PINOCCHIO and the hierarchical build-up of dark matter haloes

Giuliano Taffoni,[1]★ Pierluigi Monaco[2] and Tom Theuns[3]

[1]*SISSA, Via Beirut 2-4, 34060 Trieste, Italy*
[2]*Dipartimento di Astronomia, via Tiepolo 11, 34131 Trieste, Italy*
[3]*Institute of Astronomy, Madingley Road, Cambridge CB3 0HA*

ABSTRACT

We study the ability of PINOCCHIO (PINpointing Orbit-Crossing Collapsed HIerarchical Objects) to predict the merging histories of dark matter (DM) haloes, comparing the PINOCCHIO predictions with the results of two large N-body simulations run from the same set of initial conditions. We focus our attention on the quantities most relevant to galaxy formation and large-scale structure studies. PINOCCHIO is able to predict the statistics of merger trees with a typical accuracy of 20 per cent. Its validity extends to higher-order moments of the distribution of progenitors. The agreement is also valid at the object-by-object level, with 70–90 per cent of the progenitors cleanly recognized when the parent halo is cleanly recognized itself. Predictions are also presented for quantities that are usually not reproduced by semi-analytic codes, such as the two-point correlation function of the progenitors of massive haloes and the distribution of initial orbital parameters of merging haloes. For the accuracy of the prediction and for the facility with which merger histories are produced, PINOCCHIO provides a means to generate catalogues of DM haloes, which is extremely competitive with large-scale N-body simulations, making it a suitable tool for galaxy formation and large-scale structure studies.

Key words: galaxies: clusters: general – galaxies: formation – galaxies: haloes – cosmology: theory – dark matter.

1 INTRODUCTION

In the *hierarchical clustering scenario*, structure in the Universe forms from the aggregation and merging of smaller subunits. This theoretical picture is now substantiated, at least on a qualitative level, by a wealth of observations of the high-redshift ($z \sim 3$–5) Universe. In the most commonly discussed scenario, the 'cold dark matter' (CDM) one, hierarchical clustering is driven by the gravitational collapse of DM fluctuations, while the visible astrophysical objects are generated from baryons falling into the DM haloes (see, for example, White & Rees 1978). Thus the process of formation and evolution of DM haloes is of fundamental importance for understanding the properties of galaxies or galaxy clusters.

The formation of DM haloes involves highly non-linear dynamical processes that cannot be followed analytically. To face this problem it is necessary to resort to numerical N-body simulations. Besides this time-consuming method, one can also use analytical approximations that are able to predict with fair accuracy some relevant quantities related to the assembly of DM haloes. Moreover, the analytic methods help to shed light on the

complex gravitational problem of hierarchical clustering. The pioneers of the analytical approach were Press & Schechter (1974; hereafter PS) who derived an expression for the mass function of DM haloes. This was found to give a fair approximation of the N-body results (Efstathiou et al. 1988; see for a review Monaco 1998). The PS approach was extended by Bond et al. (1991; see also Peacock & Heavens 1990; Bower 1991; Lacey & Cole 1993) (extended PS formalism, hereafter EPS), who fixed a normalization problem of the original PS work. The EPS model can be used to also predict some properties of DM haloes, such as their formation time, survival time and merger rate. These predictions were tested against numerical simulations, again with success, by Lacey & Cole (1994). The EPS formalism has recently become a standard tool for constructing synthetic catalogues of DM haloes for galaxy formation programs (see, for example, Kauffmann, White & Guiderdoni 1993; Somerville & Primack 1999; Cole et al. 2000).

However, recent work with larger N-body simulations has revealed significant discrepancies between PS and EPS predictions and numerical results. The PS mass function has been shown to underpredict the number of massive haloes and overpredict the number of low-mass ones (Gelb & Bertschinger 1994; Governato et al. 1999; Jenkins et al. 2001; Bode et al. 2001). Similar discrepancies were observed in the reconstruction of the

★E-mail: taffoni@sissa.it

conditional mass function, i.e. the number density of haloes bound to flow into a parent halo of given mass at a subsequent time[1] (Somerville & Kolatt 1999; Sheth & Lemson 1999b). The EPS formalism is also affected by limitations, in that it does not give full information on the spatial distribution of haloes (Catelan et al. 1998; Jing 1998, 1999; Porciani, Catelan & Lacey 1999), and by inconsistencies in the use of smoothing filters and in the construction of merger trees (Somerville & Kolatt 1999; Sheth & Lemson 1999b; Cole et al. 2000). Attempts to improve this formalism, or to develop alternative ones, were reviewed by Monaco (1998). A more recent and successful extension is given by Sheth & Tormen (1999) and Sheth, Mo & Tormen (2001); their model improves significantly the fit of the mass function and the extension of dynamics to ellipsoidal collapse (EPS is based on linear theory), but does not remove the inconsistencies of the EPS approach and does not provide spatial information of haloes either. This method has also been applied to build random realizations of the merging histories of DM haloes (Sheth & Tormen 2002), but it does not provide a significant improvement with respect to the standard merger trees.

Recently, we have presented a new algorithm, called PINOCCHIO (PINpointing Orbit-Crossing Collapsed HIerarchical Objects), to generate synthetic catalogues of DM haloes with known mass, position, peculiar velocity, merger history and angular momentum (Monaco et al. 2002, hereafter Paper I; Monaco, Theuns & Taffoni 2002, hereafter Paper II). In contrast to EPS, PINOCCHIO is able both to reproduce statistical quantities, such as the mass or two-point correlation function of haloes, and to reproduce haloes on a point-by-point basis.

In this paper, we investigate in detail how well PINOCCHIO is able to recover the merger histories (or *merger trees*) of DM haloes. We compare the PINOCCHIO code with numerical *N*-body simulations and with the analytical estimates of the EPS theory. We examine the ability of PINOCCHIO to reconstruct the main statistical properties of the merger trees, extending the analysis to predict the correlation function and the initial orbital parameters of merging haloes.

Section 2 gives a brief description of the PINOCCHIO code with special attention to the extraction of the merger trees. In Section 3 we compare the statistical properties of the distribution of DM haloes at different redshifts given by PINOCCHIO with the results of numerical *N*-body simulations. Section 4 is dedicated to the study of the spatial distribution of the haloes that will form cluster-sized objects at the present time. Section 5 shows the ability of PINOCCHIO to predict the impact parameters of merging haloes. The conclusions are reported in Section 6.

2   MERGER TREES FROM PINOCCHIO

The PINOCCHIO code was presented in Paper I and described in full detail in Paper II. Here we give only a brief description of the code, necessary to discuss the procedure used to extract the merger trees.

For a given cosmological background model and a power spectrum of fluctuations, a Gaussian linear density contrast field δ_l (i.e. linearly extrapolated to $z = 0$) is generated on a cubic grid, in a way much similar to what is usually done to generate initial condition for *N*-body simulations. The linear density contrast δ_l is smoothed repeatedly with Gaussian filters of FWHM R, where R

takes values that are equally spaced (~ 20 smoothing radii usually give an adequate sampling). For each point q of the Lagrangian (initial) coordinate and for each smoothing radius R, the collapse time (i.e. the time at which the particle is predicted to enter a high-density, multistream region) is computed using Lagrangian perturbation theory (hereafter LPT; see, e.g., Catelan 1995; Bouchet 1996; Buchert 1996) and its ellipsoidal truncation (Monaco 1997). Technically, the collapse time is defined as the instant of orbit crossing (OC); this definition is discussed at length in Paper II (see also Monaco 1995, 1997). For each particle only the earliest collapse time is recorded, which amounts to recording the field

$$F_{\max}(q) \equiv \max_R \left[\frac{1}{b(t_c)} \right], \tag{1}$$

where $b(t)$ is the linear growing mode (see Padmanabhan 1993; Monaco 1998) and t_c is the OC-collapse time. Note that in an Einstein–de Sitter universe $F_{\max} = (1 + z_{\max})$, where $z_{\max}$ is the largest collapse redshift at which the particle collapses.[2]

Besides $F_{\max}$, we also record the smoothing radius $R_{\max}$ for which the maximum in equation (1) is reached, and the velocity of the particle at collapse time as given by the Zel'dovich approximation (1970) (which is the first-order term of LPT),

$$v_{\max} = -b(t)\nabla\phi(q; R_{\max}) \tag{2}$$

(in units of comoving displacement), where $\phi(q; R_{\max})$ is the rescaled peculiar gravitational potential (smoothed at $R_{\max}$), which obeys the Poisson equation $\nabla^2 \phi(q) = \delta_l$. All differentiations and convolutions are performed using fast Fourier transforms.

