the slope of the relation itself too shallow for cluster satellites and field galaxies. The FJ scatter and slope for cluster early-type galaxies are in better agreement with those obtained from the data.

(vii) Finally, we obtain a fair fit to the fundamental plane in two wavebands, reproducing the observed parameters and the scatter. This is the first time semi-analytic models have been applied to look at this important observational test.

Thus the impression that emerges from this first set of results confirms the overall robustness of the main statistical features predicted for local galaxies, with respect to previous works involving semi-analytic models or the hybrid approach, although the details of the current implementation in GALICS are fully new and original. Hierarchical galaxy formation does provide us with a natural explanation of many patterns of our local Universe, at least qualitatively: the rough percentage of each morphological type along the Hubble sequence, the fact that early-type galaxies are redder on average, the correlation between luminosity and a characteristic internal velocity, the existence of an upper limit for the luminosities of galaxies and the overall luminosity budget between the optical and the IR. These patterns appear very naturally from the interplay of a limited number of physical processes.

The quantitative fit to local galaxies is generally satisfactory, given the crudeness of the assumptions, but some failures remain. More interesting perhaps is the inner consistency of these successes and failures. The massive galaxies seem to rotate too fast and be

too blue, with a lot of ellipticals still having discs. This can be attributed to an improper modelling of cooling and/or to possible changes in the IMF, but it can be anticipated that a modification of the recipes will have simultaneous consequences on these three features. More work is needed from our reference model to clarify this point.

The well-known influence of selection effects on the observed properties of galaxies receives a new emphasis from hierarchical galaxy formation. This theory generically predicts that galaxies scatter within a wide range of properties, and that observations always unveil only the tip of the iceberg. Such a pattern considerably complicates the comparison of predictions with published data for which the exact observational selection criteria are not always fully documented. Large galaxy samples produced by extensive surveys, with a careful control of selection effects, should provide us with a better basis for the comparison with the predictions of the models.

Other papers in the series will continue to detail the successes and failures of the current model at higher redshifts, so that a consistent view on the evolution of most of the patterns can be obtained.

11 CONCLUSIONS

The semi-analytic approach has often been criticized for the number of free parameters it implies, and the differences between the various models. The impression after this fully original work here, when it is compared, for example, to KCDW, is the overall robustness of many results, despite differences in the details of the recipes that are used to model the astrophysical processes, and to handle haloes and galaxies.

After a brief description of the *N*-body simulation, and the way we build halo merging history trees, we have detailed the semi-analytic recipes that are used to model the baryons. We have then shown a series of predictions for galaxy properties in the local Universe coming from a reference model. The model produces a wide range of properties of the local galaxy population, and gives reasonable fits to the optical/IR luminosity functions, the correlation of the IR/optical luminosity ratio with IR luminosity, the morphological fractions and the colour–morphology relation. Satisfactory fits of the disc sizes, the Tully–Fisher and Faber–Jackson relations, and the fundamental plane, are also obtained. We have tried to test the influence of various selection effects on the comparison of predictions to data. We have also shown that the influence of time and mass resolution on the results can be estimated and should not affect our conclusions.

One of the ambitions of the GALICS research programme is to make predictions for a full set of observational data, at various redshifts. Clearly, such a programme will produce a host of results that will be covered by several papers in this series. Consequently, it is too early to form a full judgement of the ability of the model to capture the main statistical properties of galaxies, and their time evolution. In this first paper, we have presented our GALICS hybrid model for the study of galaxy formation. Forthcoming papers will complement the picture given here. In Paper II, we will study the influence of the cosmological and astrophysical parameters on our results, and we will predict the redshift evolution of a set of statistical galaxy properties. Paper III will be devoted to Lyman-break galaxies. Paper IV will present multiwavelength faint counts and angular correlation functions, and the construction of mock images that are the most fundamental basis for the study of selection effects. In Paper V, we will study the 3D correlation function, and examine the behaviour of the bias parameter for a variety of galaxy samples as a function of scale and time. After this global overview, subsequent papers will address more specific issues.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support extended by the IFCPAR under contract 1710-1. They thank J. Blaizot, B. Lanzoni, G. Mamon and F. Stoehr for help and criticism throughout this project and the anonymous referee and Simon White for their very careful reading of this manuscript and their constructive remarks. The *N*-body simulation used in this work was run on the Cray T3E at the IDRIS supercomputing facility. SJH acknowledges support of the EU TMR network 'Formation and Evolution of Galaxies'. The research of JEGD at Oxford has been supported by a major grant from the Leverhulme Trust.

