SISSA ISAS
SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI
INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

Formation and Evolution of Dark Matter Haloes in Hierarchical Models for Structure Formation

Thesis submitted for the degree of
"Doctor Philosophiæ"

CANDIDATE

Giuliano Taffoni

SUPERVISORS

Prof. Luigi Danese
Dr. Pierluigi Monaco

October 2002

To Elio....

Anything that happens, happens.

Anything that, in happening, causes something else to happen,
causes something else to happen.

Anything that, in happening, causes itself to happen again, happens again.

It doesn't necessarily do it in Chronological order, though.

Douglas Adams

Contents

List of Figures

Acknowledgments

In the first place, all my thanks go to Pierluigi Monaco, who always supported me with trust and friendship, and to Luigi Danese who stimulated my research activity and intellectual growth.

During the last two years I have been very fortunate to work with Monica Colpi and I want to thank her for the useful, interesting and *long* scientific discussions. She shared with me her deep physical insight and guided my research with a unique enthusiasm.

All the people I worked with deserve my gratitude. I cannot forget Fabio Governato whose practical mind focused my dispersing way of approaching scientific problems. A special place is due to Lucio Mayer, we have been working on the evolution of satellites since the beginning of my PhD. I believe he is a versatile researcher and a good friend.

Then, I am grateful to the staff of the Astrophysics sector in SISSA and in particular to Riccardo Valdarnini, for helping me every time I had a scientific problem, and to Luciano Rezzolla for his useful advice.

Part of my work in SISSA concerned computers. So how can I forget to mention the system managers? David and Davide, Luisa, Marina, Pietro and last but not least Fabio, who was always ready to install (with ding-dong results) new high-tech hardware devices on my computer.

Since I arrived in Trieste, a lot of friends shared with me good time, fun and "bora". However, I didn't forget my historical friends in Rome. I will try to list all of them.

Alessandro: since I met him (a long long time ago) he has been and he still is a constant source of dialectic exercise.

My Amnesty International group: Antonio, Alessandra, Cristina, Claudio, Eleonora, Emiliano, Luca, Massimiliano.

I did long and exciting discussions about the "meaning of life and its possible applications to the strategies of scopone scientifico" with Bobo, Cartacci, il "Direttore", Giampaolo and "Mac".

Sandro communicated to me his passion for Astronomy, probably it is also his fault if I got this PhD.

Joachim helped me to survive the first months in Trieste. He introduced me to the night life of downtown, I introduced him to the secrets of the Italian cuisine.

Pasquale shared with me office, desk and the passion for Linux: I will never forget his recipe for the red–hot–chili–pepper oil.

Olindo shared with me office, house and his wife's spaghetti sauce.

Stefano, Paola and their cats. "Lo discepolo si inchina allo ricordo de lo maestro lontano".

Manuela animated our lives with her unique way of doing.

Then Ezio "the dungeon master"; Enrichetta, Ilaria and Vanessa "the room 101 girls"; Annamaria, Carlo "the wizard", Donatella, Fabio, Francesca, Ilia, Luca, Michele, Paolo, Pedro, Ruben, Stefano, Tom; and those that I forgot.

A special thank goes to Delphine.

Obviously, I would like to thank my family. In particular, my parents for the moral and economical support and my grandfather, his recent departure has been a great loss for all of us.

Last but (of course !) not least...I thank Valentina for being as she is...wonderful.

Chapter 1

Introduction

In Dark Matter (DM) dominated cosmological models, structure grows through the gravitational amplification and collapse of small primordial perturbations, imprinted at very early times by some mechanism such as inflation. In particular, in the case of Cold Dark Matter (CDM), the formation of structures follows a hierarchical pattern, with more massive halos forming from accretion of mass and mergers of smaller objects (see e.g. Padmanabhan 1993 for a general introduction). The properties of the halo population are of fundamental importance to understand galaxy formation and evolution. Indeed, galaxies are thought to form when baryons fall into such dark matter haloes and are shocked to sufficiently high temperatures and densities that the gas can cool radiatively to form stars (Rees & Ostriker 1977; White & Rees 1978; White 1996).

The formation of DM haloes involves highly non-linear dynamical processes which cannot be followed analytically. To tackle this problem it is necessary to relay on numerical N-body simulations. A numerical simulation evolves a set of equal mass particles representing the dark matter, in a simulation box (see e.g. Efstathiou et al. 1985). A popular way of identifying 'haloes' in such calculations is via the friends-of-friends (FOF) algorithm, which links particles within a fraction b of the mean inter-particle spacing to a halo, at a density contrast $> 1/b^3$ (other halo identification algorithms generally give similar results). Despite their successful and accurate description of the hierarchical evolution of structures, the use of N-body codes have some drawbacks, mainly related to their expensive use of computer power. For example, it is difficult to use simulations to explore a wide range of cosmological models or halo masses.

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), 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 Press & Schechter approach was then extended by Bond et al. (1991) to give the so called *Extended Press & Schechter* formalism (hereafter EPS; see also Peacock & Heavens 1990; Bower 1991; Lacey & Cole 1993). The EPS model can be also used to predict some properties of DM haloes, such as their formation time, survival time and merger rate. These predictions were successfully tested against numerical simulations, by Lacey & Cole (1994).

Semi-analytical methods based on the EPS formalism

The EPS formalism has recently become a standard tool to construct synthetic catalogues of DM haloes for galaxy formation programs (see, e.g., Kauffmann, White & Guiderdoni 1993; Somerville & Primack 1999; Cole et al. 2000).

In the framework of the EPS formalism I have developed a semi-analytical code to Monte-Carlo-generate merging histories of haloes. This Merger Tree code presents many advantages when compared to the numerical approach: it is simple to implement, computationally efficient and can be used for different cosmological models. Unfortunately any EPS–based semi–analytical approach presents some systematic differences when compared with simulations, already noticed by several authors (Lacey & Cole 1993; Somerville & Kolatt 1999; Sheth & Lemson 1999). The clustering of halos of given mass in the Press & Schechter approach can be obtained analytically (Mo & White 1996; Catelan et al. 1998; Porciani, Catelan & Lacey 1999; Sheth & Tormen 1999; Sheth et al. 2001; Colberg et al. 2001), but the EPS approach is not able to produce both spatial information and merger histories at the same time. This is an important limit that also affects any EPS–based semi–analytical code.

The use of Lagrangian perturbation theory

A number of analytic or semi-analytic techniques based on Lagrangian perturbation theory (Bouchet 1997; Buchert 1997) were devised to approximate some aspects of the gravitational evolution. Those techniques have the advantage of giving insight of the dynamics of the gravitational collapse. In particular, the Lagrangian Perturbation Theory (LPT; Moutarde et al. 1992; Buchert & Ehlers 1993; Catelan 1995) and more specifically its linear term, the Zel'dovich (1970) approximation, have been used to compute many properties of the density and velocity fields in the 'mildly non-linear regime' , when the density contrast is not very high and particle trajectories still retain some memory of the initial conditions.

PINOCCHIO code

Recently, we have proposed a new approach to simultaneously obtain the spatial information and the merger history of several DM haloes (Monaco et al. 2002). In the PINOCCHIO (PINpointing Orbit-Crossing Collapsed HIerarchical Objects) formalism, following Monaco (1995; 1997a,b) and Monaco & Murante (1999), we use LPT in the context of the EPS approach to provide predictions for the collapse of fluid elements in a given numerical realization of a linear density field. The PINOCCHIO code consists of a two-step procedure. First, mass elements are assumed to collapsed after undergoing *orbit crossing*. The orbit crossing corresponds to the time at which the Lagrangian approximations break down and the density diverges *locally*. It is then reasonable to assume this time as the "collapse time" of a mass element. Secondly, collapsed points are grouped into halos using an algorithm that mimics the hierarchical growth of structures through accretion and mergers. The Zel'dovich approximation is then used to compute the Eulerian positions of halos at a given time. Some points that have undergone orbit crossing are assigned to the network of filaments and sheets that connect the halos.

The PINOCCHIO results have been compared with simulations (Monaco, Theuns & Taffoni 2002; Taffoni, Monaco & Theuns 2002): our code can accurately reproduce many properties of DM haloes from large N-body simulations that started from the same initial density field. The good agreement is not only for statistical quantities such as the mass or the correlation function, but extends to the object-by-object comparison. The mass function of haloes is recovered within a $\sim$5 per cent accuracy and the correlation function in a $\sim$20 per cent ($\sim$10 per cent error in r_0). At the object-by-object level, the $\sim$70 per cent of the simulated halos are identified by PINOCCHIO with an error of 10 per cent on position and 10–20 per cent on velocity. Moreover the distribution of the angular momentum of halos is also recovered. These results show that PINOCCHIO is a proper (but approximated) way to describe the gravitational problem, and not simply a phenomenological model able to reproduce some particular aspects of gravitational collapse.

Furthermore, I demonstrate that PINOCCHIO can be used to reconstruct merging history of haloes which agrees both from a statistical point of view and at the object-by-object level with that measured in N-body simulations (Taffoni, Monaco & Theuns 2002). The conditional mass function and the distribution of progenitors number are recovered with error $\sim$20 per cent error. The merging history of haloes is also reproduced with an error lower than 15 per cent. PINOCCHIO thus provides a significant improvement over the EPS approach, which is known to be approximately valid only in a statistical sense (Bond et al. 1991; White 1996).

Substructure of dark matter haloes

Halo clustering leads to merging of their embedded galaxies. At each merging event various processes occur, such as morphological transformation of the stellar and gaseous components. Understanding the dynamical evolution of DM haloes is a fundamental step in any theory of galaxy formation.

N-body simulations are widely used to study the dynamical evolution of cosmic structures. A detailed study of the merging between haloes and or harassment and tidal stirring among substructures requires extremely high resolution simulations (e.g. Huang & Carlberg 1997; Naab, Burkert & Hernquist 1999; Moore et al. 1999; Lewis et al. 2000; Jing & Suto 2000; Fukushige & Makino 2001).

A different approach can be found on the semi-analytical codes for halo formation, such as the PINOCCHIO code or EPS Monte Carlo merger trees. The evolution of substructures in semi-analytical models is followed in a rather schematic way: merging events between unequal mass haloes take place when the lighter halo reaches the center of the more massive one. The time scale for this to happen is obtained from the local application of the Chandrasekhar's formula (1943) for dynamical friction.

However, as the magnitude of the frictional drag depends on the mass of the satellite and this is a time-dependent quantity, we expect stripping to ultimately affect the orbital decay rate. Somerville & Primack (1999) include a simple recipe which accounts for mass stripping, and gradually reduces the mass of the satellite by re-calculating its tidal radius while it spirals towards the center along a circular orbit.

Colpi, Mayer & Governato (1999; hereafter CMG) quantify the interplay between dynamical friction and tidal stripping for a selected sample of orbits and satellite masses. Using fairly high-resolution N-body simulations, they show that small satellites (with initial masses 50 times smaller than that of the primary halo) undergo tidal mass loss and that their orbit decays as if they had an "effective mass" $\sim$ 60 per cent lower than the initial one. On typical cosmological orbits they never sink into the center of the primary halo because the magnitude of the drag is drastically reduced for such small effective masses. The fraction of mass lost by the satellite is strictly related to the particular orbital parameters and halo density profile assigned to the haloes. In order to improve this recipe and make it more physically motivated, it is necessary to recognize that mass loss is the consequence not only of the initial tidal truncation but also of the repeated gravitational shocks occurring at each pericenter passage (Weinberg 1994a,b,c; Gnedin, Hernquist & Ostriker 1999; Taylor & Babul 2000). This regime of disruption is completely neglected by semi-analytical models of galaxy formation.