The collapsed medium is then 'fragmented' into isolated objects using an algorithm designed to mimic the accretion and merger events of hierarchical collapse. Collapsed particles may belong to relaxed haloes or to lower-density filaments. At the instant a particle is deemed to collapse, we decide which halo, if any, it accreted on to. The candidate haloes are those that already contain one Lagrangian neighbour of the particle (on the initial grid q of Lagrangian positions, the six particles nearest to a given one are its 'Lagrangian neighbours'). The particle will accrete on to the halo if its distance (in the Eulerian space at the collapse time) from the centre of mass of the candidate halo is smaller than a given fraction of the halo size ($R_N = N^{1/3}$ in grid units, where N is the number of particles in the halo), otherwise they are catalogued as filaments. If a particle has more than one candidate halo, we check whether these haloes should merge. The merging condition is very similar to the accretion one: two groups merge if their distance in the Lagrangian space is smaller than a fraction of the size of the largest halo. If a particle is a local maximum of $z_{\max}$ it is considered as the seed of a new halo. Finally, filament particles are accreted on to a halo when they neighbour in the Lagrangian space an accreting particle; this is done to mimic accretion of filaments on to haloes.

The fragmentation algorithm requires the introduction of free parameters, which are analogous to those required by any clump-finding algorithm applied to *N*-body simulations. These parameters specify the level of overdensity at which the halo is defined, or are introduced to fix resolution effects. They are discussed in Paper II (to which we refer for all details) and chosen by requiring the fit of the mass function of haloes selected with the friends-of-friends (FOF) algorithm with linking length 0.2 times the mean interparticle

---

[1] In the following, the 'final' haloes at $z = 0$ (or occasionally at higher redshift) will be called *parent*, while the higher-redshift haloes that flow into the parent will be called *progenitors*.

[2] Taking the largest redshift (or *F*-value) of collapse is analogous to considering the largest collapse radius, as done in the EPS formalism.

distance at $z = 0$. Here for the parameters we use the values found in Paper II.

To give a taste of the speed of PINOCCHIO with respect to simulations, a 256^3 particles realization needs ~ 6 h on a Pentium III 450 MHz computer, with a RAM requirement of ~ 512 Mb. When comparing with the corresponding simulation the statistical quantities such as the mass function or the two-point correlation function are reproduced with a typical error of ~ 10 per cent or smaller. At the object-by-object level, the accuracy of PINOCCHIO depends on the degree of non-linearity that is reached at the grid level, and degrades in time. Typically, 70–99 per cent of the objects are reproduced with an error on the mass of 30–40 per cent, an error on the position of ~ 0.5–2 grid points and a one-dimensional error on the velocity of ~ 150 km s^{-1} at $z = 0$.

It is noteworthy that merging events in PINOCCHIO are not restricted to being binary: in principle up to six objects can merge together at the same time, even if, as expected, the number of mergers that involves more than three haloes is a very small fraction of the total.

The merger histories of haloes are evaluated directly by PINOCCHIO. At each merger the largest halo retains its identification number (ID) which will become the ID of the merger, while the other haloes are labelled as expired. The mass of each halo involved in the merging event is recorded together with the redshift at which the merger takes place. For each expired (progenitor) halo we keep track at all times of the (parent) halo they are presently incorporated within. Even though accretion is defined rigorously as the entrance of a single particle into the object, the merger of a halo with another one with less than 10 particles is always considered as an accretion event.

The merger trees extracted from PINOCCHIO provide a more complete description of the merging histories of haloes than the EPS one. They not only follow the time evolution of the mass and number distribution of the progenitors, but also their distribution in space, their velocities and angular momenta.

3 STATISTICS OF THE PROGENITORS

3.1 The simulations

In order to test the ability of PINOCCHIO in predicting the statistics of the merger trees, we compared the results of two *N*-body simulations with those of PINOCCHIO applied to the same initial density field. The simulations were already presented in Papers I and II, to which we refer for all the details. They are a standard CDM model (SCDM), run with the PKDGRAV code on a large box of $500\,h^{-1}$ Mpc^3 with 360^3 particles (Governato et al. 1999), and a ΛCDM model, run with the HYDRA code (Couchman, Thomas & Pearce 1993) on a smaller box of $100\,h^{-1}$ Mpc with 256^3 particles. The reason why we use two different simulations is to check the method for different cosmologies, boxes, resolutions and codes. For the present purpose, the ΛCDM simulation is more suitable as the higher mass resolution allows one to reconstruct the merger tree to higher redshifts and lower masses, but the SCDM allows one to test the merger trees for the more massive haloes.

The haloes are identified using a standard FOF algorithm with a linking length of 0.2 times the interparticle distance. Note that, following the suggestion by Jenkins et al. (2001), we do not change the linking length with the cosmology. In this paper, we adopt 10 particles as the minimum mass of the haloes when we analyse the

conditional mass function. This is to test the effect of the degrading of the agreement at small masses. In general, at least 30 particles are necessary to identify reliably a halo both in the simulations and in PINOCCHIO, so we consider a threshold mass of 30 particles for the other statistical analysis.

The merger trees for the FOF haloes at final time z_0 are constructed as follows. Progenitors are defined as those haloes that at the higher redshift z contain some of the particles of the parent halo at z_0. As noted by some authors (see, e.g., Somerville et al. 2000), some particles that are located in a progenitor are not included later into the parent. This reflects the actual dynamics of the haloes that suffer stripping and evaporation events, and make the progenitor identification process more ambiguous. We then adopt two simple rules.

(i) If a parent halo contains less than 90 per cent of the mass of all its progenitors at redshift z, then it is excluded from the analysis (this happens in a few per cent of cases).

(ii) We assign to the progenitor the mass of all its particles that will flow into the parent at z_0.

In this way we force mass conservation in the merger tree and reject some extreme cases when the progenitor is strongly affected by these 'evaporation' effects.

Owing to the limited number of available outputs, the merger trees obtained from our simulations are very coarse-grained in time.

This highlights one of the advantages of using a code such as PINOCCHIO to produce the merger trees. In fact, in PINOCCHIO we follow the merging of haloes in real time, and then we can link each progenitor to its parent after each merging event, while in the simulations (where haloes are identified *after* the run) it is necessary to analyse and cross-correlate a large number of outputs to follow the merger histories. In other words, the generation of the merger trees is far less expensive (in term of CPU time, disc space and human labour) in PINOCCHIO than in a simulation; in fact, PINOCCHIO automatically computes the merging history of haloes and it does not need any further analysis.

3.2 Progenitor mass function

The progenitor (conditional) mass function $dN(M, z|M_0, z_0)/dM$ is the number density of progenitors of mass M at redshift z that merge to form the parent M_0 at redshift z_0. An estimate of this quantity based on the PS formalism was found by Bower (1991), while Bond et al. (1991) gave the basis for computing it in the EPS formalism (as done by Lacey & Cole 1993).