REFERENCES

Appel A., 1981, PhD thesis, Princeton Univ.
Appel A., 1985, SIAM J. Sci. Stat. Comput., 6, 85
Bardeen J., Bond J., Kaiser N., Szalay A., 1986, ApJ, 304, 15
Barnes J.E., 1990, J. Comp. Phys., 87, 161
Barnes J., Efstathiou G., 1987, ApJ, 319, 575
Barnes J., Hut P., 1986, Nat, 324, 446
Baugh C.M., Cole S., Frenk C.S., 1996, MNRAS, 283, 1361
Baugh C., Cole S., Frenk C., Lacey C., 1998, ApJ, 504
Benson A.J., Cole S., Frenk C.S., Baugh C.M., Lacey C.G., 2000, MNRAS, 311, 793
Binney J.J., Evans N.W., 2001, MNRAS, 327, L27
Binney J., Tremaine S., 1987, Galactic Dynamics. Princeton Univ. Press, Princeton, p. 747
Blumenthal G.R., Faber S.M., Primack J.R., Rees M.J., 1984, Nat, 311, 517
Bond J.R., Kaiser N., Cole S., Efstathiou G., 1991, ApJ, 379, 440
Bouchet F.R., Hernquist L., 1988, ApJ Suppl., 68, 521
Bower R.G., 1991, MNRAS, 248, 332
Bruzual A.G., Charlot S., 1993, ApJ, 405, 538
Bruzual G., 1981, Spectral Evolution of Galaxies. Univ. California Berkeley
Buta R., Mitra S., de Vaucouleurs G., Corwin H.G., 1994, AJ, 107, 118
Carilli C.L. et al., 2000, ApJ, 533, L13
Cavaliere A., Fusco-Femiano R., 1976, A&A, 49, 137
Cole S., Aragon-Salamanca A., Frenk C., Navarro J., Zepf S., 1994, MNRAS, 271, 781
Cole S., Lacey C., 1996, MNRAS, 281, 716
Cole S., Lacey C.G., Baugh C.M., Frenk C.S., 2000, MNRAS, 319, 168 (CLBF)
Cole S., Norberg P., Baugh C.M., Frenk C.S., The 2dF GRS Team, 2001, MNRAS, submitted (astro-ph/0012429)
Cross N., Driver S., Couch W., The 2dF GRS Team, 2001, MNRAS, accepted (astro-ph/0012165)
Dalcanton J.J., Spergel D.N., Summers F.J., 1997, ApJ, 482, 659
David L.P., Jones C., Forman W., 1996, ApJ, 473, 692
Davis M., Efstathiou G., Frenk C.S., White S.D.M., 1985, ApJ, 292, 371
de Jong R.S., Lacey C., 2000, ApJ, 545, 781
Dekel A., Silk J., 1986, ApJ, 303, 39
Désert F.X., Boulanger F., Puget J.L., 1990, A&A, 237, 215
Devriendt J.E.G., Guiderdoni B., 2000, A&A, 363, 851
Devriendt J.E.G., Guiderdoni B., Sadat R., 1999, A&A, 350, 381 (DGS)
Djorgovski S., Davis M., 1987, ApJ, 313, 59
Domínguez-Tenreiro R., Tissera P.B., Sáiz A., 1998, ApJ, 508, L123
Draine B.T., Lee H.M., 1984, ApJ, 285, 89
Dressler A., Lynden-Bell D., Burstein D., Davies R.L., Faber S.M., Terlevich R., Wegner G., 1987, ApJ, 313, 42
Efstathiou G., Frenk C.S., White S.D.M., Davis M., 1988, MNRAS, 235, 715
Eke V.R., Cole S., Frenk C.S., 1996, MNRAS, 282, 263
Eke V.R., Navarro J.F., Frenk C.S., 1998, ApJ, 503, 569
Elbaz D. et al., 1999, A&A, 351, L37