The full dynamical evolution of the satellites must be studied using haloes that have cuspy density profiles similar to those forming in cosmological simulations. In this thesis, I consider haloes with Navarro, Frenk & White (1996; 1997) profiles while in previous works the analysis was restricted to isothermal spheres with cores (CMG). However, more recent higher resolution simulations (Moore et al. 1999; Ghigna et al. 2000; Bullock et al. 2000; Jing & Suto 1999) find the inner slope of the density profile to be even steeper than the Navarro, Frenk & White one.

Following the same philosophy as in CMG, I use semi-analytical methods to describe the orbital evolution and mass loss of satellites, and compare the results with high resolution N-body simulations. In particular, I use the theory of linear response to model dynamical friction and study orbital decay (Colpi 1998; Colpi & Pallavicini 1998) and apply the theory of gravitational shocks developed by Gnedin, Hernquist & Ostriker (1999) to model both the tidal mass loss and the disruption of satellites (Taylor & Babul 2001; Hayashi et al. 2002).

Initial satellite sizes, masses, orbital energies, and eccentricities are selected as predicted by hierarchical models of structure formation. The interplay between dynamical friction and tidal mass loss/evaporation in determining the final fate of the satellite is then explored. Finally, I provide a quick, user–friendly, expression for the dynamical friction timescale of a *live* (i.e. mass loosing) satellite and for the disruption time–scale. These expressions can then be coupled with semi–analytical codes for structure formation to follow in detail the evolution of substructures.

The structure of the thesis

The layout of this thesis is the following. In Chapter 2, I introduce some basic elements of modern cosmology and devote special attention to recent observational constraints on cosmological parameters. I also review the theory of gravitational instability, paying particular attention to the theory of collapse of the initial density perturbations. Chapter 3 deals with the issue of hierarchical clustering. I present a Monte Carlo code to generate catalogues of haloes based on the EPS formalism, and then compare its results with numerical simulations. In Chapters 4, 5, and 6 PINOCCHIO code is presented and tested. First, I describe the analytical backbones of this algorithm and then test the code against numerical simulations in order to verify its ability of estimating the statistical properties of the hierarchical clustering of haloes. I will show that PINOCCHIO can also reproduce the numerical experiments to an object-by-object level. In the last two Chapters, I study the evolution of satellites in DM haloes. Chapter 7 derives a simple formula for the decaying time of rigid satellites orbiting in Navarro Frank & White haloes, while Chapter 8 studies the fate of mass loosing satellites which sink in the main DM halo. Conclusions are presented in the last sections of the chapters containing original results (namely, Chapter 3,4,5,6,7,8) and summarized in Chapter 9.

Chapter 2

Hierarchical Cosmology

> *If the doors of perception were cleansed*
> *everything would appear as it is,*
> *infinite*
>
> W. Blake

The cosmology of the last century was driven by an extraordinary increase of theoretical and observational results which contributed to the formalization of the standard description of the cosmic evolution: the "hot Big Bang" model. The main kinematic and dynamical properties of the model are based on the Einstein's theory of General Relativity and on the assumption that the "cosmological fluid" is homogeneous and isotropic on large scales (*cosmological principle*). These hypothesis, even if confuted by the strong inhomogeneity on galaxy scales, are supported by various observational constraints on much larger scales, for instance the isotropy of the temperature of the *cosmic microwave background radiation*.

The isotropic and homogeneous Universe can be described by an appropriate choice of the space-time metric: the Robertson-Walker metric.

In this framework, small amplitude perturbations in the initial density field are thought to be the seeds of the cosmic structures observed today. They grow by gravitation instability and evolve through a series of complex highly non-linear dynamical processes.

This Chapter is devoted to illustrate the cosmological models and the process of structure formation according to these models. The treated arguments are standard in cosmology and they are described in details in various books like Weinberg (1972), Peebles (1980), Peebles (1993) Lucchin & Coles (1995), Peacock (1999).

2.1 The theoretical model

The cosmological principle allows to construct a proper theoretical model based on the Robertson-Walker space-time metric

$$ds^2 = c^2dt^2 - a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right] . \tag{2.1}$$

Here k gives the sign of the curvature of the spatial hypersurface and it assumes values $k = +1, 0, -1$. The scale factor $a(t)$ is a multiplicative term that transforms the dimensionless distances computed using coordinates on homogeneity hypersurface into physical ones; $R^{-2} = c^2k/a^2$ is the curvature parameter of the spatial section.

Assuming that the Universe can be described as a perfect fluid, it is possible to resort to the General Relativity to describe the dynamics of the scale factor $a(t)$ in terms of the Friedmann equations:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} , \tag{2.2}$$

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2} = \frac{8\pi G}{3}\rho + \frac{\Lambda c^2}{3} , \tag{2.3}$$

where Λ is the cosmological constant, historically introduced as a constant of nature.

To determine the evolution of $a(t)$, it is necessary to give an equation of state for the matter or radiation present in the Universe. For example for non-relativistic pressureless matter $p = 0$ and $\rho(t) \propto a^{-3}$. In most cases the equation of state can be written as $p = w\rho$ and the energy density evolves as

$$\rho(t) \propto a^{-3(1+w)} . \tag{2.4}$$

Expansion and contraction of the universe are expected to give strong observational signatures. For example as a consequence of Hubble expansion the light sources in the Universe are redshifted, and the cosmological redshift is related to the expansion factor a by:

$$1 + z(t) = \frac{a(t_0)}{a(t)} , \tag{2.5}$$

where t_0 is the present time and the redshift z is measured for an object at the cosmological epoch t.

The normalization of the scale factor is arbitrary; it is usually normalized as $a(t_0) = 1$. The Hubble parameter is defined as:

$$H(t) = \frac{\dot{a}}{a} ; \tag{2.6}$$

the Hubble constant is then $H(t_0) = H_0$. If $\Lambda = R = 0$, the Universe is flat (Einstein-de Sitter model), the background density in this case is called critical density:

$$\rho_{\rm cr}(t) = \frac{3H^2}{8\pi G} . \tag{2.7}$$

It is then convenient to define, for any Friedman-Robertson-Walker (FRW) cosmology, the following density parameter:

$$\Omega(t) = \rho / \rho_{\rm cr} . \tag{2.8}$$

$\Omega = 1$ denotes the Einstein-de Sitter cosmology; in this case it remains constant in time. If $\Omega < 1$, and the cosmological constant is null or not large enough, then the Universe is open; if $\Omega > 1$ (and there is no negative cosmological constant), the Universe is closed. The cosmological density parameter Ω_0 is:

$$\Omega_0 = \Omega(t_0) . \tag{2.9}$$

The three constants H_0, Ω_0 and Λ, together with the normalization of the scale factor, define uniquely the FRW background model. The Hubble constant is usually parameterised as:

$$H_0 = h \cdot 100 \, \mathrm{km \, s^{-1} \, Mpc^{-1}} , \tag{2.10}$$

where h contains our ignorance on the real value of H_0. Recent measures of the Hubble parameter done by the HST Key Project give $h = 0.72 \pm 0.8$ (Freedman e al. 2001).

In the following, only three families of FRW cosmological models will be considered (the others are not of physical interest):

1) flat models with no cosmological constant: $\Omega = 1, \Lambda = 0$;

2) open models with no cosmological constant: $\Omega < 1, \Lambda = 0$;

3) flat models with positive cosmological constant: $\Omega < 1, \Lambda \neq 0, \Omega + \Omega_\Lambda = 1$, where $\Omega_\Lambda = \Lambda / 3H_0^2$.

In model (1), the FRW equation becomes:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho \,, \tag{2.11}$$

whose solution is:

$$a(t) = \left(\frac{t}{t_0}\right)^{2/3} \,. \tag{2.12}$$

In model (2):

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho\left[1 + \left(\frac{1}{\Omega_0} - 1\right)a\right] \,, \tag{2.13}$$

and the $a(t)$ evolution can be expressed through the following parametric representation:

$$\begin{aligned} a(\eta) &= \frac{\Omega_0}{2(1-\Omega_0)}(\cosh\eta - 1) \\ t(\eta) &= \frac{\Omega_0}{2H_0(1-\Omega_0)^{3/2}}(\sinh\eta - \eta) \,. \end{aligned} \tag{2.14}$$

In model (3):

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho\left[1 + \left(\frac{1}{\Omega_0} - 1\right)a^3\right] \,, \tag{2.15}$$

$$a(t) = \left(\Omega_0^{-1} - 1\right)^{-1/3}\sinh^{2/3}\left(\frac{3}{2}\sqrt{\frac{\Lambda}{3}}t\right) \,. \tag{2.16}$$

2.2 A brief history of the early Universe: theory and observations

In the hot "Big Bang" model the Universe evolves from a hot and dense early stage to the present state. In the following we will describe the main stages that characterise this evolution, with a special attention on some precise epochs, that will be important for the discussion of structure formation and evolution. We will not enter into details of the physical processes that characterise the first moments after the "Big Bang" which can be described using the Grand Unified Theory developed in quantum field Physics. We just notice that a large variety of particles is produced by symmetry braking phase transitions, most of them disappears with the progressive cooling of the Universe caused to Hubble expansion (see for ex. Weinberg 1972 or Peacock 1999).

2.2.1 Epochs in the Universe

When the reaction rate of a given particle interaction becomes smaller than the Hubble expansion rate, the interaction is no longer occurring and the number density of the involved species is "frozen". The knowledge of the time scale for particle interaction allows to identify important moments in the thermal evolution of the Universe.

The Radiative era

The radiative era begins when the typical energy per particle becomes lower than 0.5 MeV: the electron-positron pairs annihilate and the content of the Universe consists of a radiative component (photons and non interactive mass-less neutrinos) and of a non-relativistic matter mixture (neutrons, protons, electrons). In the radiative era, which accounts for the first $\sim$100 seconds of the Universe, the photonic component dominates and the first simple nuclei form. The outcome of the nucleosynthesis can be computed with very small uncertainties (see Weinberg 1972, Peacock 1999) and compared with the observational measures. About 24% of the baryonic mass goes into ^{4}He, and protons constitutes the remaining fraction. A tiny fraction of the total baryons converts into light isotopes (D, ^{3}He or ^{7}Li) whose abundances are strongly related to the baryon density $\Omega_{\rm b}$ and to the

Hubble parameter h. A determination of these abundances would then constrain the value of $\Omega_{\rm b}h^2$:

$$\Omega_{\rm b}h^2 = 0.019 \pm 0.0095 \ , \tag{2.17}$$

(Kaplinghat & Turner 2001). This result is compatible with the estimates of $\Omega_{\rm b}$ based on recent observations of the CMB power spectrum:

$$\Omega_{\rm b}h^2 = 0.022 \pm 0.004 \ , \tag{2.18}$$

(de Bernardis et al. 2001).

The Matter-Radiation equality

When the energy content of the Universe comes to be dominated by the non-relativistic matter which is still coupled with the radiation counterpart by bound-free and free-free interaction, we enter the epoch of Matter-Radiation equality. The time scale for this event is $z_{\rm eq} \sim 2.3 \cdot 10^4/(\Omega_{\rm b}h^2)^2$.

The Re-Combination

When $z_{\rm eq} \sim 1300$ matter reduces its coupling with photons and starts to become neutral.

Decoupling and the CMB

Decoupling of radiation with matter inhibits the gravitational collapse of the density perturbations which freeze until the "decoupling time". At $z_{\rm dec} \sim 1100$ the photons decouple completely from the baryons and subsequently cool following the Hubble expansion. They set the perturbation free to collapse and the gravitational instability forms the first bound structures in the Universe: the large scale structures (LSS).

The mean free path of the radiation rapidly increases so that photons are no longer scattered and they "reach" the present Universe as a low temperature black-body radiation. This background radiation was discovered by Penzias & Wilson (1965).