Let $\delta(z)$ be the critical density for a spherical perturbation to collapse at redshift z and $\Lambda(M) = \sigma^2(M)$ be the variance of the initial density field when smoothed over regions that contain on average a mass M. The fraction of mass of the parent halo that was in the progenitors of mass M at early times is

$$f(M, \delta|M_0, \delta_0)\,dM = \frac{1}{\sqrt{2\pi}} \frac{(\delta - \delta_0)}{[\Lambda(M) - \Lambda(M_0)]^{3/2}}$$
$$\times \exp\left\{ -\frac{(\delta - \delta_0)^2}{2[\Lambda(M) - \Lambda(M_0)]} \right\} d\Lambda, \qquad (3)$$

and the conditional mass function is

$$\frac{dN}{dM}(M, z|M_0, z_0)\,dM = \left(\frac{M_0}{M}\right)^2 f(M, \delta|M_0, \delta_0)\,dM. \qquad (4)$$

---

[3] The Hubble constant is assumed to be $H_0 = 100\,h$ km s^{-1} Mpc^{-1}.
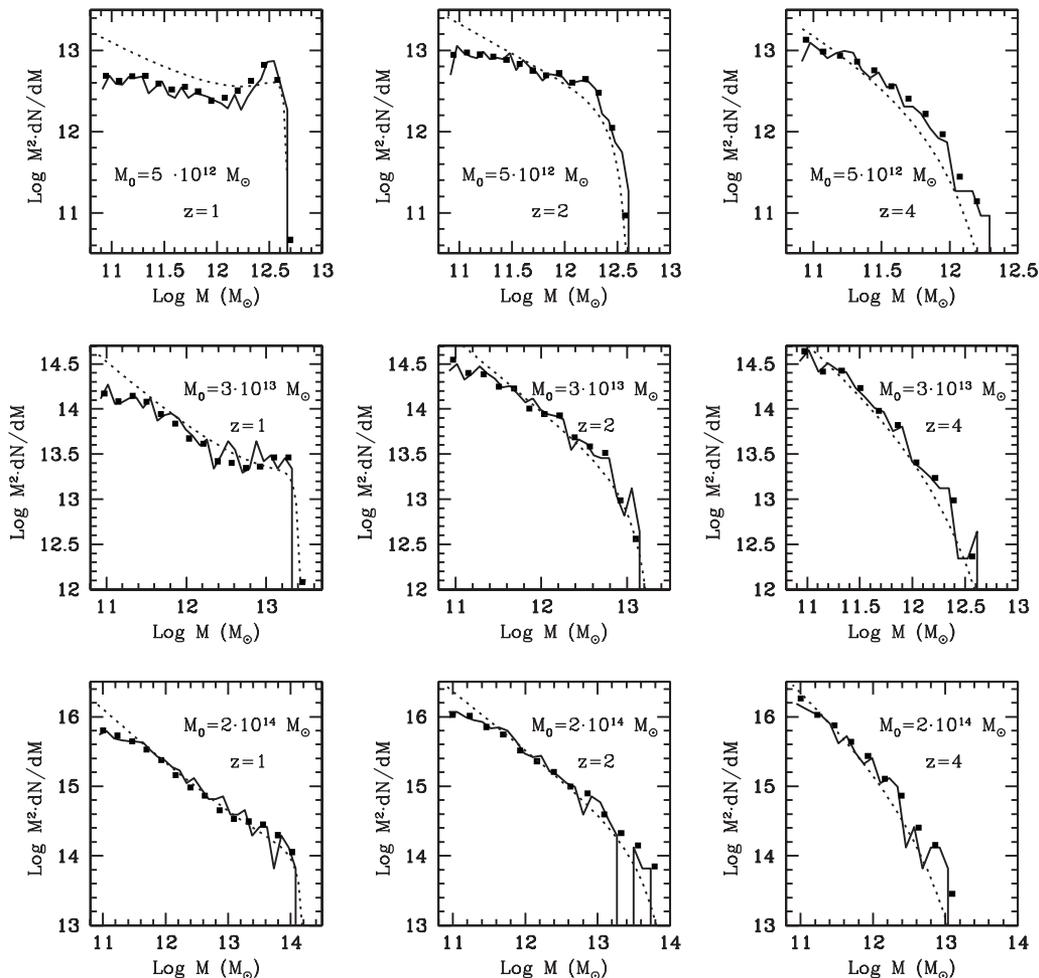
Figure 1. Conditional mass functions in the ΛCDM case for parent haloes identified at $z = 0$. The mass threshold is fixed at $M_{\rm th} = 7.6 \times 10^{10}\,{\rm M}_\odot$ (10 particles), the redshift increases from left to right and covers the values: $z = 1, 2, 4$. The mass of the parent halo increases from top to bottom, the adopted values are: $M_0 = 5.0 \times 10^{12}$, 3.0×10^{13} and $2.0 \times 10^{14}\,{\rm M}_\odot$. The points represent the simulation data while the solid lines are the prediction of PINOCCHIO; the dashed lines are the analytical predictions of the EPS formalism.

To identify the parent mass for both the PINOCCHIO conditional mass function and that obtained from the simulations we consider a mass interval around the parent mass $\log M_0$ of 0.01 dex. We are actually evaluating the conditional mass function not for a single parent but for all haloes of approximately that mass, taking care of using a mass interval small enough not to distort the distribution.

In Figs 1 and 2 we compare the conditional mass functions obtained from the EPS formalism, PINOCCHIO and the simulations for the ΛCDM and the SCDM case, respectively. The bottom panels of Fig. 1 show for the ΛCDM case the results for a cluster-sized parent of $M_0 = 2 \times 10^{14}\,{\rm M}_\odot$, the case of haloes corresponding to small groups ($M_0 = 3 \times 10^{13}\,{\rm M}_\odot$) and galaxies ($M_0 = 5 \times 10^{12}\,{\rm M}_\odot$) are presented in the mid and upper panels. In Fig. 2 we show the results for parents with masses comparable to massive clusters ($M_0 = 1 \times 10^{15}$ and $5 \times 10^{15}\,{\rm M}_\odot$) extracted from the SCDM simulation. The dotted lines show the EPS analytical prediction and the points show the expected value computed from the simulations.

The conditional mass function predicted using PINOCCHIO (the solid lines in the plots) shows a very good agreement when compared with the simulations. In Figs 1 and 2 we show that the PINOCCHIO prediction fits the simulations data with similar accuracy for all the considered parent mass and redshifts and we

identify a discrepancy between the two distributions, which in general is less than 25 per cent. This means that PINOCCHIO reproduces the conditional mass function with better accuracy that the EPS prediction and is almost constant in mass and redshift.

On the other hand, the figures show a discrepancy already pointed out by other authors for the mass function of haloes (Gelb & Bertschinger 1994; Governato et al. 1999; Jenkins et al. 2001; Bode et al. 2001): the EPS prediction overestimates the number of low-mass progenitors and underestimates the number of high-mass progenitors. This discrepancy is less evident at high redshift and it ranges from 30 per cent to a factor of 2 or more depending on the mass of the parent halo.