Faber S.M., Jackson R.E., 1976, ApJ, 204, 668
Fall S.M., Efstathiou G., 1980, MNRAS, 193, 189
Fixsen D.J., Dwek E., Mather J.C., Bennett C.L., Shafer R.A., 1998, ApJ, 508, 123
Forbes D.A., Ponman T.J., 1999, MNRAS, 309, 623
Giovanelli R., Haynes M.P., Herter T., Vogt N.P., da Costa L.N., Freudling W., Salzer J.J., Wegner G., 1997, AJ, 113, 53
Governato F., Ghigna S., Moore B., Quinn T., Stadel J., Lake G., 2001, ApJ, 547, 555
Granato G.L., Lacey C.G., Silva L., Bressan A., Baugh C.M., Cole S., Frenk C.S., 2000, ApJ, 542, 710
Gross M.A.K., Somerville R.S., Primack J.R., Holtzman J., Klypin A., 1998, MNRAS, 301, 81
Guiderdoni B., Rocca-Volmerange B., 1987, A&A, 186, 1
Guiderdoni B., Bouchet F.R., Puget J., Lagache G., Hivon E., 1997, Nat, 390, 257
Guiderdoni B., Hivon E., Bouchet F.R., Maffei B., 1998, MNRAS, 295, 877
Hauser M.G. et al., 1998, ApJ, 508, 25
Hernquist L., 1987, ApJ Suppl., 64, 715
Hernquist L., 1990, ApJ, 356, 359
Hernquist L., Bouchet F.R., Suto Y., 1991, ApJ Suppl., 75, 231
Huang J.-S., Glazebrook K., Cowie L.L., Tinney C., 2003, ApJ, 584, 203
Jernigan J.G., 1985, IAU Symposia, 113, 275
Kauffmann G., 1995, MNRAS, 274, 153
Kauffmann G., 1996a, MNRAS, 281, 487
Kauffmann G., 1996b, MNRAS, 281, 475
Kauffmann G., Charlot S., 1998, MNRAS, 297, L23
Kauffmann G., Haehnelt M., 2000, MNRAS, 311, 576
Kauffmann G., White S.D.M., 1993, MNRAS, 261, 921
Kauffmann G., White S., Guiderdoni B., 1993, MNRAS, 264, 201
Kauffmann G., Guiderdoni B., White S.D.M., 1994, MNRAS, 267, 981
Kauffmann G., Nusser A., Steinmetz M., 1997, MNRAS, 286, 795
Kauffmann G., Colberg J.M., Diaferio A., White S.D.M., 1999, MNRAS, 303, 188 (KCDW)
Kennicutt R.C., 1983, ApJ, 272, 54
Kennicutt R.C., 1989, ApJ, 344, 685
Kennicutt R.C., Jr, Tamblyn P., Congdon C.E., 1994, ApJ, 435, 22
Kent S.M., 1981, ApJ, 245, 805
Kent S.M., Dame T.M., Fazio G., 1991, ApJ, 378, 131
Koyama K., Takano S., Tawara Y., 1991, Nat, 350, 135
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lacey C., Cole S., 1994, MNRAS, 271, 676
Lacey C., Silk J., 1991, ApJ, 381, 14
Lacey C., Guiderdoni B., Rocca-Volmerange B., Silk J., 1993, ApJ, 402, 15
Loveday J., 1996, MNRAS, 278, 1025
Loveday J., Peterson B.A., Efstathiou G., Maddox S.J., 1992, ApJ, 390, 338
Loveday J., Peterson B.A., Maddox S.J., Efstathiou G., 1996, ApJS, 107, 201
Lucy L.B., Danziger I.J., Gouiffes C., Bouchet P., 1989, Supernovae. Springer-Verlag, Berlin, p. 82
Ma C.P., Bertschinger E., 1995, ApJ, 455, 7
Magorrian J. et al., 1998, AJ, 115, 2285
Makino J., 1990, J. Comp. Phys., 87, 148
Makino J., Hut P., 1997, ApJ, 481, 83
Mamon G.A., 1992, ApJ, 401, L3
Mamon G.A., 1995, in de Vega H.N.S., eds, 3rd Paris Cosmology Coll. World Scientific, Singapore, p. 95 (astro-ph/9511101)
Mamon G.A., 2000, in ASP Conf. Ser. Vol. 197, Dynamics of Galaxies: from the Early Universe to the Present. Astron. Soc. Pac., San Francisco, p. 377
Martin C.L., Kobulnicky H.A., Heckman T.M., 2002, ApJ, 574, 663
Mathis J.S., Mezger P.G., Panagia N., 1983, A&A, 128, 212
Mazzei P., Xu C., de Zotti G., 1992, A&A, 256, 45
Mo H.J., Mao S., White S.D.M., 1998, MNRAS, 295, 319
Navarro J.F., Frenk C.S., White S.D.M., 1995a, MNRAS, 275, 56
Navarro J.F., Frenk C.S., White S.D.M., 1995b, MNRAS, 275, 720
Ninin S., 1999, PhD thesis, Université Paris 11
Nulsen P.E.J., Fabian A.C., 1995, MNRAS, 277, 561