The COsmic Background Explorer (COBE) satellite confirmed and improved the results pointed out by Penzias & Wilson (1965), measuring the thermal distribution of photons at temperature

$$T = 2.725 \pm 0.002 \ {\rm K} \ ,$$

95 per cent confidence (Mather et al. 1999).

The brightness temperature of the CMB is very uniform across the sky but it reveals some important anisotropies. Some of them are related to the proper motion of the Earth or of the Galaxy (the so called dipole anisotropy of one part in 10^3; Lineweaver et al. 1996), but others on large, intermediate or small scales are due to physical effects that took place at recombination. Gravitational (Sachs-Wolfe) or intrinsic (adiabatic) perturbations of the CMB temperature are the signature on large and intermediate angular scales, of the presence of fluctuations in the initial density field.

The large scale Sachs-Wolfe perturbations were measured by COBE which revealed a temperature anisotropy of about one part in 10^5. Recent, successful, balloon experiments, BOOMERanG and MAXIMA, mapped the the CMB on $\sim$1 degree scale verifying the presence of small scale anisotropy providing a strong evidence for primeval density inhomogeneities of amplitude $\sim 10^{-5}$ (de Bernardis et al. 2000; Hanany et al. 2000).

2.2.2 The inflationary paradigm

How did primordial density fluctuations generate?
A plausible explanation is given by the so called *inflationary scenario* (Linde 1990). The main idea is that at a certain time in the early Universe, after the Plank epoch, a causally connected part of the Universe underwent an accelerated expansion (the inflation) caused by a homogeneous scalar field whose vacuum energy happened to dominate, giving rise to negative pressure. Quantum fluctuations of this scalar field generated the seeds of the observed density perturbations. The statistical properties of the fluctuations are related to the properties of the inflation itself. In general, the density perturbations form a Gaussian random field with a nearly scale invariant power spectrum.

The inflationary paradigm has been invoked as a possible mechanism to solve a number of other important questions:

- The flatness problem: why the Universe is so flat?

- The horizon problem: how is it possible that even though the particle horizon on the last scatter surface sub-tends about one degree on the sky the

CMB is homogeneous all over the sky?

- The relic problem: why are we not detecting any observational signature of the presence of topological defects which should be created during the cosmological phase transition in the early Universe?

There are several inflationary models: all of them answer the previous questions. During inflation the cosmological density is driven toward 1 and the Universe flattens, moreover the density of the unwanted relics is extremely diluted by the enormous amount of expansion. Finally the horizon problem is solved because all the observable Universe is part of a larger region which was causally connected in the past.

2.3 Statistics of the primordial fluctuations

The outcome of the inflationary scenario is a flat Universe where randomly distributed density fluctuations do not affect the space-time metric on large scales. The statistical properties of the primordial density perturbations depend on the matter content of the Universe and on the inflation itself.

We define the density contrast in the point x as:

$$\delta(\mathrm{x}) = \frac{\rho(\mathrm{x}) - \bar{\rho}}{\bar{\rho}} , \tag{2.19}$$

this dimensionless quantity is bound to be larger than -1, while it can grow up to infinite (very large) positive values. Its Fourier transform is:

$$\tilde{\delta}(\mathbf{k}) = \frac{1}{(2\pi)^{3/2}} \int d\mathrm{x} \delta(\mathrm{x}) \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{x}} . \tag{2.20}$$

We can define the power spectrum of the field to be the autocovariance of $\tilde{\delta}(\mathbf{k})$:

$$P(k) = \langle \tilde{\delta}(\mathbf{k}) \tilde{\delta}(\mathbf{k}') \rangle \delta_D(\mathbf{k} + \mathbf{k}') , \tag{2.21}$$

where δ_D is a Dirac delta function which accounts for the symmetry and reality of $P(k)$.

In the case of a Gaussian random field the power spectrum provides a complete statistical characterisation of the density field. With Gaussian statistics, phases are randomly distributed.

Many physically-interesting functions are related to the power spectrum. For example the two-point correlation function, $\xi(x)$, which is simply the Fourier transform of the power spectrum:

$$\xi(x) = \frac{1}{(2\pi)^{3/2}} \int d\mathbf{k} P(k) e^{i\mathbf{k}\cdot\mathbf{x}} , \tag{2.22}$$

or the moments of the power spectrum:

$$\sigma_n^2 \propto \int P(k)\, k^{2n}\, d^3k \ . \tag{2.23}$$

It is in the common use to resort to a more physical quantity: the mean square fluctuation smoothed on scale R, *the mass variance*:

$$\sigma^2(R) = \int d\mathbf{k} P(k) \tilde{W}(\mathbf{k}; R) , \tag{2.24}$$

where all the perturbations on scales smaller than R are filtered out by the Fourier transform of the filter function $W(x)$.

To complete our review of the statistical properties of the primordial density field we need to specify the shape of the initial power spectrum. The inflationary scenario produces a power law form:

$$P(k) = Ak^n , \tag{2.25}$$

where the primordial spectral index n is predicted to be equal to 1 by different inflationary models.

Comment: Even if the inflation scenario was not responsible for the formation of density perturbations a power law is a useful phenomenological parametrization for the power spectrum.

The normalization of the power spectrum is usually empirically assigned. Before the observations of the CMB by the COBE satellite, the spectrum was normalised by requiring that the mass variance on the scale of $8h^{-1}$Mpc, $\sigma_8 = 1$ as suggested by the analysis of the galaxy catalogues. After COBE flight the spectrum is normalised using the measures of the CMB fluctuations.

2.4 The growth of perturbations: linear theory

After recombination, the density perturbations grow and evolve passing through a number of different states. During the first epoch of their evolution the fluctuations are still small compared to the background ($\delta \ll 1$) and their evolution can be successfully described using the linearised equation of motion for the gravitational instability.

Linear theory is presented in full detail in Peacock (1999), here we briefly present its main features.

2.4.1 Perturbation

The evolution of perturbations is followed by approximating cosmological matter as a perfect pressureless fluid. It is convenient to subtract the effect of Hubble expansion from the evolution of perturbations. If $\mathbf{r}$ is the physical coordinate, the comoving coordinates x is defined as:

$$\mathbf{r} = a\mathrm{x} \ . \tag{2.26}$$

In this way, a point comoving with the background has fixed comoving coordinates x. The peculiar velocity is defined as:

$$\mathrm{v} = \dot{\mathbf{r}} - H\mathbf{r} = a\dot{\mathrm{x}} \ . \tag{2.27}$$

It is possible to construct a Newtonian peculiar potential for matter fluctuations:

$$\nabla_{\mathrm{x}}^2 \Phi = 4\pi G(\rho - \bar{\rho})a^2 \ . \tag{2.28}$$

This Poisson equation, which connects the peculiar potential to the matter distribution, is one of the equations for the evolution of matter perturbations. The subscript x in the ∇ operator indicates that the differentiation is performed with respect to the comoving coordinate.

The gradient of the peculiar potential gives the gravitational force acting on fluid elements. The Euler equation of motion for a generic fluid element is:

$$\frac{\partial \mathrm{v}}{\partial t} + \frac{1}{a}\left(\mathrm{v} \cdot \nabla_{\mathrm{x}}\right)\mathrm{v} + \frac{\dot{a}}{a}\mathrm{v} = -\frac{\nabla_{\mathrm{x}}\Phi}{a} \ . \tag{2.29}$$

The last evolution equation is the continuity one:

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla_{\mathrm{x}} \cdot (1+\delta)\mathbf{v} = 0 \; . \tag{2.30}$$

2.4.2 Eulerian perturbation theory

When the density contrast $\delta \ll 1$ and peculiar velocities are small enough to satisfy $(vt/d)^2 \ll \delta$, where t is the cosmological time and d is the coherence length of the matter field, the system of equations (2.28 – 2.30) can be linearised, leading to the equation:

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G\bar{\rho}\delta \; . \tag{2.31}$$

This second-order equation has two solutions, representing a growing and a decaying mode. The growing mode, denoted by $b(t)$, is the one of cosmological interest, responsible for the growth of small perturbations. It is always assumed that the decaying mode has already faded away. For this reason, the relation between the velocity field and the density contrast can be simplified as follows:

$$\begin{aligned} \nabla_x \cdot \mathbf{v} &= -a\frac{\partial \delta}{\partial t} \\ \nabla_x \times \mathbf{v} &= 0 \; . \end{aligned} \tag{2.32}$$

An interesting consequence is that any primordial vorticity is damped out by linear evolution.

In the following, the solutions of Eq. (2.31) for the growing modes relative to the three background models of §2.2.1, are reported:

$\Omega = 1$:

$$b(t) = a(t) \; . \tag{2.33}$$

$\Omega < 1, \Lambda = 0$: it is useful to use the time variable

$$\tau = \frac{1}{\sqrt{1-\Omega(t)}} = \sqrt{\left[a\left(\frac{1}{\Omega_0} - 1\right)\right]^{-1} + 1} \; . \tag{2.34}$$

Then:

$$b(\tau) \propto \left[1 + 3(\tau^2 - 1)\left(1 + \frac{\tau}{2}\ln\left(\frac{\tau - 1}{\tau + 1}\right)\right)\right] . \tag{2.35}$$

$\Omega < 1, \Lambda \neq 0$, flat: it is useful to use the time variable

$$h = \coth(3t\sqrt{\Lambda/3}/2) \ . \tag{2.36}$$

Then,

$$b(\tau) = h \int_h^{\infty} (x^2(x^2-1)^{1/3})^{-1} dx \ . \tag{2.37}$$

Growing modes are normalised to give $b(t) \simeq a(t)$ at early times, and $a(t_0) = 1$.

It is convenient to define the quantity:

$$\delta_l \equiv \delta(t_i)/b(t_i) \ . \tag{2.38}$$

This is the initial density contrast linearly extrapolated to the time at which $b(t)$=1, which, in an Einstein-de Sitter background, is the present time; it will be used in next chapters.

Using the growing mode $b(t)$ as a time variable, it is possible to write the equations of motion (2.28 – 2.30) in a more compact way. Defining the peculiar velocity $\mathbf{u}$ as

$$\mathbf{u} = \frac{d\mathbf{x}}{db} = \frac{\mathbf{v}}{a\dot{b}} \ , \tag{2.39}$$

the Lagrangian (convective) derivative d/db as

$$\frac{d}{db} = \frac{\partial}{\partial b} + \mathbf{u} \cdot \nabla_{\mathbf{x}} \ , \tag{2.40}$$

and the rescaled peculiar gravitational potential as

$$\varphi = 2a\Phi/3bH_0^2\Omega_0 \ , \tag{2.41}$$

the following system of equations is obtained:

$$\frac{d\mathbf{u}}{db} + \frac{3}{2b}\frac{\Omega}{f^2(\Omega)}\mathbf{u} = -\frac{3}{2b}\frac{\Omega}{f^2(\Omega)}\nabla_{\mathbf{x}}\varphi \tag{2.42}$$

$$\frac{d\delta}{db} + (1+\delta)\nabla_{\mathbf{x}} \cdot \mathbf{u} = 0 \tag{2.43}$$

$$\nabla_{\mathbf{x}}^2\varphi = \frac{\delta}{b}. \tag{2.44}$$

The function $f(\Omega) = \dot{b}/Hb \simeq \Omega^{0.6}$ is defined in Peebles (1980); note that $\Omega/f^2(\Omega) \simeq \Omega^{-0.2}$ is weakly time-dependent.

2.5 Choosing a cosmological model

A given cosmological model is specified by a set of cosmological parameters. A choice of those parameters constrains not only the geometry and the matter-energy content of the Universe, but also for example the initial condition which seeded the formation of structures. In this Section, we propose recent observational and theoretical results which constrain some of the cosmological constants and allow us to make a choice of the cosmological framework where structure forms and evolves.