3.3 Higher-order analysis of the progenitor distribution

We evaluate the distribution of the mass of the largest progenitor M_1 (i.e. the most massive halo that flows into the parent) for each of the parent haloes analysed previously. The histograms in Figs 3 and 4 show the distribution of the mass of the larger progenitor normalized to the parent mass, M_1/M_0, predicted by PINOCCHIO for the ΛCDM and SCDM case (in the following the mass threshold is always set to 30 particles). The symbols connected with lines denote the corresponding simulation results. The agreement
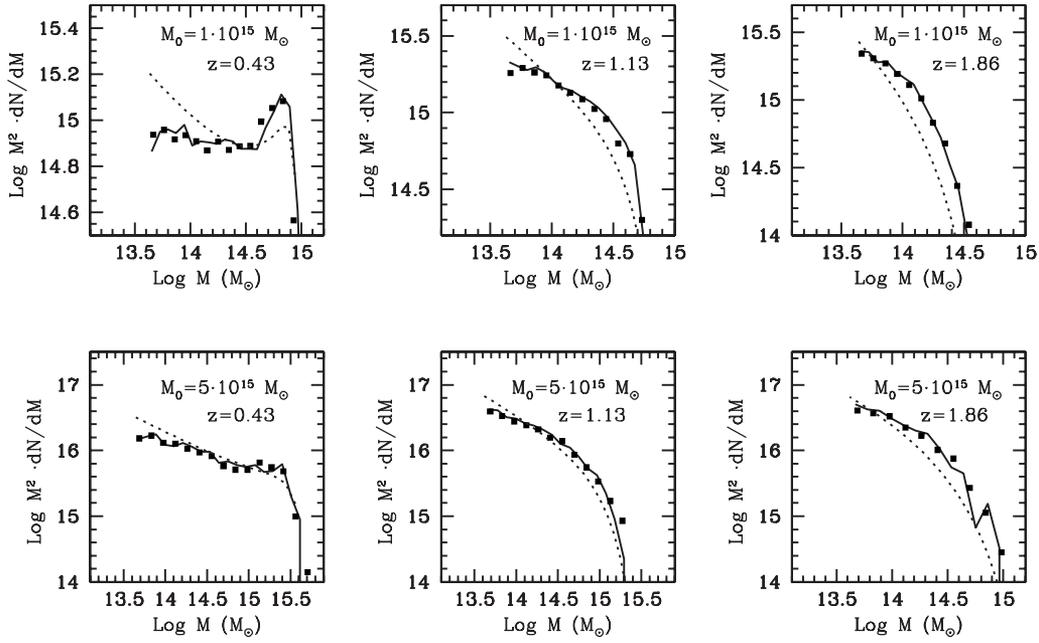
Figure 2. Same as in Fig. 1 but for the SCDM case. The mass threshold is $M_{\rm th} = 1.49 \times 10^{13}$ M$_\odot$ (10 particles).
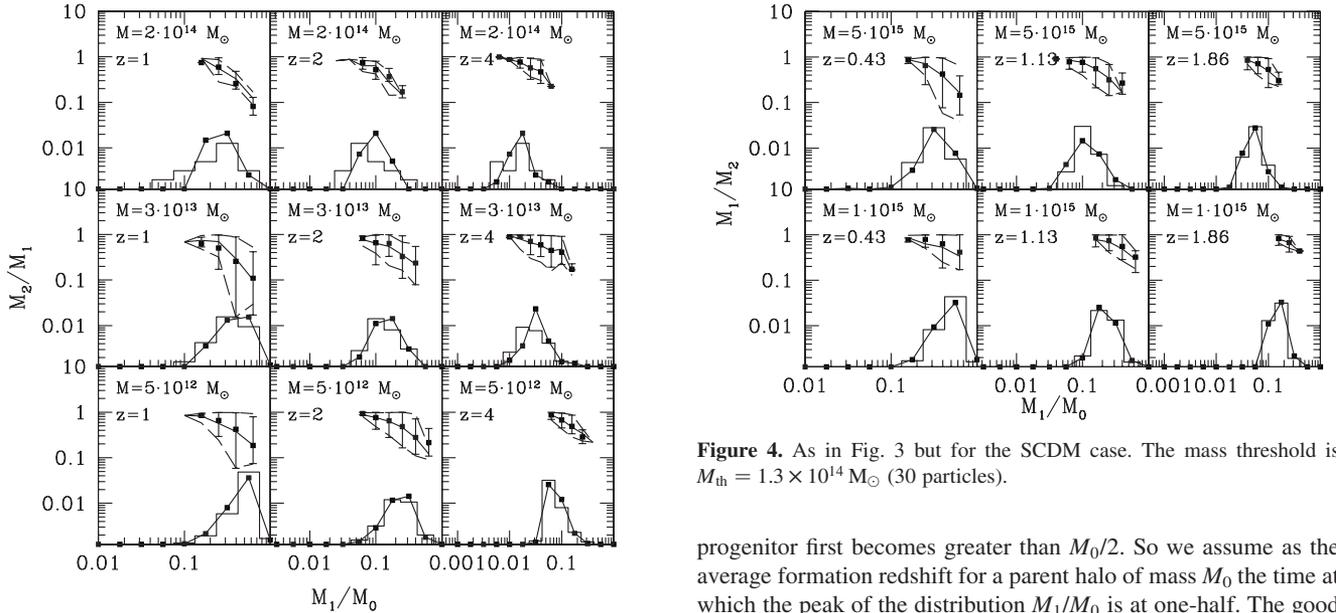
Figure 3. The distribution of the mass of the largest progenitor M_1 for the ΛCDM case with mass threshold $M_{\rm th} = 2.3 \times 10^{11}$ M$_\odot$ (30 particles). The histograms are the PINOCCHIO predictions and the points connected with solid lines are the simulations. The quantity plotted in the upper part of each box is the mean of the distribution of the mass ratio of the second largest progenitor M_2 to the first largest progenitor M_1 versus the mass ratio of the largest progenitor to the parent halo. The solid line denotes the PINOCCHIO result and the dashed lines show its 1σ variance. The points with error bars are the simulation data.

between the numerical experiment and PINOCCHIO is very good. Both the mean value and the width of the distribution are reproduced with good accuracy at all redshifts.

The distribution of M_1/M_0 also provides a hint on the formation time of the parent. In fact, one possible definition of *formation time* for a halo of mass M_0 is the epoch at which the size of its largest

Figure 4. As in Fig. 3 but for the SCDM case. The mass threshold is $M_{\rm th} = 1.3 \times 10^{14}$ M$_\odot$ (30 particles).

progenitor first becomes greater than $M_0/2$. So we assume as the average formation redshift for a parent halo of mass M_0 the time at which the peak of the distribution M_1/M_0 is at one-half. The good agreement of PINOCCHIO with the simulations can thus also be extended to the halo formation times. For instance, Fig. 4 suggests that, in this SCDM cosmology, a halo of 1×10^{15} M$_\odot$ forms at $z \sim 0.43$ or later. Note that a more detailed analysis of formation times is hampered by the small number of simulation outputs available.

In the upper part of the plots of Figs 3 and 4 the distribution of M_2/M_1 (the ratio of the second largest progenitor and largest ones) given M_1/M_0 is shown. The points are the mean value of the distribution and the error bars are the corresponding 1σ variance, both measured in the simulations. The solid lines and the dashed lines are the same quantities predicted by PINOCCHIO. Again the agreement is very good.

The results reported in this section and in the previous one suggest that the merging histories of haloes produced by PINOCCHIO reproduce with very good accuracy the statistical

properties of the masses of those extracted from numerical simulations. We note that the EPS-based algorithms for producing merger trees (Lacey & Coles 1994; Somerville & Kolatt 1999; Sheth & Lemson 1999a,b; Cole et al. 2000) are by construction forced to reproduce the EPS analytical distributions, and they suffer from the same discrepancy noted for the EPS analytical prediction (Figs 1 and 2).

3.4 The progenitors in number

In this section we analyse the statistical properties of the distribution of the number of progenitors of a halo of mass M_0.