Osterbrock D.E., 1989, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei. University Science Books Mill Valley
Pagel B.E., 1998, Nucleosynthesis and Chemical Evolution of Galaxies. Cambridge Univ. Press Cambridge
Pahre M.A., Djorgovski S.G., 1997, in ASP Conf. Ser. Vol. 116, The Nature of Elliptical Galaxies, 2nd Stromlo Symp. Astron. Soc. Pac., San Francisco, p. 154
Pearce F.R., Jenkins A., Frenk C.S., White S.D.M., Thomas P.A., Couchman H.M.P., Peacock J.A., Efstathiou G., 2001, MNRAS, submitted (astro-ph/0010587)
Peebles P., 1982, ApJ, 263, L1
Porter D.H., 1985, PhD thesis, California Univ., Berkeley
Prantzos N., Aubert O., 1995, A&A, 302, 69
Press W.H., Schechter P., 1974, ApJ, 187, 425
Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., 1992, Numerical Recipes in FORTRAN, 2nd edn. Cambridge Univ. Press, Cambridge
Puget J., Abergel A., Bernard J., Boulanger F., Burton W.B., Desert F., Hartmann D., 1996, A&A, 308, L5
Puget J.L. et al., 1999, A&A, 345, 29
Quilis V., Moore B., Bower R., 2000, Sci, 288, 1617
Renzini A., 1999, in Rosa M.R., eds, Chemical evolution from zero to high redshift, Proc. ESO workshop held at Garching, Germany, 14–16 October 1998/Jeremy R. Walsh. Springer-Verlag, Berlin (ESO Astrophysics Symposia), p. 185 (astro-ph/9902361)
Renzini A., 2000, in Plionis M., Georgantopoulos I., eds, Large Scale Structure in the X-ray Universe, Proc. 20–22 September 1999 Workshop, Santorini, Greece. Atlantisciences, Paris, p. 103
Roberts M.S., 1963, ARA&A, 1, 149
Roukema B.F., Peterson B.A., Quinn P.J., Rocca-Volmerange B., 1997, MNRAS, 292, 835
Salpeter E.E., 1955, ApJ, 121, 161
Scalo J.M., 1986, Fund. Cosmic Phys., 11, 1
Scodeggio M., Gavazzi G., Belsole E., Pierini D., Boselli A., 1998, MNRAS, 301, 1001
Silk J., 2001, MNRAS, 324, 313
Silk J., 2003, MNRAS, submitted
Silva L., Granato G.L., Bressan A., Danese L., 1998, ApJ, 509, 103
Simien F., de Vaucouleurs G., 1986, ApJ, 302, 564
Smail I., Ivison R.J., Blain A.W., 1997, ApJ, 490, L5
Soifer B.T., Neugebauer G., 1991, AJ, 101, 354
Soifer B.T., Sanders D.B., Madore B.F., Neugebauer G., Danielson G.E., Elias J.H., Lonsdale C.J., Rice W.L., 1987, ApJ, 320, 238
Solanes J., García-Gómez C., González-Casado G., Giovanelli R., Haynes M., 2001, ApJ, 97
Somerville R.S., Kolatt T.S., 1999, MNRAS, 305, 1
Somerville R.S., Lemson G., Kolatt T.S., Dekel A., 2000, MNRAS, 316, 479
Somerville R.S., Primack J.R., 1999, MNRAS, 310, 1087 (SP99)
Somerville R.S., Primack J.R., Faber S.M., 2001, MNRAS, 320, 504
Springel V., White S.D.M., Tormen G., Kauffmann G.A.M., 2001, MNRAS, 328, 726
Steinmetz M., Bartelmann M., 1995, MNRAS, 272, 570
Suginohara T., Suto Y., Bouchet F.R., Hernquist L., 1991, ApJ Suppl., 75, 631
Sugiyama N., 1995, ApJ Suppl., 100, 281
Sutherland R.S., Dopita M.A., 1993, ApJ Suppl., 88, 253
Toomre A., Toomre J., 1972, ApJ, 178, 623
Tully R.B., Fisher J.R., 1977, A&A, 54, 661
Tytler D., Fan X.-M., Burles S., 1996, Nat, 381, 207
van den Bosch F.C., 1998, ApJ, 507, 601
Walker I.R., Mihos J.C., Hernquist L., 1996, ApJ, 460, 121
White M., 2001, A&A, 367, 27
White R.E., Day C.S.R., Hatsukade I., Hughes J.P., 1994, ApJ, 433, 583
White S.D.M., Davis M., Efstathiou G., Frenk C.S., 1987, Nat, 330, 451
White S.D.M., Frenk C.S., 1991, ApJ, 379, 52
White S.D.M., Rees M.J., 1978, MNRAS, 183, 341
Wu K.K.S., Fabian A.C., Nulsen P.E.J., 2000, MNRAS, 318, 889