2.5.1 The presence of dark matter

It is usually accepted that the matter content of the Universe is in the form of two non-interacting fluids: one of them is made up of baryons, the other is assumed to be composed by dissipationless and non baryonic particles.

The need to assume the presence of a non-baryonic matter arises because of the necessity to generate large enough perturbations to account for the presence of galactic haloes, from sufficiently small initial values of the density contrast δ.

In pure baryonic models a perturbation grows linearly just after the recombination epoch, with the expansion factor $a(t)$. The maximum permitted growth for perturbations is therefore $\delta/\delta_{rec} \simeq 1100$ (for $\Omega \leq 1$). Consequently, the necessary but not sufficient condition for the formation of gravitationally bound systems (such as galaxies), that $\delta > 1$ today, leads to the primordial amplitude: $\delta \approx 10^{-4}$ at recombination. That kind of perturbations should induce an anisotropy of $\sim 10^{-3}$ on the temperature of the CMB radiation on arc minutes scale which has not been observed. On the other hand, perturbations in a non–baryonic component, can start growing soon after matter–radiation equality (i.e. by $z \sim 10^4$ in the case of an $\Omega = 1$ Universe) with the result that large final perturbations can develop from small initial values without violating the CMB constraints on small scales.

The existence of non–baryonic matter is also strongly supported by other observational results, which suggest that the most of the matter in the Universe is not detectable by present instrumentation and reveals its presence just through gravitational interaction.

We briefly review the main observational evidences for the presence of the *dark matter* (DM).

- Combining the most recent nucleosynthesis constrain on the baryonic contribution to the density parameter (§2.2.1) and the uncertainties on H_0 we obtain $0.026 \leq \Omega_b \leq 0.08$. However, there is a ample evidence that the total Ω is at least 0.2 (§2.5.2).
- The flatness of the rotational curves of spiral galaxies up to scales much larger than the typical length of their luminosity distribution (Persic & Salucci 1995).
- X-ray observation of extended haloes of hot gas around elliptical galaxies presumably retained by a large DM halo around the luminous matter (Fabian 1986).
- Gravitational lensing effects, recent results (e.g. Grogin & Narayan 1996; Sheldon, et al. 2002) suggest that lens galaxies required dark haloes.

To conclude this Section, we notice that several attempts have been made to measure the amount of mass (baryonic and DM) present in the Universe. Even if it is not possible to fix a value for Ω_m, many observations (e.g. Myers et al. 1997; Mohr et al 1999; Borgani et al. 2001) allow to place it in the range bounded by:

$$0.15 < \Omega_m < 0.5 \ . \tag{2.45}$$

Large scale data on galaxy clustering consistently assign $\Omega_m h \sim 0.20 \pm 0.03$ (Percival et al. 2001).

These results and the determination of the baryonic matter strongly suggest that the dominant component for the mass in the Universe has to be some sort of DM. Particle Physics provides some viable candidates for the DM particles: a light massive neutrino (with mass $\sim$ 10 eV), a very light axion (with mass $\sim$ 10^{-5} eV), and the SUSY particle (possibly the neutralino with mass in the range of 10 – 500 GeV). The nature of the DM particles strongly affects the evolution of the cosmic fluid and consequently the structure formation process: the large velocity dispersion of a hot DM component (like the massive neutrino) smear out

Table 2.1: Current constraints (68% c.l.) from CMB data plus prior information from high redshift type Ia supernovae

Dataset	Ω_Λ	Ω_m
BOOMERanG + COBE + SNe Ia	$0.72^{0.05}_{0.04}$	$0.37^{0.07}_{0.07}$
MAXIMA + COBE + SNe Ia	$0.60^{0.07}_{0.07}$	$0.37^{0.07}_{0.07}$
MAXIMA + BOOMERanG + COBE + SNe Ia	$0.75^{0.06}_{0.07}$	$0.35^{0.07}_{0.07}$

all small scale perturbations which can form by subsequent fragmentation of large scale structures (top-down scenario). On the other hand the presence of a cold DM (CDM) particle (like the neutralino) preserves small perturbations and allows the formation of structures with $\sim 10^6 \mathrm{M}_\odot$ soon after recombination which merge to form bigger structures as time passes (bottom-up scenario).

2.5.2 Evidence for a cosmological constant

SUPERNOVAE are individual objects that can be resolved inside distant galaxies; SNe Ia are almost homogeneous in their properties, and can be used as "standard candles" in distance measures. The key assumption is that the maximum luminosity of these supernovae does not change as a function of cosmic age. The measure of distances using this kind of indicators evidences an accelerated expansion for the Universe which is consistent with the existence of a non zero cosmological constant. For example, Perlmutter et al. (1999) find that $\Omega_\Lambda \neq 0$ with a confidence level of 99%, their results can be synthesised as:

$$0.8\Omega_m - 0.6\Omega_\Lambda \simeq -0.2 \pm 0.1 \, . \tag{2.46}$$

A fundamental test to constrain the value of Ω_Λ comes from the study of the CMB radiation. In particular the position of the first acoustic peak in the CMB power spectrum is sensitive to the value of $\Omega_m + \Omega_\Lambda$.

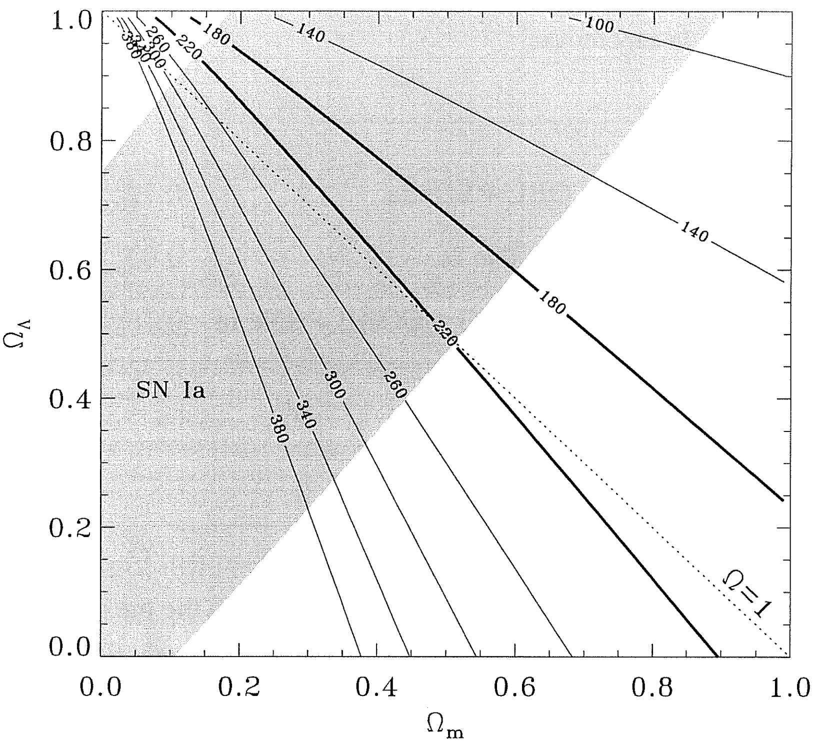

Figure 2.1: Degeneracy lines in the Ω_m–Ω_Λ plane, assuming $h = 0.7$ and $\Omega_b h^2 = 0.03$ (Balbi 2001). The lines correspond to models having approximately the same angular scale of sound horizon at decoupling, and are labelled by the approximate position of the first peak in the CMB angular power spectrum. The ticker lines give a reference interval corresponding to $180 \leq \ell_{peak} \leq 220$. Flat models lie on the diagonal dotted line ($\Omega = 1$). The shaded curve represents the 95 % confidence levels from high redshift type Ia supernovae.

BOOMERanG (de Bernardis et al. 2000) and MAXIMA (Hanany et al. 2000) balloon missions dataset were used independently to set constrains on the value of a cosmological constant (e.g. Balbi 2001). The intersection of CMB detection results with the observational constraint from SNe Ia are summarised in Tab (2.1). Those results strongly suggest the presence of a non-zero cosmological constant Ω_Λ, even with some degeneracy (see e.g. Fig. [2.1]) in the choice of the preferred value.

Comment: The values proposed are in agreement with a post inflationary Universe, nearly flat, with initial density perturbations characterised by a scale-invariant power spectrum. The main matter component is a CDM fluid.

We conclude this Section noting that in this thesis we will assume that formation and evolution of structures takes place in a standard CDM (SCDM) and ΛCDM scenario.

2.6 The growth of perturbations: non-linear regime

As time proceeds the density fluctuations continue to grow sustained by gravitational instability until they turn non-linear: this is the moment when a collapsed structure forms. Structure formation is a complicated process which involves highly non-linear gravitational dynamics; moreover the non-locality of gravity makes the problem very difficult to solve.

Different methods have been developed to follow the evolution of the perturbations through the complicated transient configurations that characterise the non-linear regime. A complete review of those methods is beyond the scope of this thesis, here we will discuss the simplest class of non-linear approximations which involves extrapolations of the linear properties.

2.6.1 The spherical collapse

To identify collapsed, virialized structures we need to follow the evolution of the initial density perturbations. The simplest approach is to assign to each perturbation a simple geometrical shape, for example a sphere. In spherical symmetry the gravitational collapse can be solved exactly (Coles & Lucchin 1995). In fact, as a consequence of Birkhoff's theorem, a spherical perturbation evolves as a FRW Universe with density equal to the mean density inside the perturbation.

Let us consider a spherical density perturbation growing in an Homogeneous and Isotropic background Universe. This fluctuation is characterised by a density parameter Ω', which differs from the density parameter of the background

Universe Ω. For $\Omega' > 1$, the fluctuation expands to a maximum radius then turns around at a time $t_{\rm m}$ and collapses to a point at $t_{\rm c} = 2t_{\rm m}$. The spherical overdensity has a radius R and initial overdensity $\delta_{\rm i}$, let its mass be $M = 4\pi R^3\bar{\rho}(1+\delta_{\rm i})/3$ where $\bar{\rho}$ is the density of the background Universe; M is conserved during the collapse. Under the hypothesis that spherical shells of radius $R(t)$ do not cross during the evolution, we can write the equation of motion

$$R = A[1 - \cos(\theta)] \tag{2.47}$$

$$t = B[\theta - \sin(\theta)], \tag{2.48}$$

with A, and B function of $\delta_{\rm i}$.

Comment : The time evolution of a spherical perturbation is identical to that of spatially close FRW universe: the radius R evolves with time according to the cycloid equations.

We now want to compute the overdensity in each mass shell: using the mass conservation

$$\rho = \frac{3M}{4\pi R^3} = \frac{3M}{4\pi A^3(1-\cos\theta)^3}\;, \tag{2.49}$$

while for a flat matter–dominated Universe the background density scales as

$$\bar{\rho} = \frac{1}{6\pi G t^2} = \frac{1}{6\pi G B^2(\theta\sin\theta)^2}\;, \tag{2.50}$$

so that

$$\delta(\theta) = \frac{9}{2}\frac{(\theta\sin\theta)^2}{(1-\cos\theta)^3} - 1\;. \tag{2.51}$$

From the previous equations we find that at the maximum expansion $\delta(\pi) \simeq 4.6$ while at recollapse $\delta \to \infty$.

Eq.(2.51) allows us to recover the linear limit for small θ and consequently for small t:

$$\delta_{\rm l}(\theta) \sim \frac{3\theta^2}{20} \sim \frac{3}{5}\delta_{\rm i}\left(\frac{t}{t_{\rm i}}\right)^{2/3} \propto (\theta - \sin\theta)^{2/3}\;. \tag{2.52}$$

From this equation we obtain that $\delta_{\rm l}(\pi) \simeq 1.063$ at the turnaround and $\delta_{\rm l}(2\pi) \simeq 1.689$ at recollapse.