In Figs 5 and 6 we show the probability $P(N, M_0)$ that a halo of mass M_0 has N progenitors. The average of these distribution gives (with suitable normalization) the integral of the conditional mass
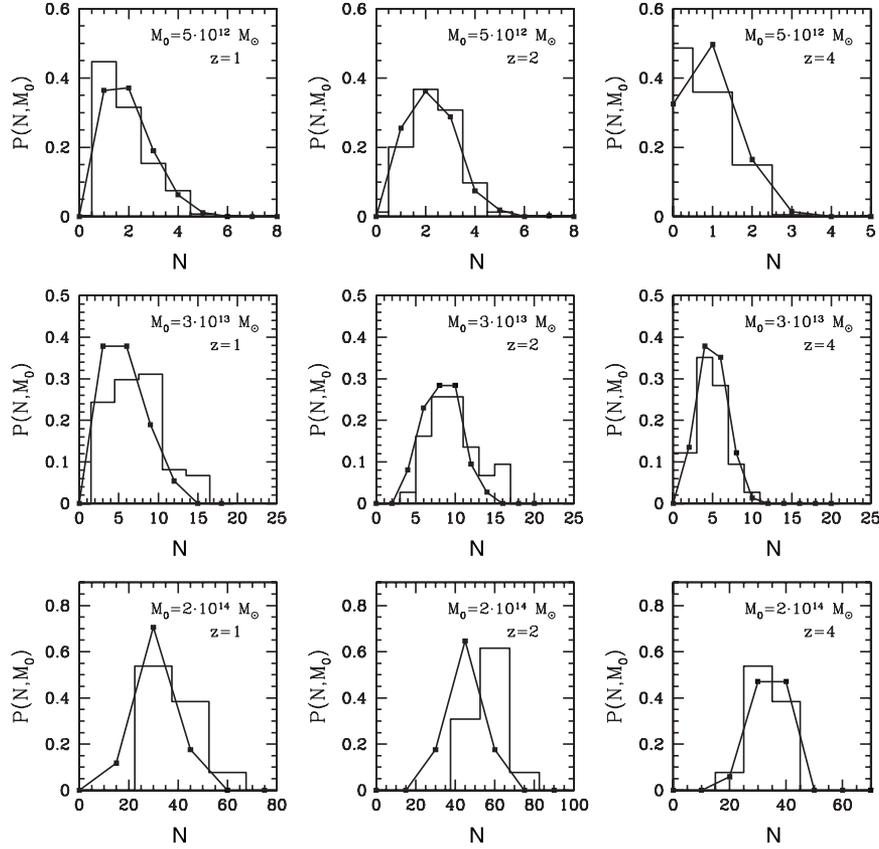
Figure 5. Probability that a halo M_0 at $z = 0$ has N progenitors for the ΛCDM case. The threshold mass is $M_{th} = 2.3 \times 10^{11}$ M$_\odot$ (30 particles). The points connected with solid lines represent the simulation data while the histograms are the prediction of PINOCCHIO.
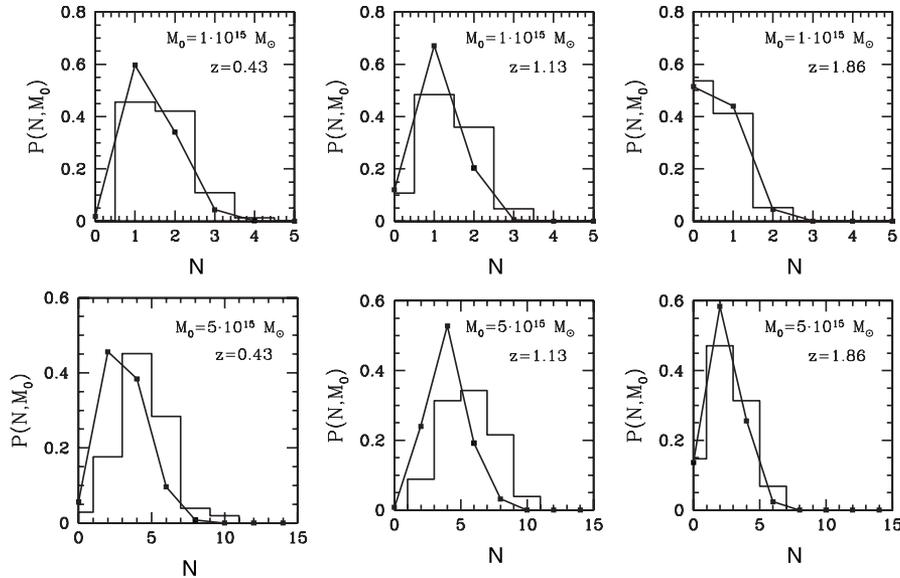
Figure 6. Same as in Fig. 5 for the SCDM case. The threshold mass is $M_{th} = 1.3 \times 10^{14}$ M$_\odot$ (30 particles).
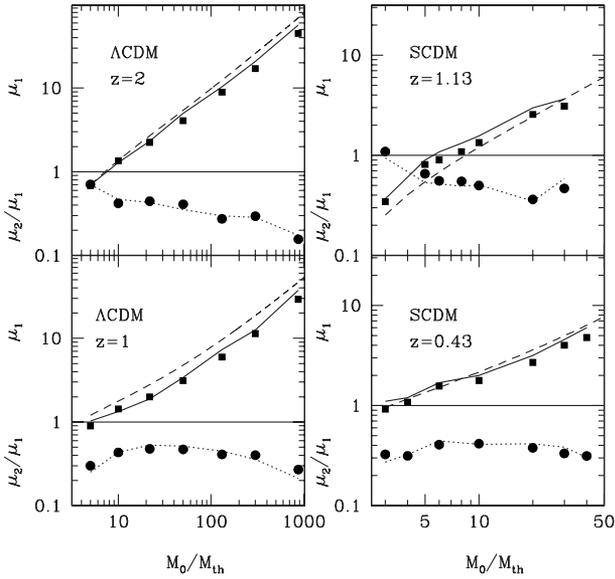
Figure 7. The first two moments of the distribution of the number of progenitors $P(N, M_0)$ as a function of the parent mass M_0. Lines represent the PINOCCHIO results and symbols denote the simulation data. The left-hand plots show the ΛCDM case at redshifts $z = 1$ and 2. The threshold mass is $M_{th} = 2.3 \times 10^{11} M_\odot$ (30 particles) and we plot the mean (squares and solid line) and the rescaled variance (circles and dotted line) up to $M_0 = 1000 M_{th}$. The dashed line is the EPS analytical prediction for the mean. The right-hand plots show the SCDM case at redshifts of $z = 0.43$ and 1.13. The threshold mass is $M_{th} = 1.3 \times 10^{14} M_\odot$ (30 particles), and we plot the mean and the rescaled variance up to $M_0 = 50 M_{th}$.

Figure 8. Fraction f_p of cleanly assigned progenitors for redshift $z = 1, 2$ and 4. The left-hand side represents the fraction of cleanly assigned progenitors as function of the parent mass. The three lines correspond to the different redshifts. The tree plots on the right are the scatter plots of f_p as function of the progenitor mass M normalized to the parent mass M_0, for the parent haloes of mass $10^{11} M_\odot < M_0 < 10^{15} M_\odot$; the redshift increases from top to bottom.
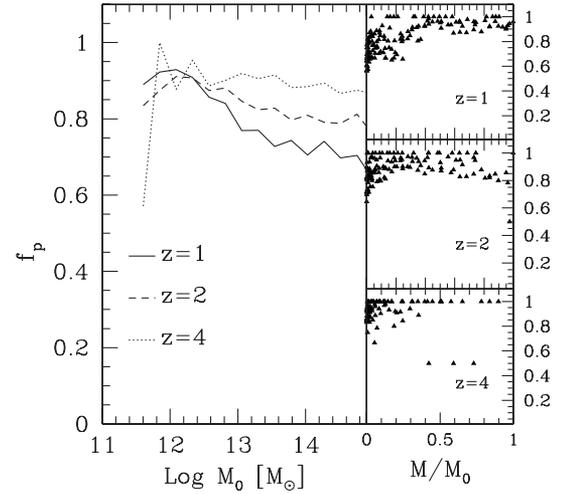
function to the threshold mass, and is dominated by the more numerous small-mass objects.