APPENDIX A: THE TREECODE

A1 The hierarchical approach

When computing the gravitational potential acting on a particle, it is possible to ignore the details of the internal structure of a group of distant particles. Replacing several particle–particle interactions by a one particle–distant group interaction is much more efficient than a particle–particle code, and one can still control the accuracy of the computations. This is the key simplification of the hierarchical approach.

The first codes of this kind were written by Appel (1981, 1985), Jernigan (1985) and Porter (1985), who were trying to follow the spatial distribution of mass as closely as possible. Barnes & Hut (1986) used a simpler algorithm, employing a hierarchical decomposition of space, allowing them to measure the accuracy of the force estimation. This code was vectorized by Hernquist (1987), Makino (1990) and Barnes (1990) and checked on vectorial architectures such as the Cray X-MP. The adaptation for cosmological cases, especially the adding of periodic boundary conditions (using the Ewald summation method) was performed by Bouchet & Hernquist (1988), Hernquist, Bouchet & Suto (1991) and Suginohara et al. (1991).

The tree-code uses a multipolar expansion to compute the interaction of a particle and a group of particles. We can do this if the size of the group is small enough compared with the distance of the particle to this group: a variety of different criteria are in practice used for that condition.

This concept is implemented by a hierarchical division of space, usually called a tree, which, in our case is a binary tree. The tree is constructed by starting from the entire box and dividing the space along one direction (say x first) in two equal cells. Each of those cells is divided along another direction (y), and this process (division along x, y, z, alternately) continues until each cell has one or zero particles inside.