Warning: The exact spherical collapse is only an approximation, in reality the radius cannot reduce to a point: dynamical relaxation and shocks ensure that the system reaches virial equilibrium with finite density. At the equilibrium we can apply the virial theorem and we get $2R_{\rm vir} = R_{\rm m}$, $R_{\rm vir}$ is the radius of the equilibrium configuration, and $\rho_{\rm vir} = 8\rho_{\rm m}$.

While the internal density increases the density of the background Universe decreases, so – for an Einstein-de Sitter Universe – the ratio internal to external density at the equilibrium reaches a value of about 180. On the contrary from the linear theory, we got a contrast $\delta \simeq 1.686$.

2.6.2 The Press & Schechter approach and the mass function

If primordial fluctuations survive on small scales, the first structures to be formed after recombination come from the gravitational collapse of sub-galactic mass units. Those units cluster together in a hierarchy, forming successively more massive systems as time progresses. The evolution of those structures is roughly autosimilar in time. Sub-units are erased after they merge to form a new structure, but the total mass at each time is conserved. This picture is called *Hierachichal Clustering Scenario* (HCS).

The picture described in the HCS is too complicated to be dealt with any variant of the linear theory. However there is a method which attempts to explain some aspects of the Hierachichal Clustering in terms of properties of the linear density field. This method was developed by Press & Schechter (PS). It is based on the critical assumption that even if the field is non linear the amplitude of large-wave modes in the final density field is close to that predicted from linear theory. This means that large-scale power must exceed that generated from non-linear coupling of small-scale modes. PS produced a "numerical recipe" to obtain the so called cosmological mass function, i.e. the number density of collapsed structures at different redshift. We will discuss this method in some detail.

A massive clump is going to collapse if it is overdense with respect to the background; in other word the average overdensity in a volume containing that mass should exceed a fixed threshold $\delta_{\rm c}$. The spherical collapse model gives us a value for this threshold: $\delta_{\rm c} = 1.686$.

The density contrast grows in time proportionally to the linear growing factor $b(t)$; it is more convenient to switch the time dependency to the threshold $\delta_{\rm c}$ which evolves proportional to $1/b(t)$.

We consider the density field at some initial early time, and we smooth the field on a scale $R_{\rm f}$ using a filter function $W_{\rm R_f}(x)$:

$$\delta(\mathbf{r}, R_{\rm f}) = \int W_{\rm R_f}(\mathbf{r} - r')\delta(\mathbf{r}')d\mathbf{r}' \; . \tag{2.53}$$

If the filter is a top-hat function, $R_{\rm f}$ is the typical size of the filtered fluctuation, and it is reasonable to identify the mass associated to each fluctuation as:

$$M = \frac{4}{3}\pi\rho R_{\rm f}^3 \; . \tag{2.54}$$

We remind that the mass variance of the field σ^2 is connected to the Fourier transform of the filter function so it can be used as a resolution variable instead of $R_{\rm f}$. Note that σ^2 decreases when R increases.

If the density contrast is Gaussianly distributed, the probability that the linearly evolved density is larger then $\delta_{\rm c}$ is:

$$\Omega(\sigma) = \int_{\delta_c}^{\infty} P(\delta)d\delta = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\delta_c}{\sqrt{2\sigma^2}}\right)\right] \; . \tag{2.55}$$

$\Omega(< \sigma)$ is the fraction of collapsed mass while eq.(2.54) determines the mass associated with the scale σ^2. However, there is a problem here: as σ^2 diverges (the perturbation vanishes) $\Omega_{\delta_c} \to 0.5$. This is in contrast with what we expect in an hierarchical scenario where as $\sigma^2 \to \infty$ all the mass should be collapsed in objects of some scale. PS arbitrarily solved the problem by multiplying the mass fraction by a factor of 2.

The so called “golden rule” connects the mass function $N(M)$ to the fraction of collapsed mass:

$$N(M)MdM = \rho_0 \left|\frac{d\Omega}{dM}\right| dM = \rho_0 n(\sigma^2) \left|\frac{d\sigma^2}{dM}\right| dM \; , \tag{2.56}$$

in the previous equation we can isolate two terms: one, $n(\sigma^2) \equiv d\Omega/d\sigma^2$, connected with the statistic of the initial density field and with the dynamic of the

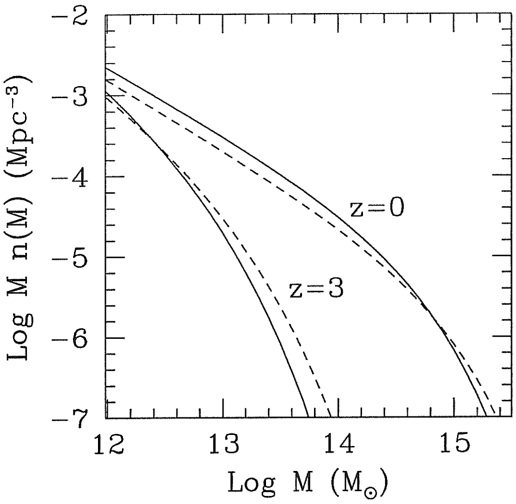

Figure 2.2: An example of mass function for a power-law type spectrum: $n = -2$. The solid lines are the PS mass function given by eq. (2.57) and the dashed lines are the Sheth & Tormen (1999) mass function (eq. [2.64]). We plot the mass function at two different redshifts.

collapse (through δ_c) and another, $d\sigma^2/dM$, with the spectrum of the perturbation.

From eq.(2.56) we have:

$$M^2 N(M) = \delta_c \rho_0 \sqrt{\frac{2}{\pi}\sigma^2} \exp\left(-\frac{\delta_c^2}{2\sigma^2}\right) \left|\frac{d\log\sigma^2}{d\log M}\right| . \tag{2.57}$$

which is the famous Press & Shechter formula.

For example if we have a power-law type spectrum:

$$\sigma = \left(\frac{M}{M_c}\right)^{-a} , \tag{2.58}$$

with $a = (n+3)/6$ (n is the spectral index), the PS formula gets a simple analytical form

$$N(M)dM = 2a\frac{\rho_0}{M_c^2}\frac{\delta_c}{\sqrt{2\pi}}\left(\frac{M}{M_c}\right)^{a-2} \exp\left[-\left(\frac{M}{M_c}\right)^{2a}\frac{\delta_c}{2}\right] dM . \tag{2.59}$$

The PS derivation is based on some sensible hypothesis, which are difficult to be theoretically justified. The great success of this "numerical recipe" is mainly due to its simple formulation, moreover for a long time it was thought to fit accurately N-body data (Efstathiou et al. 1988).

It is important to identify the main weakness of the PS derivation:

- Dynamical approximations (Zel'dovich or Adhesion for example) show that the matter distributes on a three dimensional framework made of sheets, filaments and nodes. It is evident that the distribution of the overdensities does not remain homogeneous at each epoch as PS theory states.

- The relation between σ and M is a crucial point. The simple eq. (2.54) is just a zero-order approximation. The collapsing clumps have an irregular geometry so a more sophisticated mass assignment has to be developed.

- Last but not least, the so called "cloud-in-cloud" (Bardeen et al. 1986) problem. As noticed before, the PS formula predicts that only half of the Universe is part of a lump of any mass. PS speculated that non-collapsed (underdense) regions have a finite probability of being included in larger collapsed ones, then the missing mass would accrete onto the collapsed lumps. They found that the net effect is to double the mass of every halo without altering the shape of the mass distribution.

2.6.3 The peak theory

If we assume that structures form in the high-density regions of the initial density field (Doroshkevich 1970) we can study the clustering properties of the non-linear objects looking at the clustering of the peaks of the linear field.

We define a peak as a region of the field $\delta(x)$ which reaches some critical value $\delta_{\rm c}$ (tipically $\delta_{\rm c} = 1.689$), it is common to call this region "excursion region". An ensemble of those regions is identified as an "excursion set". We need to select peaks that originate the non–linear structures. A first problem that arises is that the field δ has an infinite number of small scale peaks, but we do not want to take into account of the of very high frequency modes which correspond to scales much smaller that the galactic ones. If we assume that the large scale modes are not influenced by the small scale ones but they evolve linearly anyway, we can erase the unwanted frequencies smoothing the density field with some kind of low-pass filter.

Starting from a Gaussian field we can evaluate the correlation function of two excursion regions:

$$Q_2 = \int_{\delta_c}^{\infty} \int_{\delta_c}^{\infty} P_2(\delta_1, \delta_2) \mathrm{d}\delta_1 \mathrm{d}\delta_2. \tag{2.60}$$

with

$$P_N(\delta_1, ..\delta_N) = \frac{||\mathbf{C}||^{1/2}}{(2\pi)^{N/2}} \exp\left(-\tfrac{1}{2}\vec{V}^T \mathbf{C} \vec{V}\right), \tag{2.61}$$

where $\mathbf{C}$ is the correlation matrix and V is the column vector with δ_i components.

We will not enter into details of the peak theory (see e.g. Kaiser 1984, Katz, Quinn & Gelb 1997) anyway we notice that it does not seem a good approach to study the LSS formation. The first problem is that two close peaks of the density field that collapse at different time should generate two different galaxies. This process is not confirmed by the observations, so we need to introduce some mechanism to avoid the formations of those kind of structures: we select only the isolated peaks as seeds of the collapsed structures.

In spite of this, the peak theory is still an approximated approach to structure formation. It does not account neither for the dynamical processes that move the objects from their initial positions, nor for the tidal perturbations which could destroy the peaks. Moreover, the mass associated to a peak is not unique but it is strongly related to the theoretical approach we use (see e.g. Peacock & Heavens, 1985; Colafrancesco, Lucchin & Matarrese, 1989).

A possible extension of the peak formalism which account for dynamical processes is the *peak patch formalism* developed by Bond & Myers (1996a,b). They identified structures by considering peaks of an initial field, filtered on a hierarchy of scales. They defined the patch which is going to collapse with the peak, as the matter contained by a homogeneous ellipsoid which can collapse along its three axes. The peaks were then moved according to Zel'dovich approximation. This formalism correctly takes into account a number of important dynamical events, such as the effect of the shear on collapse dynamics (through the ellipsoidal model, see also §2.7.3); moreover, it has been found (Bond & Myers 1996b) to reproduce correctly the structures present in N-body simulations, but this time not only from a statistical point of view. This method is then able to generate Monte Carlo object catalogs in a faster way than N-body simulations, unfortunately it is too complex for an analytical or semi-analytical treatment.

2.6.4 The excursion set derivation

An alternative derivation of the PS formula is due to Bond et al. (1991). The so-called "excursion set" approach (see also Peacock & Heavens 1990; Cole 1991) overcomes the cloud-in-cloud problem and sets the mass function derivation on a more robust theoretical framework.

We want to study the statistical properties of the density fluctuation field at some resolution scale R. The statistical properties of the smoothed field $\delta(R)$ clearly depend on the choice of the window function. For a generic window no analytical calculation is possible, that because $\delta(R)$ contains correlation between different scales. But if we apply a filtering function which is a top hat in the Fourier space (*sharp k-filter*) the correlation simplifies: varying R corresponds to add k-space shells which are independent of each other if the fluctuation field is Gaussian.

Comment: A first problem now is how to associate a mass to the smoothed volume. In the real space the sharp k-filter is a diffuse function, so we cannot use eq. (2.54). A possible solution is to multiply the average density by the volume enclosed by the filter $M = 6\pi^2\rho_0 k_f^{-3}$ (Lacey & Cole 1993) where k_f is the width of the top-hat. Of course there is some arbitrariness in this choice, so we propose another procedure which guarantee good agreement with numerical simulations of clustering growth (Lacey & Cole 1994). If $\sigma^2_{\rm th}$ is the mass variance for a top-hat filter in the real space, we define $R_{\rm th}$ as $\sigma^2(R) = \sigma^2_{\rm th}(R_{\rm th})$ and the mass associated to the filter is given by eq. (2.54) with $R = R_{\rm th}$. **Warning:** Any top-hat-kind mass assignment to the sharp k-filter is arbitrary and actually inconsistent.