The histograms show the distribution of the number of progenitors evaluated from PINOCCHIO for different parent masses and redshifts. The filled symbols connected with lines are the distribution extracted from the simulation. We note that PINOCCHIO also reproduces the distributions fairly well for the more massive haloes and at all redshifts.

The ability of PINOCCHIO to predict the distribution of the number of progenitors can be quantified by comparing the first and second moments measured in the simulations with their values predicted by PINOCCHIO. In Fig. 7 we show the average μ_1 and the rescaled variance μ_2/μ_1 as a function of the parent halo mass for different redshifts. The lines are the prediction of PINOCCHIO and the symbols are the same quantities measured from the simulations. The dashed lines are the EPS analytical prediction for μ_1 computed by integrating equation (4). Note that for arbitrary initial conditions the EPS formalism cannot evaluate the higher moments of the distribution analytically.

The agreement between PINOCCHIO and the simulations varies from 5 per cent of the ΛCDM to 10 per cent of the SCDM case, but it does not depend on the redshift. Again PINOCCHIO is found to improve with respect to EPS. In particular, at low redshifts the EPS predictions underestimate the mean value by a factor that ranges from 20 to 30 per cent.

Our results can be compared with those given by Sheth & Lemson (1999b) and Somerville et al. (2000) for EPS-based merger trees and with Sheth & Tormen (2002) who elaborate an excursion set model based on ellipsoidal collapse. In general, PINOCCHIO reproduces the statistical properties of progenitor distributions with better accuracy than the other methods. It is

remarkable that the tests based on parent haloes with different mass ranges give very similar results, reproducing the simulations with a comparable accuracy.

3.5 Object-by-object comparison

We finally test the degree of agreement between PINOCCHIO and the simulations at the object-by-object level for the number of progenitors that are reconstructed cleanly. In Papers I and II a pair of haloes from the two catalogues (PINOCCHIO and FOF) were defined as cleanly assigned to each other if they overlapped in the Lagrangian space for at least 30 per cent of their volume and no other object overlapped with either of them to a higher degree. The cleanly assigned haloes were shown to overlap on average at the 60–70 per cent level over all redshifts. The fraction of cleanly assigned haloes was found to depend on the degree of non-linearity reached by the system, decreasing from almost 100 per cent to 70 per cent, at worst, at later times. This is caused by the lower accuracy of the Zel'dovich approximation in predicting the displacements as the density field becomes more and more non-linear (see Paper II).

We now quantify the number of PINOCCHIO progenitors that are cleanly assigned to FOF progenitors for each cleanly assigned parent halo. For this analysis we restrict ourselves to the ΛCDM case, which gives a wider mass range but a higher level of non-linearity.

In Fig. 8 we show, for the parents that are cleanly identified, the fraction in number f_p of the progenitors that are cleanly identified as well. This quantity is shown both as a function of the parent mass M_0 and as a function of the progenitor mass M/M_0 in units of the parent mass. The number of cleanly identified progenitors ranges from 60 to 100 per cent, with an average value of between 80 and 90 per cent. The fraction f_p is in general higher at higher redshift, when the object-by-object agreement between PINOCCHIO and the simulation is better. As a function of M_0 larger parent haloes tend to be reconstructed with worse accuracy, especially at $z = 1$. This is mainly caused by the small progenitors, as the

right-hand panels of Fig. 8 show. The progenitors that carry a mass of less that ~ 20 per cent are those that are worst reconstructed. We conclude that PINOCCHIO is able to reconstruct correctly the main branches of the merger trees, while secondary branches, which are especially present in the larger haloes, are reconstructed in a noisier way.

4   THE SPATIAL PROPERTIES OF MERGING HALOES

One notable limitation of EPS is the lack of spatial information for the haloes. Several authors (Mo & White 1996; Mo, Jing & White 1997; Catelan et al. 1998; Porciani et al. 1998) found approximate analytical expressions for the bias of haloes of fixed mass, i.e. for the ratio between the two-point correlation function of haloes and that of the underlying matter field. Such analytical estimates have been found to agree with the results of simulations to within ~ 40 per cent (Mo & White 1996; Jing 1998; Porciani et al. 1999; Sheth et al. 2000; Colberg et al. 2000). In this approach it is not possible to know how the bias changes for haloes with different merger histories. This piece of information is important for producing predictions on the bias of galaxies of different types, which typically have different merger histories.

As shown in Papers I and II, PINOCCHIO haloes have the same correlation length r_0 as FOF haloes to within a 10 per cent error. Having knowledge of both merger histories and halo positions, PINOCCHIO can provide information on the relation between clustering and merging. To show this, we select PINOCCHIO and FOF haloes in the ΛCDM cosmology at $z = 0$ with masses greater than $10^{14}\,\mathrm{M_\odot}$. We check their merging histories at $z = 1$, 2 and 4, and we evaluate the two-point correlation functions for their progenitors. In Fig. 9 the solid lines represent the two-point

correlation function of progenitors, $\xi_p(r)$, evaluated in PINOCCHIO compared with the same quantity measured in the simulation. The plots show that PINOCCHIO reproduces such correlation functions to within ~ 20 per cent error.

We also compare this function with the average correlation function, $\xi_h(r)$, at the same redshifts. It is apparent that PINOCCHIO reproduces correctly the larger clustering amplitude of haloes that flow into a cluster-sized one. The bias between the two halo populations is defined as: $b^2(r, z) = \xi_p(r)/\xi_h(r)$. In the bottom row of plots in Fig. 9 We compare the bias measured in the simulation with the PINOCCHIO results. The bias is recovered to within ~ 20 per cent and the scale dependence is correctly reproduced.