The computation of the force for one particle starts from the 'root' (that is the entire simulation box) of the tree. The selection criterion depending on the size of the cell as compared with the particle–cell distance is applied, to determine whether we can replace the cell by its centre of mass. If so, we add to the 'interaction list' the term corresponding to this interaction. If not, the same check is performed, recursively, on subcells, until each term of the force is computed.

This is the basic idea for the 'sequential' version of the code. We now describe the implementation of our parallel approach.

A2 Parallelization

Our code is written specifically for a T3D or T3E Cray architecture, which has a distributed memory type. Fast Cray libraries such as SHMEM allow us to use shared memory routines, such as remote reads or remote writes, which are rather faster than the usual message passing routines of PVM or even MPI. The code could be easily translated into MPI-2 routines.

At each time-step, a global sort of particles is performed on all processors. This global sort is based on four keys, which are defined by interleaving the binary digits of the x-, y- and z-positions of particles, in such a way that two physically close particles are, most probably, on the same processor, except for particles at the borders. This sorting operation thus amounts to a 'pre-building' of the tree, and also helps to decrease communications between processors during the force computation. The tree is then built, simultaneously on all processors. The binary tree is a shared one, with its root on processor 0. To decrease communications, cells that contain particles on the same processor will also be on this processor, most of the time; that is to say, going down along branches of the tree, we find that cells tend to be on the same processors as the particles they 'control'.

The code also takes into account periodic boundary conditions by using the Ewald summation method, and copying the grid of periodic replicas on to each processor.

The force computation is finally performed in a parallel way using the shared memory library of the Cray T3E. Once forces are computed, we can derive accelerations, and then update particle velocities and positions, using a standard integration scheme.

We describe a few finer details of the code below.

(i) *Sharing of particles on nodes.* First of all, to balance load, each processor is assigned approximately the same number of particles. More precisely, if $N_{\rm part}$ is the total number of particles and $N_{\rm procs}$ is the number of processors, we perform the Euclidean division

$$N_{\rm part} = q N_{\rm procs} + r. \tag{A1}$$

Each processor has at least q particles and the first r processors have an additional one.

(ii) *Sorting particles.* Each particle is then assigned a key, composed of four integer numbers. These keys are based on the idea of the bijection between R^3 and R. For computers, this is actually a bijection between N^3 and N, because of the limited precision of real numbers. The building and sorting of particles is essentially based on Cantor's construction of this bijection, which we briefly review here. We obtain the first three integers by interleaving the binary digits of the three spatial coordinates of the particles, where all the coordinates are normalized between 0 and 1. Let those coordinates be written as, for example: $x = 0.x_1\,x_2\,x_3\,x_4\,x_5\,x_6\,x_7\,x_8\,x_9$, $y = 0.y_1\,y_2\,y_3\,y_4\,y_5\,y_6\,y_7\,y_8\,y_9$, $z = 0.z_1\,z_2\,z_3\,z_4\,z_5\,z_6\,z_7\,z_8\,z_9$. The first three integer keys will be: $k_1 = x_1\,y_1\,z_1\,x_2\,y_2\,z_2\,x_3\,y_3\,z_3$, $k_2 = x_4\,y_4\,z_4\,x_5\,y_5\,z_5\,x_6\,y_6\,z_6$, $k_3 = x_7\,y_7\,z_7\,x_8\,y_8\,z_8\,x_9\,y_9\,z_9$. The fourth number used is just the original particle number, such that in the rare event that two particles have the same position, they will not have the same keys.

The keys are used to sort particles on the processors in increasing lexicographic order. This is a global sort: after sorting, particles on processor number $n + 1$ will all have keys greater than those of the particles on processor n.

(iii) *Building the tree.* A global binary tree shared on all processors is then built. The entire box is the root of the tree. We then divide recursively all cells of the tree alternately along the (x, y, z) directions, until each cell is either empty or contains only one particle. As the tree is shared on all nodes, the top of the tree is divided on all processors. So, the beginning of the tree-building process requires a certain amount of communication between them. However, as the tree building proceeds, most of the subcells of a cell will be on the same processor (as that controlling the cell), and this will finally decrease interprocessor communication considerably.