Let us consider for each point $\mathbf{x}$ of the space the trajectory $\delta(R)$ as a function of R. The largest R at which $\delta(R)$ upcrosses the threshold $\delta_{\rm c}(z_{\rm c})$ correspond to

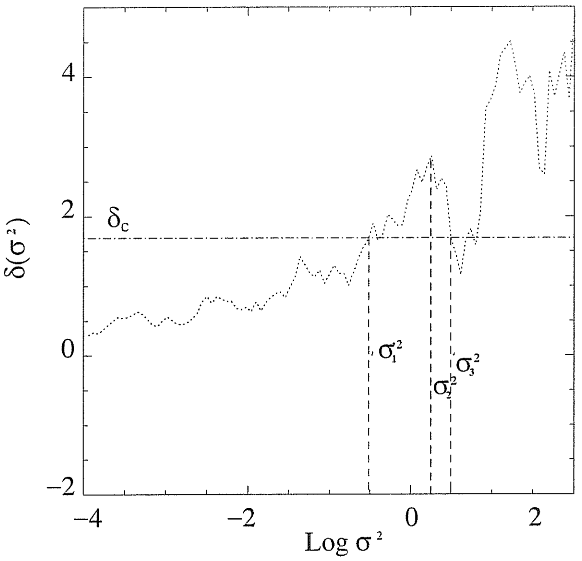

Figure 2.3: Random walk with absorbing barrier.

the largest mass collapsed at that point at redshift z_c, all substructures been erased. So the computation of the mass function is related to the calculation of the fraction of trajectories that first upcrosses the threshold $\delta_\mathrm{c}(z_\mathrm{c})$ as the scale M decreases.

Once again, it is convenient to use the mass variance as resolution variable instead of M. Because of the properties of the sharp k-filter, $\delta(\sigma^2)$ executes a Brownian trajectory as σ^2 increases (M decreases). In such case, we have to solve the Fokker-Plank equation for the probability density $P_{\delta_\mathrm{c}}(\delta,\sigma^2)d\delta$ that the stochastic process at σ^2 assume a value in the interval $\delta, \delta+d\delta$, with an absorbing boundary condition $P_{\delta_\mathrm{c}}(\delta_\mathrm{c},\sigma^2)=0$. We can write the probability that a random walk is absorbed by the barrier δ_c in the interval $\sigma^2, \sigma^2+d\sigma^2$ as:

$$f(\sigma^2,\delta_\mathrm{c}) = -\frac{\partial}{\partial\sigma^2}\int_{-\infty}^{\delta_\mathrm{c}} P_{\delta_\mathrm{c}}(\delta,\sigma^2)\,d\delta = \frac{\delta_\mathrm{c}}{\sqrt{2\pi\sigma^3}}\exp\left(-\frac{\delta_\mathrm{c}^2}{2\sigma^2}\right) . \qquad (2.62)$$

$f(\sigma^2,\delta_\mathrm{c})$ also correspond to the probability that a fluid element is a part of a collapsed halo with mass $M(\sigma^2+d\sigma^2), M(\sigma^2)$. Then we can easily recover the PS mass function, including the factor of 2 which is now naturally justified.

2.6.5 The Excursion Set and the Ellipsoidal Collapse

The great success of the PS formula was mainly due to to its accuracy in reproducing the N–Body data (Efstathiou et al. 1988). However, recent high resolution simulations revealed a discrepancy between the analytical formula and the numerical experiments (Gelb & Bertschinger 1994; Governato et al. 1999; Jenkins et al. 2001; Bode et al. 2001; see also §3).

This discrepancy is the result of the various oversimplifications of the Excursion Set model, in particular, the spherical collapse approximation. Sheth & Tormen (1999; Sheth, Mo & Tormen 2002) suggest that incorporating the effect of a non–spherical collapse in the excursion set formalism improves the agreement between the predicted mass function and that measured from the simulations.

The mass function is connected to the dynamic of the collapse through the collapse time $\delta_{\rm c}$. In the Excursion Set derivation the collapse time is just a function of redshift, so the absorbing boundary condition is constant in σ^2. Then, to incorporate the effects of the non–spherical collapse we need to determine the shape of the barrier associated to the dynamics of a non–spherical perturbation.

To describe the collapse of overdense regions we can use the ellipsoidal collapse model (Peebles 1980; Lee & Shandarin ; Bond & Myers 1996a). Following Bond & Myers (1996a), it is possible to derive an expression of the collapse time of a primordial elliptical perturbation of mass M:

$$\delta_{\rm ec}(\sigma^2, z) \propto \delta_{\rm c}(z) \left\{ 1 + \beta \left[\frac{\sigma^2(M)}{\delta_{\rm c}^2(z)} \right]^{\gamma} \right\} , \tag{2.63}$$

where $\beta = 0.47$ and $\gamma = 0.615$ (Sheth, Mo & Tormen 2002)

The random walk of the excursion set model is now absorbed by a 'moving' barrier $\delta_{\rm ec}(\sigma^2, z)$. We need to evaluate the probability that a random walk is absorbed by the barrier $\delta_{\rm ec}$ in the interval $\sigma^2, \sigma^2 + d\sigma^2$. Fitting the results of Monte Carlo independent and unconstrained realizations of random walks we can estimate the absorption probability. We can then rewrite eq. (2.62) in the case of the ellipsoidal collapse:

$$f_{\rm ec}(\sigma^2, z) = 2\,A \left[1 + \left(\frac{\sigma^2}{\delta_{\rm c,1}^2(z)} \right)^{q} \right] \frac{\delta_{\rm c,1}(z)}{\sqrt{2\pi}\sigma^3} \exp\left(-\frac{\delta_{\rm c,1}^2(z)}{2\sigma^2} \right) , \tag{2.64}$$

where $q = 0.3$, $A \approx 0.3222$ and $\delta_{\rm c,1}(z) = 0.8408\ \delta_{\rm c}(z)$, are fitted to reproduce N-body numerical data.

This simple modification to the PS formula provides a good fit to the mass function measured in the numerical simulation (Sheth & Tormen 1999; Sheth, Mo & Tormen 2002) curing the discrepancy of the original PS formula (see Fig. [2.2]).

2.7 Lagrangian approach

The local extensions of the linear theory provide a qualitative first step in describing the non linear regime of the evolution. Although a deeper insight into the gravitational clustering is provided by the dynamical approximations. The Lagrangian approach provides us with a qualitative understanding of the gravitational dynamics until the formation of the first collapsed structures. In this Section we will review the main features of this dynamical approach.

2.7.1 The Lagrangian Framework

Let us consider a fluidodynamic approach to follow the motion of a fluid element[1] of a set of collisionless self-gravitating particles in an expanding Universe. If we describe the system by means of observers comoving with the matter (the *Lagrangian* formulation of fluid dynamics, e.g. Shandarin & Zel'dovich 1989) we can label a given patch of the fluid with its Lagrangian comoving position q. The Lagrangian coordinate $\mathrm{q} = \mathrm{x}(t_{\rm i})$ represent the actual (Eulerian) comoving position at some initial time $t_{\rm i}$.

The trajectory $\mathrm{x}(\mathrm{q}, t)$ of a fluid element identified by the "Lagrangian label" q can be expressed in terms of a displacement vector field $\mathrm{S}(\mathrm{q}, t)$ as:

$$\mathrm{x}(\mathrm{q}, t) = \mathrm{q} + \mathrm{S}(\mathrm{q}, t)\ . \tag{2.65}$$

This equation recasts the problem of the evolution of a matter field in terms of a mapping procedure between the Lagrangian and the Eulerian space. When the time is small one expects the map $\mathrm{q} \to \mathrm{x}$ to be one-to-one, but as time passes the probability that two fluid elements get to the same point increases. This event

---

[1] Here we use "particle" and "fluid element" interchangeably.

is called *orbit crossing* (hereafter OC). When this happens the mapping becomes singular and a region with infinite density forms at x: the caustic. The situation after OC is called *multi-streaming* and the regions where OC happens are called *multi-stream regions*. From a mathematical point of view, OC corresponds to the instant when the Jacobian determinant $J \equiv \det(\partial \mathrm{x}/\partial \mathrm{q})$ vanishes.

Now the key quantity is the displacement vector field. The Eulerian-Poisson system of equations can be recast into an equivalent set of equations for $\mathrm{S}(\mathrm{q}, t)$ (see e.g. Catelan 1995), which becomes the only dynamical variable. For example, as the fluid element contains by construction a fixed but vanishingly small mass, the continuity equation becomes:

$$1 + \delta(\mathbf{q}, t) = \frac{1}{J(\mathbf{q}, t)} \ . \tag{2.66}$$

The effects of perturbations and the evolution of the kinematic and dynamical quantities relative to the fluid element can be singled out by defining the *deformation tensor* $\Delta = \partial \mathrm{S}/\partial \mathrm{q}$; $J = \det(I + \Delta)$ where I is the identity matrix. If the displacement field is irrotational then the deformation tensor is symmetric.

Comment: The main advantage of the Lagrangian formulation is that any Eulerian framework, which is intrinsically based on $\delta \ll 1$ (see §2.4.2), breaks down well before the OC, on the contrary the Lagrangian formulation allows to accurately follow the evolution of the matter field for a longer period of time.

The Lagrangian approach can be applied to smoothed versions of the initial field, in the hypothesis that small scale multi-streaming does not affect the evolution of larger scales. This turns crucial if we want to apply the Lagrangian theory to follow the evolution of the cosmic fluid; in fact, if the initial field is perturbed at small scales, than multi-stream regions dominates on small scales after recombination.

2.7.2 Lagrangian perturbation theory

In the Lagrangian framework the evolution of the cosmic fluid is followed using the standard equations of fluid dynamics: the continuity, energy and Poisson equation. The Lagrangian perturbative approach looks for solution of fluid equations for small displacement (i.e. the weak non-linear regime). Expanding S, and consequently J, in series:

$$\begin{aligned} \mathrm{S}(\mathrm{q},t) &= \mathrm{S}^{(1)}(\mathrm{q},t)+\mathrm{S}^{(2)}(\mathrm{q},t)+\mathrm{S}^{(3)}(\mathrm{q},t)+... & (2.67)\\ J(\mathrm{q},t) &= 1+J^{(1)}(\mathrm{q},t)+J^{(2)}(\mathrm{q},t)+J^{(3)}(\mathrm{q},t)+... & (2.68)\end{aligned}$$

we express them as the sum of a finite number of terms. In a spatially flat matter dominated Universe $\mathrm{S}^{(i)}$ are separable at every order: $\mathrm{S}^{(i)}(\mathrm{q},t)=\delta^{(i)}(t)\cdot\mathrm{S}^{(i)}(\mathrm{q})$.

The first order solution is the well-known **Zel'dovich approximation** (Zel'dovich 1970; Buchert 1992):

$$\mathrm{S}(\mathrm{q}\,t)=-b(t)\,\mathbf{u}(\mathbf{q})\;, \tag{2.69}$$

here $\mathbf{u}$ the peculiar velocity field of the fluid element, eq. (2.39), and $\varphi(\mathbf{q})$ is the rescaled peculiar gravitational potential, which obeys the Poisson equation:

$$\nabla^2\varphi(\mathrm{q})=\delta(\mathrm{q},t_i)/b(t_i)\equiv\delta_l(\mathrm{q}), \tag{2.70}$$

where t_i is an initial time at which linear theory holds (see §2.4.1).