5   ORBITAL PARAMETERS OF THE MERGING HALOES

In the hierarchical clustering scenario a merging event between two or more haloes corresponds to the loss of identity of the single primitive units that merge to form a new halo. However, high-resolution *N*-body simulations show that the dynamical evolution after an encounter is more complicated than this idealized picture: the haloes may retain their identity, and become substructure for the new system (Moore, Katz & Lake 1996; Tormen 1997; Ghigna et al. 1998; Tormen, Diaferio & Syer 1998). Indeed, this is in line with the same evidence of galaxies within galaxy groups or clusters. The life of these substructures is affected by the various dynamical effects that contribute to their disruption. The dynamical friction force drives the satellites towards the centre of mass of the system where they can merge with the central object or among themselves. While a satellite orbits inside the main halo, the tidal forces exerted by the background induce its evaporation and reduce its mass (Gnedin & Ostriker 1997; Gnedin, Hernquist &
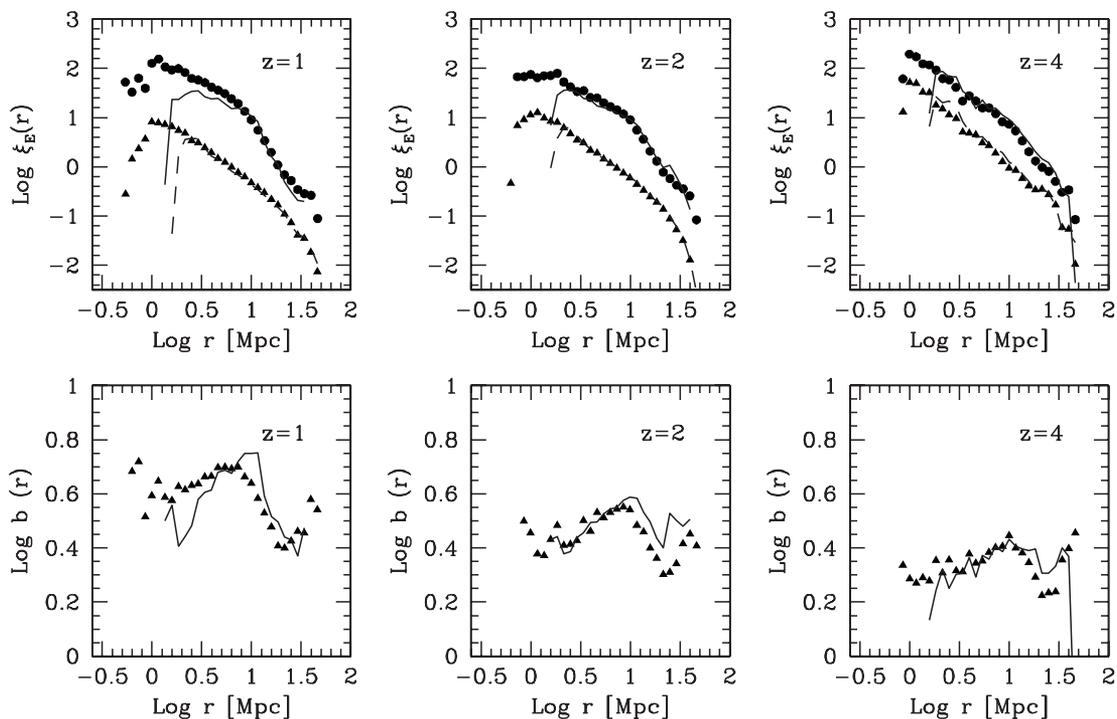
Figure 9. Top panels: correlation function for the progenitors of haloes with a mass greater than $10^{14}\,\mathrm{M_\odot}$ at $z = 0$, ξ_p (circles and solid lines), and for all haloes larger than the threshold mass $M_{th} = 2.3 \times 10^{11}\,\mathrm{M_\odot}$, ξ_h (triangles and dashed lines), at various redshifts indicated in the panels. Symbols refer to FOF-selected haloes obtained from the ΛCDM simulation, lines to the corresponding PINOCCHIO prediction. Bottom panels: bias ξ_p/ξ_h for simulations (symbols) and PINOCCHIO (lines). PINOCCHIO is able to predict accurately the clustering of the selected halo types.

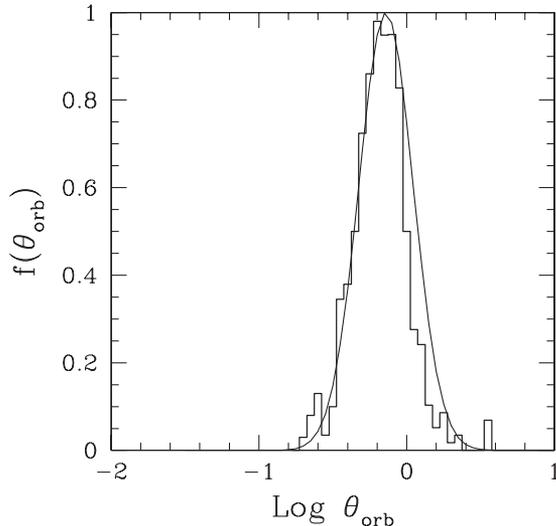Ostriker 1999; Taylor & Babul 2000; Taffoni et al. 2001, in preparation).

The evolution of substructure is one of the crucial points in modelling galaxy formation. An important aspect of this is the prediction of the initial orbital parameters, i.e. the energy and the angular momentum of the orbit of the satellites infalling into the main halo. The large-scale simulations such as those we use in the present paper lack enough resolution to address such processes. High-resolution simulations are necessary to describe the evolution of satellites (Tormen 1997; Ghigna et al. 1998), at the cost of simulating one cluster at a time. It is then useful to resort again to analytic modelling of the dynamical friction and tidal stripping (Chandrasekhar 1943; see, e.g., Binney & Tremain 1987; Lacey & Cole 1993; van den Bosch et al. 1999; Colpi, Mayer & Governato 1999). These analytical models require knowledge of the initial orbital parameters. In the semi-analytical, EPS-based codes for galaxy formation, these parameters are in general Monte Carlo extracted from some distributions obtained from high-resolution simulations (Tormen 1997; Ghigna et al. 1998).

Within the PINOCCHIO code, it is possible to predict the impact parameters of the merging satellites, as the infall velocities and the relative distances are known. Note that this calculation is analogous to that of the angular momentum of haloes presented in Paper II. Given the impact (Zel'dovich) velocity $\Delta \boldsymbol{v}$ and the relative distance Δr the angular momentum and the energy are computed as:

$$\boldsymbol{J} = \Delta \boldsymbol{r} \times \Delta \boldsymbol{v} \tag{5}$$

$$E = \frac{1}{2}(\Delta \boldsymbol{v})^2 + \phi(|\Delta \boldsymbol{r}|). \tag{6}$$

$\phi(|\Delta \boldsymbol{r}|)$ can be evaluated as the gravitational potential of a point mass that touches the external layer of a spherical halo of mass $M : \phi(|\boldsymbol{r}|) = GM/|\boldsymbol{r}|$. The linear growth of the relative velocity is stopped at a physical time equal to half of the merging time (see Paper II).

To study the ability of PINOCCHIO in predicting the orbital parameters of DM substructures we compute the distribution of the orbital parameters of the satellites that merge with a halo of mass $M = 2 \times 10^{14}\,\mathrm{M}_\odot$. We express the angular momentum and the energy per unit mass in terms of the circularity $\epsilon \equiv J/J_c(E)$, where $J_c(E) = V_c r_c(E)$ is the angular momentum and $r_c(E)$ is the radius of the circular orbit with the same energy (see, e.g., van den Bosch et al. 1999). To do that we assume the density profile of Navarro, Frenk & White (1996) for the main halo and we calculate the associated potential energy profile $\phi_{\mathrm{NFW}}(R)$ and the associated circular velocity profile $V_c(R)$ (see, e.g., Navarro et al. 1996; Klypin et al. 1999; Taffoni et al. 2002). We consider a particular combination of the orbital parameters

$$\Theta_{\mathrm{orb}} = \epsilon^{0.78} \left[\frac{r_c(E)}{R_{\mathrm{vir}}} \right]^2, \tag{7}$$

introduced by Cole et al. (2001). They use the results of Tormen (1997) to derive the distribution for the Θ_{orb} factor and they find that this distribution can be fitted with a log normal function with mean value $\langle \log_{10}(\Theta_{\mathrm{orb}}) \rangle = -0.14$ and dispersion $\langle [\log_{10}(\Theta_{\mathrm{orb}}) - \langle \log_{10}(\Theta_{\mathrm{orb}}) \rangle]^2 \rangle^{1/2} = 0.26$.

The results of our analysis are presented in Fig. 10, we compare the distribution of the Θ_{orb} factor measured in PINOCCHIO (histogram) with the theoretical fit derived by Cole et al. (2001) (solid line). We note that the distribution measured from PINOCCHIO reproduces with good accuracy the log normal function. The average value derived by our analysis is $\langle \log_{10}(\Theta_{\mathrm{orb}}) \rangle = -0.18$ and the dispersion is $\langle [\log_{10}(\Theta_{\mathrm{orb}}) - \langle \log_{10}(\Theta_{\mathrm{orb}}) \rangle]^2 \rangle^{1/2} = 0.23$.

6 CONCLUSIONS

We have tested the predictions of the PINOCCHIO code, presented in Papers I and II, regarding the hierarchical nature of halo formation, with particular attention being paid to those aspects that are mostly relevant for galaxy formation. We have compared the results of PINOCCHIO with those of two large N-body simulations (ΛCDM and SCDM cosmologies) drawing the following conclusions.