(iv) *Tree properties.* We compute the mass and centre-of-mass positions of each cell. This is performed by starting from the lowest cells in the tree, just above the particle level. A synchronization of the computation is made between each level of the tree, because the results of the previous level are required to compute the mass and centre-of-mass position of this level.

(v) *Force computation.* The force computation is performed by starting from the top of the tree, going down recursively until each

term of the force has been computed: each cell is either 'opened' or accepted in the interaction list, if the opening criterion is satisfied.

(vi) *Particles update.* A conventional leapfrog algorithm, taking into account the expansion of the Universe, is used to update positions and velocities of particles.

(vii) *Optimization.* Several tricks have been used to optimize the code. First, we 'derecursified' most of the routines of the code. A copy of the top of the tree (the first $N \approx 12$ levels), is also saved on each node, thus decreasing communications in the tree descent. To decrease cache misses, we also modify the tree, so that the right pointer previously used for the 'right child' is instead used, in the force computation, as a 'sibling pointer': this modification goes with the derecursification of the tree descent.

APPENDIX B: HALO POTENTIAL ENERGIES

As mentioned in Section 2.3, the potential energy of large dark matter haloes is computationally expensive using a simple sum over pairs. We compare the measured potential with that from applying the formula used by Steinmetz & Bartelmann (1995) for the potential energy of an ellipsoid,

$$W = -\frac{3}{5} G M^2 R_{\rm F}(a^2, b^2, c^2), \tag{B1}$$

where $R_{\rm F}$ is Carlson's form for the elliptic integral of the first kind (see, e.g., Press et al. 1992),

$$R_{\rm F}(x, y, z) = \frac{1}{2} \int_0^\infty \frac{{\rm d}t}{\sqrt{(t+x)(t+y)(t+z)}}. \tag{B2}$$

This procedure does not correct for the non-inverse square nature of the softened gravitational potential in the simulations, but, as shown in Fig. B1, there is a good correlation between the true potential energy and the energy estimated in this way. We note from this figure that the scatter in the relation is fairly small, of the order of 5 per cent. It will also be seen that the median depends on the number of particles in the halo. It ranges from being a slight overestimate of the true potential energy at low halo mass, to a direct correspondence at around 25 particles, before flattening off towards large particle number with an underestimate of the magnitude of the potential energy of around 15 per cent.

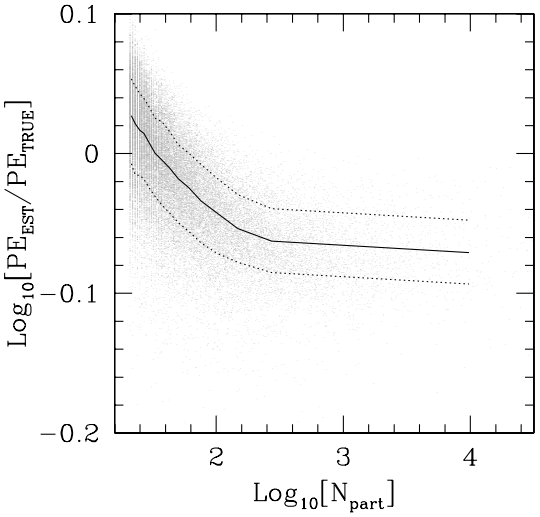

Figure B1. Relationship between the true, measured potential energy from pair counting and that found by applying equation (B1) to the mass distribution. The faint dots show the relation for each halo, while the solid line is the median and the dotted lines show the scatter at plus and minus one standard deviation.

In order to use an unbiased recipe, whilst still speeding up the computation, we apply this formula, to haloes of more than 1000 particles, using the full double-sum method for smaller haloes. The potential energies of these haloes are then accurate only to 5 per cent. Since the spin parameter depends on the square root of the total energy, this additional scatter will have little effect on the distribution of spin parameters.

It should be noted that this cut-off means that the energies are computed properly (i.e. without this approximation) for the majority of haloes (given the mass distribution of Fig. 1), and for almost all the haloes that were found to have positive total energy (since their mass function has a steeper slope).

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