A peculiarity of the Zel'dovich Approximation is that, using $b(t)$ as a time variable, eq. (2.69) looks just like the inertial motion of particles moving with constant velocities[2], as a result, particles intersect their trajectories leading the formation of singularities in the density field.

The deformation tensor of the first order solution is symmetric, and λ_1, λ_2 and λ_3 are its three eigenvalues, ordered according to:

$$\lambda_1\geq\lambda_2\geq\lambda_3 \tag{2.71}$$

(note that, because of Poisson equation, λ_1+λ_2+λ_3=δ_l). The Jacobian determinant J and the density contrast δ can be written in terms of the three eigenvalues, and evolve with time according to:

$$\begin{aligned} J(\mathrm{q},t) &= (1-b(t)\lambda_1)(1-b(t)\lambda_2)(1-b(t)\lambda_3)\;, & (2.72)\\ \delta(\mathrm{q},t) &= \frac{1}{(1-b(t)\lambda_1)(1-b(t)\lambda_2)(1-b(t)\lambda_3)}-1\;. & (2.73)\end{aligned}$$

---

[2] We notice that particles do accelerate in physical coordinates.

When the Jacobian determinant vanishes, the density formally goes to infinity. This corresponds to the formation of a caustic, a process discussed in detail by Shandarin & Zel'dovich (1989). After OC the approximation fails, particles maintain their velocity and the collapsed structures are smeared out.

The second order in the Lagrangian perturbation is the so called **post-Zel'dovich approximation**, where

$$\mathbf{S}(\mathbf{q},t) = -b(t)\mathbf{S}^{(1)}(\mathbf{q}) - \frac{3}{14}b^2(t)\mathbf{S}^{(2)}(\mathbf{q}) \, , \tag{2.74}$$

and the third order (**post-post-Zel'dovich approximation**, here after PPZ)can be written as

$$\begin{aligned} \mathbf{S}(\mathbf{q},t) = b(t)\mathbf{S}^{(1)}(\mathbf{q}) + \frac{3}{14}b^2(t)\mathbf{S}^{(2)}(\mathbf{q}) + \frac{1}{9}b^3(t)\mathbf{S}^{(3a)}(\mathbf{q}) \\ -\frac{5}{42}b^3(t)\mathbf{S}^{(3b)}(\mathbf{q}) - \frac{1}{14}b^3(t)\mathbf{S}^{(3c)}(\mathbf{q}) \, . \end{aligned} \tag{2.75}$$

The Lagrangian approximations break down after OC; to extend the approximation also to later epochs it is possible to filter the scales which are going nonlinear by means of a cutoff in the original power spectrum. The results are the so called truncated approximations: the truncated Zel'dovich approximation and the truncated post–Zel'dovich approximations (eg. Melott,Pellman, Shandarin, 1994; Buchert, Melott & Weiss 1994; Melott, Buchert & Weib 1995; Sahni & Coles 1995)

2.7.3 The definition of collapse

The OC corresponds to the instant when the Lagrangian approximations break down: the caustics form and the density "locally" diverges. It is reasonable, from the point of view of the mass element, to define such instant as its *collapse time* (Monaco 1995; 1997a). In this way collapse is well defined and easy to compute using the LPT which remains valid up to that point.

In the Zel'dovich approximation the mass element $\mathbf{q}$ collapses when $b(t_c) = 1/\lambda_1$ (see eq. [2.73]). The first-order LPT overestimates the growing mode at collapse time by nearly a factor of two[3]; this discrepancy is easy to understand:

---

[3] In the spherical case, the growing mode at collapse time is $3/\delta_l$, instead of the exact $1.69/\delta_l$ value.

the Zel'dovich approximation is exact (up to OC) in the case of planar symmetry and then it is able to describe the collapse of pancake-like structures (Shandarin & Zel'dovich 1989) but in the spherical limit, relevant for the collapse of high peaks, it underestimates the collapse rate (Monaco 1997a).

The post-Zel'dovich approximation is ill–behaved in under–densities, it can give false solutions which can make even a spherical void collapse (Sahni & Shandarin 1996). In conclusion, it is necessary to require the third-order LPT to calculate the collapse time of generic mass elements.

The ellipsoidal collapse

A homogeneous triaxial ellipsoid is characterised by its mean overdensity and its axial ratios; it can experience a global expansion, a deformation or a global rotation. It can be recognised that its properties are analogous to that of a mass element. The Lagrangian perturbative series can be truncated so as to resemble formally the collapse of an ellipsoid in an external shear field (Bond & Myers 1996a; Monaco 1995; Monaco 1997a):

Following Bond & Myers (1996a), to extract an ellipsoid from a perturbed potential field in a point $\mathbf{q}_0$, it suffices to expand the potential around that point in a Taylor series:

$$\varphi(\mathbf{q}) = \varphi(\mathbf{q}_0) + \frac{\partial \varphi}{\partial q_i}(\mathbf{q}_0) q_i + \frac{1}{2}\frac{\partial^2 \varphi}{\partial q_i \partial q_j}(\mathbf{q}_0) q_i q_j + \ldots \tag{2.76}$$

The first term is an unimportant constant; the second term produces a bulk motion of the mass element, but does not influence the internal properties of the ellipsoid. The third, quadratic term is the first one which is relevant for internal dynamics; it is then possible to approximate the potential as a quadratic form. The next step is to split the potential into an internal and an external term:

$$\varphi = \varphi_{\rm int} + \varphi_{\rm ext} \tag{2.77}$$

(this corresponds to the "extraction" of the ellipsoid). The second term, divergenceless, is supposed to give external tides. It can be held constant in the evolution of the ellipsoid: it is accurately constant in the linear and quasi-linear regime, while it becomes negligible, with respect to the internal potential, in the collapse phase (see Bond & Myers 1996a). In the principal frame of $\partial^2\varphi(\mathbf{q}_0)/\partial q_i \partial q_j \, q_i q_j$, the first term (the *internal ellipsoidal potential*), can be written as:

$$\varphi(\mathbf{q}) = \frac{1}{2}(\lambda_1 q_1^2 + \lambda_2 q_2^2 + \lambda_3 q_3^2) \ , \tag{2.78}$$

where λ_i are the three eigenvalues of $\partial^2\varphi/\partial q_i \partial q_j$.

Warning: The homogeneous ellipsoidal collapse has already been used in the cosmological context. White & Silk (1979) first proposed the ellipsoidal collapse model to describe the evolution of *extended* collapsing regions (see also Barrow & Silk 1981; Bartlemann, Ehlers & Shneider 1993; Eisenstein & Loeb 1995).

The initial conditions for the ellipsoid semi axes a_i at the initial time t_i are:

$$a_i = a(t_i)(1 - b(t_i)\lambda_i) \ , \tag{2.79}$$

where $a(t)$ is the scale factor. Note that at the initial time the ellipsoid is an infinitesimally perturbed sphere. With these initial conditions, the exact equations of ellipsoidal collapse can be integrated numerically (Monaco 1995; Bond & Myers 1996a). However, it is easier to solve exactly the third-order LPT equations in the ellipsoidal case of eq. (2.78), as only first and second derivatives of the peculiar potential are retained. This LPT solution gives a very good approximation to the numerical integration in all cases with the exception of the spherical limit. A small numerical correction is sufficient to recover properly this limit; this is described in Appendix B of Monaco (1997a). Apart from describing accurately the collapse of an ellipsoid, this solution gives a general approximation for the LPT evolution of a mass element under the action of gravity. This approximation, which will be denoted by ELL in the following, is easy to implement as it requires only the computation of the deformation tensor, while full third-order LPT requires the solution of many Poisson equations, thereby introducing numerical noise. Moreover, 3rd-order LPT still under predicts the quasi-spherical collapse of the highest peaks (a simple correction, as in the ELL case, is not feasible in this case), and consequently also the high-mass tail of the mass function. In general, ELL is an adequate approximation to compute the OC collapse time of generic mass elements.

In conclusion, it is worth stressing that in this context ELL is purely a convenient truncation of LPT; no constraint is put on the shape of the collapsing objects, nor on the 'shape' of the mass elements (which is simply a meaningless concept).

First-axis versus third-axis collapse

Recently there has been extensive discussion in the literature about whether the collapse of the first axis is enough to characterise gravitational collapse, or whether all three axes should reach vanishing size (Bond & Myers 1996a; Audit et al. 1997; Lee & Shandarin 1999; Sheth , Mo & Tormen 2002). Here we try to clarify this issue, showing that apparently contradictory claims result from different interpretations of ellipsoidal collapse, and from the choice of smoothing window.

As described previously, ellipsoidal collapse can be considered as a truncation of LPT, a convenient description of the dynamical evolution of a mass element. In other words, ELL does not attempt to describe the collapse of an *extended* ellipsoidal peak, rather, it operates on the infinitesimal level. Given this, OC appears as the most sensible choice for the collapse condition. OC corresponds to collapse along the first axis, which means that the ellipsoid has undergone pancake collapse. However, this does not imply that the extended region is flattened as well. Indeed, as the example in Monaco (1998) illustrates, in the collapse of a spherical peak with decreasing density profile, all mass elements (except for the one in the centre) collapse as needles pointing to the centre. This is because the spherical symmetry guarantees that the first and second axis collapse together. Yet the collapse of the peak is *not* that of a filament but of a sphere. This shows how misleading the local geometry of collapse is for understanding the global geometrical properties of the collapsing matter.

Alternatively, ellipsoidal collapse can be used to model extended regions associated to a particular set of points, such as density peaks (Bond & Myers 1996; Sheth, Mo & Tormen 2001). In this case, first-axis collapse truly corresponds to the formation of a flattened structure, while third-axis collapse corresponds to the formation of a spheroidal object. For instance, in the case of the spherical peak mentioned above, the peak point is collapsing in a spherical way both locally and globally. It is clear that in such cases a satisfactory definition of collapse must be related to third-axis collapse. Sheth et al. (2001) showed that indeed this collapse condition improves the agreement with simulations when the centres of mass of FOF objects are considered (a particular set of points analogous to the peaks), but does not help much when general unconstrained points are considered.

The two definitions of collapse are very different from many points of view. First-axis collapse is on average faster than the spherical one (Bertschinger & Jain 1994), while third-axis collapse is correspondingly slower. Moreover, while 50 per cent of mass is predicted to collapse at very late times by linear theory (starting from a density field with finite variance and not taking into account the cloud-in-cloud problem), $23/25 \sim 92$ per cent of mass is predicted to undergo first-axis collapse, but only 8 per cent third-axis collapse. This is very important when computing the mass function with a PS-like approach: while first-axis collapse more or less reproduces the correct normalisation (Monaco 1997b), third-axis collapse requires a large 'fudge factor' $\sim$12 (Lee & Shandarin 1998), as only 8 per cent of mass is available for collapse.

Within the framework of the excursion set approach, it is interesting to understand whether the introduction of ellipsoidal collapse is going to improve the statistical agreement between simulations and PS. Monaco (1997b) and Sheth et al. (2001) showed that ellipsoidal collapse can be introduced through a 'moving' barrier which depends on the variance σ^2 of the smoothed field. Third-axis collapse gives longer collapse times than spherical collapse, and this corresponds to a barrier which rises with σ^2, while the opposite is true for first-axis collapse. In the case of *sharp k-space smoothing*, the fixed barrier reproduces the PS mass function and hence overestimates the number of low mass objects. Sheth et al. (2001) showed that using the moving barrier appropriate for third-axis collapse leads to the formation of fewer low mass objects, and hence improves the mass function. However, when *Gaussian smoothing* is used, the fixed-barrier solution is different from PS, and the number of small mass haloes is now severely *underestimated.* Monaco (1997b, 1998b) showed that in this case first-axis collapse (with no free parameter to tune) produces a reasonable fit to the simulations, with some improvement with respect to PS.