(i) The merger histories of the PINOCCHIO haloes resemble closely those found by applying the FOF algorithm to the N-body simulations. The agreement is valid at the statistical level for groups of at least 30 particles (good results are obtained even for haloes of ten particles).

(ii) Statistical quantities such as the conditional mass function, the distribution of the largest progenitor, the ratio of the second largest to largest progenitors, and the higher moments of the progenitor distributions are recovered with a typical accuracy of ~ 20 per cent.

(iii) The agreement is also good at the object-by-object level, as

Figure 10. The distribution of the Θ_{orb} for the satellites that merge with a halo of mass $M = 2 \times 10^{14}\,\mathrm{M}_\odot$ at $z = 0$.

Table 1. Parameters of the considered numerical simulations.

Cosmology	Ω_0	Ω_Λ	h_0	σ_8	n	L_{box}	N_p	M_{part}	Output redshifts
ΛCDM	0.3	0.7	0.65	0.9	1.0	100 Mpc h^{-1}	256^3	$7.64 \times 10^{10}\,\mathrm{M}_\odot$	0, 0.25, 0.5, 0.75, 1, 2, 3, 4, 5
SCDM	1.0	0.0	0.5	1.0	1.0	500 Mpc h^{-1}	360^3	$1.49 \times 10^{12}\,\mathrm{M}_\odot$	0, 0.43, 1.13, 1.86

PINOCCHIO cleanly reproduces $\gtrsim 70$ per cent of the progenitors when parent haloes are cleanly recognized themselves. The agreement slowly degrades with time.

(iv) The increased noise recovered in the object-by-object agreement, as time progresses and non-linearity grows, does not influence the accuracy of the predictions in a statistical sense.

(v) The correlation function of higher-redshift haloes that are progenitors of lower-redshift massive haloes is correctly reproduced to within an accuracy of ~ 10 per cent in r_0. The scale-dependent bias of these with respect to the total halo population is also reproduced to within an accuracy of 20 per cent or better.

(vi) PINOCCHIO gives an estimate of the initial orbital parameters of merging haloes as well, which is found to be in reasonable agreement with the available results from high-resolution *N*-body simulations.

(vii) The fit of the statistical quantities achieved by PINOCCHIO is much better than the PS and EPS estimates, which show discrepancies of up to a factor of 2 (see Governato et al. 2000; Jenkins et al. 2001; Bode et al. 2001; Somerville et al. 2000; Sheth & Lemson 1999b; Cohn, Bagla & White 2001).

(viii) The validity of PINOCCHIO extends to the object-by-object level, in contrast to EPS (Bond et al. 1991).

(ix) PINOCCHIO is not affected by the inconsistencies of the EPS approach, which can only be corrected by means of heuristic recipes (Somerville & Kolatt 1999; Sheth & Lemson 1999; Cole et al. 2001).

(x) PINOCCHIO provides much more useful information on the haloes, such as positions, velocities and angular momenta and initial orbital parameters at merger; at the same time it is not more computationally demanding than an EPS-based code in generating the merging histories of haloes.

These results confirm the validity of PINOCCHIO as a fast and flexible tool for studying galaxy formation or for generating catalogues of galaxies or galaxy clusters, suitable for large-scale structure studies. In fact, PINOCCHIO reproduces, in a much quicker way and to a very good level of accuracy, all the information that can be obtained from a large-scale *N*-body simulation with pure dark matter, without needing all the post-processing necessary to obtain the merger trees of haloes. PINOCCHIO is available at http://www.daut.univ.trieste.it/PINOCCHIO.

REFERENCES

Binney J., Tremaine S., 1987, Galactic Dynamics. Princeton Univ. Press, Princeton
Bode P., Bahcall N. A., Ford E. B., Ostriker J. P., 2001, ApJ, 551, 15B
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Bouchet F., 1996, in Bonometto S.et al., ed., Dark Matter in the Universe. IOS, Amsterdam
Buchert T., 1996, in Bonometto S.et al., ed., Dark Matter in the Universe. IOS, Amsterdam
Bower R., 1991, MNRAS, 248, 332
Catelan P., 1995, MNRAS, 276, 115
Catelan P., Lucchin F., Matarrese S., Porciani C., 1998, MNRAS, 297, 692
Chandrasekhar S., 1943, ApJ, 97, 255
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Cohn J. D., Bagla J. S., White M. J., 2001, MNRAS, 325, 1053
Colberg J. M. et al., 2000, MNRAS, 319, 209
Colpi M., Mayer L., Governato F., 1999, ApJ, 525, 720
Couchman H., Thomas P., Pearce F., 1995, ApJ, 452, 797
Efstathiou G. P., Frenk C. S., White S. D. M., Davis M., 1988, MNRAS, 235, 715
Gelb J. M., Bertschinger E., 1994, ApJ, 436, 467
Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 1998, MNRAS, 300, 146
Gnedin O. Y., Ostriker J. P., 1997, ApJ, 474, 223
Gnedin O. Y., Hernquist L., Ostriker J. P., 1999, ApJ, 514, 109
Gnedin O. Y., Lee H. M., Ostriker J. P., 1999, ApJ, 522, 935
Governato F., Babul A., Quinn T., Tozzi P., Baugh C. M., Katz N., Lake G., 1999, MNRAS, 307, 949
Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372
Jing Y. P., 1998, ApJ, 503, 9
Jing Y. P., 1999, ApJ, 515, 45
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Klypin A., Gottlober S., Kravtsov A. V., Khokholov A. M., 1999, ApJ, 516, 530
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lacey C., Cole S., 1994, MNRAS, 271, 676
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Mo H. J., Jing Y. P., White S. D. M., 1997, MNRAS, 284, 189
Monaco P., 1995, ApJ, 447, 23
Monaco P., 1997, MNRAS, 287, 753
Monaco P., 1998, Fund. Cosm. Phys., 19, 153
Monaco P., Theuns T., Taffoni G., Governato F., Quinn T., Stadel J., 2002, ApJ, 564, 8 (Paper I)
Monaco P., Theuns T., Taffoni G., 2002, MNRAS, 331, 587 (Paper II)
Moore B., Katz N., Lake G., 1996, ApJ, 457, 455
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Padmanabham T., 1993, Structure Formation in the Universe. Cambridge Univ. Press, Cambridge
Peacock J. A., Heavens A. F., 1990, MNRAS, 243, 133
Porciani C., Catelan P., Lacey C., 1999, ApJ, 513, 99
Porciani C., Matarrese S., Lucchin F., Catelan P., 1998, MNRAS, 298, 109
Press W. H., Schechter P., 1974, ApJ, 187, 425 (PS)
Sheth R. K., Lemson G., 1999a, MNRAS, 304, 767
Sheth R. K., Lemson G., 1999b, MNRAS, 305, 946
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Sheth R. K., Tormen G., 2002, MNRAS, 329, 61
Somerville R. S., Kolatt T. S., 1999, MNRAS, 305, 1
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Somerville R. S., Lemson G., Kolatt T. S., Dekel A., 2000, MNRAS, 316, 479
Taffoni G., Mayer L., Colpi M., Governato F., 2002, in Fusco-Fermiano R., Matteucci F., eds, ASP Conf. Ser. Vol. 253, Chemical Enrichment of Intracluster and Intergalactic Medium. Astron. Soc. Pac., San Francisco, in press (astro-ph/0109029)
Taylor J., Babul A., 2001, ApJ, 559, 716
Tormen G., 1997, MNRAS, 290, 411
Tormen G., Diaferio A., Seyer , 1998, MNRAS, 299, 728
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
van den Bosch F., Lewis G., Lake G., Stadel J., 1999, ApJ, 515, 50
Zel'dovich Y. A. B., 1970, Astrofizika, 6, 319 (translated in 1973, AP, 6, 164)

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