From these considerations, it is clear that a successful definition of collapse depends on many technical details, such as the kind of dynamics considered (mass elements versus extended regions) and the type of smoothing used (sharp k-space versus Gaussian smoothing).

Chapter 3

Structure formation: a statistical approach

I never wanted to write this words down for you
With the pages of phrases of things we'll never do.

C. Cornell

A detailed understanding of the formation processes of DM haloes is a crucial point in the study of galaxies and galaxy clusters: the way haloes merge and accrete mass determines many properties of the visible matter. Unfortunately, those processes represent the highly non-linear regime of the dynamical evolution and no exact analytic calculation is possible. It is then 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.

We have discussed the PS model (§ 2.6.2) and its extended derivation (§2.6.4, EPS), the EPS formalism can be also used to predict some average statistical properties of DM haloes such as the progenitor mass function, the formation times, survival times and merger rate (Lacey & Cole 1993). These predictions were first tested against numerical simulations, with success, by Lacey & Cole (1994; see also Sheth & Lemson 1999; Somerville et al. 2000)

However, those statistical quantities are not enough if our final goal is to model how structures form and evolve: we need to trace the history of individual objects. The knowledge of the *merging history* of DM haloes goes (fortunately not so far) beyond a simple analytic derivation. Once again the versatile EPS formalism help us to provide the theoretical backbone for semi-analytic Monte Carlo realizations of the merging history of a DM halo. The EPS formalism then becomes a standard tool to construct synthetic catalogs of DM haloes for galaxy formation programs (see, e.g., Kauffmann, White & Guiderdoni 1993; Somerville & Primack 1999; Cole et al. 2000).

As other authors before us, we venture upon the realization of a EPS based semi-analytic code to generate merging histories of haloes. This Chapter is devoted to the description of the theoretical framework and to the practical difficulties we tackled to create the code. First we briefly review the EPS formalism (§3.1) and provide the expression of the conditional mass function. Then we present the Monte Carlo methods to generate synthetic catalogs of haloes (§3.3). In §3.3.3 we introduce our recipe for the EPS based random realization of the merging histories (or *merger trees*) of DM haloes. §3.4 shows the tests of our semi–analytical code against N-body simulations.

3.1 Extended PS formalism

The EPS formalism recasts the structure formation problem in terms of trajectories, this allows to evaluate directly the *conditional mass function*, i.e. the number density of haloes at redshift z with mass M, bound to flow into a halo of given mass $M_0 > M$ at a subsequent time z_0.

Notation: In the following, the 'final' haloes at z_0 will be called *parent*, while the higher-redshift haloes that flow into the parent will be called *progenitors*.

Together with the progenitor distribution the same formalism can be manipulated to obtain other statistical average quantities (Lacey & Cole 1993):

- the *merger rate*;
- the distribution of *formation times* of haloes at different redshifts, defined as the time when the largest of its progenitors first has mass $M_0/2$;
- the distribution of *survival times*.

The analytical derivation of all these quantities goes beyond the aim of this thesis. We will just concentrate on the progenitor mass function which is the keystone of the EPS based Merger Trees.

3.1.1 The progenitor mass function

Using the excursion set approach (§2.6.4) we evaluate the fraction of trajectories in progenitor haloes of mass M at z that are in a parent halo of mass M_0 at later time z_0. The derivation of the conditional mass function is similar to that of the cumulative distribution.

Notation: Using the same notation of Lacey & Cole (1993) we introduce the variable $S(M) \equiv \sigma^2(M)$.

In this case the trajectory does originate from the point $[S(M_0),\ \delta(z_0)]$ and with a simple variable substitution eq. (2.62) becomes:

$$\begin{aligned} f(M,\delta|M_0,\delta_0)\mathrm{d}M &= \frac{1}{\sqrt{2\pi}}\frac{(\delta-\delta_0)}{[S(M)-S(M_0)]^{3/2}} \\ &\times \exp\left\{-\frac{(\delta-\delta_0)^2}{2[S(M)-S(M_0)]}\right\}\mathrm{d}\,S\,, \end{aligned} \tag{3.1}$$

where $\delta_0 = \delta(z_0)$ (Lacey & Cole 1993). With some algebra it is possible to evaluate the conditional mass function, i.e. the number density of haloes bound to flow into a parent halo of given mass at a subsequent time:

$$\frac{\mathrm{d}\,N}{\mathrm{d}\,M}(M,z|M_0,z_0)\mathrm{d}M = \left(\frac{M_0}{M}\right)^2 f(M,\delta|M_0,\delta_0)\,\mathrm{d}M\,, \tag{3.2}$$

here the fraction M/M_0 converts the counting from mass weighting to number weighting.

3.1.2 Merging histories

The knowledge of the progenitor mass function is the first step in the determination of a semi-analytic procedure to generate merging histories of haloes. Then, we have to decide which kind of "history procedure" we want to follow. In fact, we could work backwards in time starting from the parent and braking it into progenitors (*disintegration procedure*) or we could assembly the halo starting from its progenitors. Both methods are equally applicable, but the standard choice is to use the "disintegration" approach, which does not require any particular modification of the EPS formalism.

In the semi-analytic code we pick up progenitors from the number weighted probability function numerically. Form eq. (3.2) we notice that the number of haloes diverges as mass goes to zero, so it is necessary to introduce a threshold mass $M_{\rm th}$ which acts as a resolution variable.

Before giving any interpretation to $M_{\rm th}$ we notice that, when we introduce a threshold mass, the total number of progenitors at some redshift z is a finite number which strongly depends on the mass resolution. For a parent M_0 identified at z_0 the number of progenitors at redshift z is:

$$\bar{N}(z, M_{\rm th}) = \int_{M_{\rm th}}^{M_0} \frac{M_0}{M} \frac{{\rm d}N}{{\rm d}M}(M, z|M_0, z_0){\rm d}M \; . \tag{3.3}$$

Let us simplify the complex dynamics that affect the life of a halo in accretion and merging events. Any mass smaller than the mass resolution is treated as diffuse mass which accretes on a halo. As a consequence the progenitors must be more massive than $M_{\rm th}$.

Comment: The life of a halo is followed backward in time, so it is more appropriate to identify a de-accretion and a splitting event. This is the origin of the concept of Merger Tree: a splitting event corresponds to the identification of two trajectories one for each halo. It is in the common use to identify the trajectory that corresponds to the more massive halo as "main branch". The disintegration of a halo generates a tree of trajectories

An important quantity is the fraction of accreted mass onto the parent. If we define the average mass fraction of the parent contained in the progenitors at z

$$\bar{f}_{\rm p} = \int_{M_{\rm th}}^{\infty} \frac{{\rm d}N}{{\rm d}M}(M, z|M_0, z_0){\rm d}M \;, \tag{3.4}$$

the fraction of accreted mass is

$$\bar{f}_{\rm acc} = 1 - \bar{f}_{\rm p} \;. \tag{3.5}$$

Unfortunately, a crucial missing quantity is the accretion probability, which models the amount of mass accreted by a halo in an interval of time. In this case the EPS formalism is not able to provide any valid expression. Even if some reasonable guesses for its functional form were proposed by Somerville and Kolatt (1999) and Sheth & Lemson (1999), they are all based on strong hypotheses on the definition of the accretion process.

3.2 The Cosmological Model

In the EPS model the progenitor mass function is determined by the shape of the absorbing barrier and by the relation between S and M. Different choices of the cosmological model influence both the rate the barrier moves and the function $S(M)$.

The Cosmological model we have chosen is a ΛCDM with $\Omega = 0.3$, $\Omega_\Lambda = 0.7$ and H_0=50 km/s/Mpc. The primordial spectral index is $n = 1$ and $\sigma_8 = 1$. The power spectrum is calculated with the Bardeen et al. (1986) formula. We also tested our results with other models and they are equally valid in all cases.

3.3 The Monte Carlo procedure

We can now introduce in detail the semi-analytic procedure to generate merger histories of haloes. The first attempt to use a Monte Carlo approach is due to Lacey & Cole (1993). They developed a simple but efficient numerical recipe, easy to implement, which is the "progenitor" of any other more refined procedure.

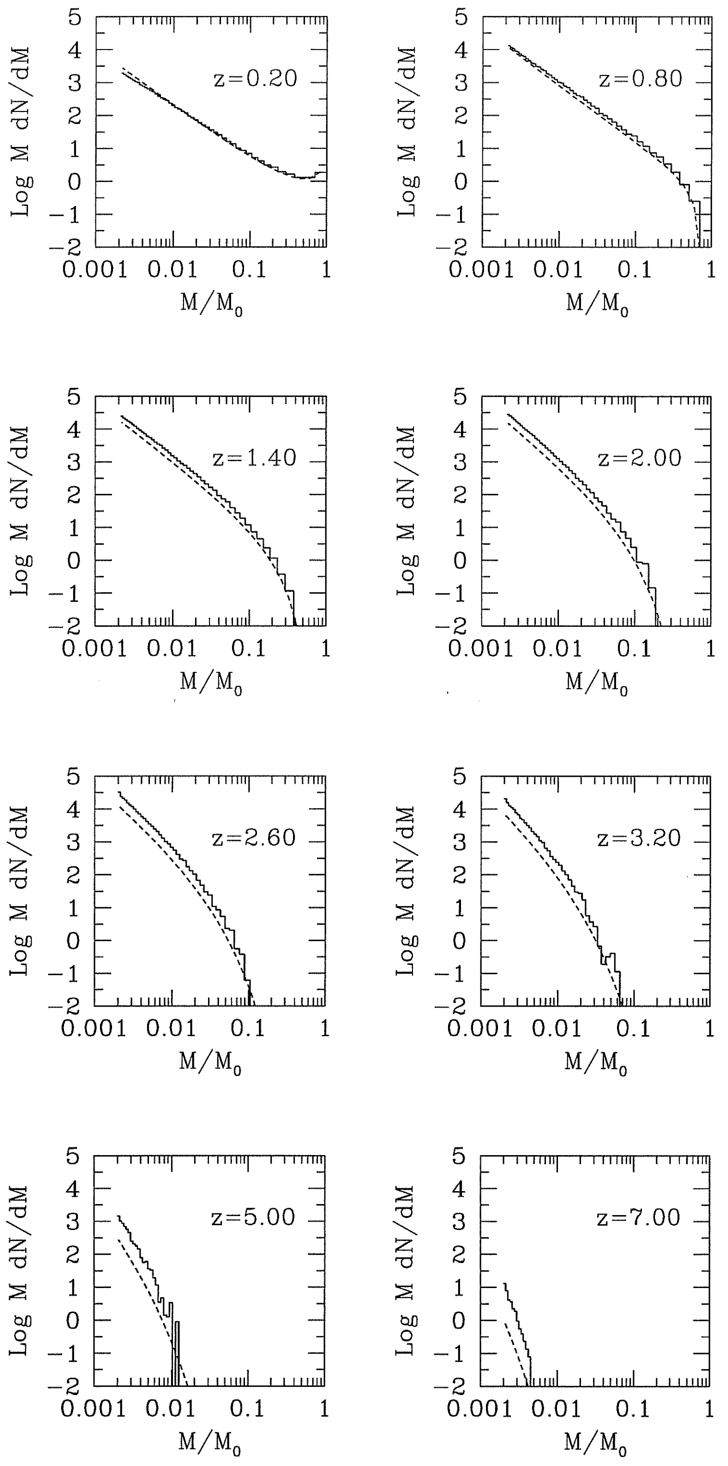

Figure 3.1: Conditional mass function for a parent of mass $M_0 = 500\ M_{\rm th}$ calculated with 1000 realizations of the original LC recipe. The dashed lines are the EPS predictions while the histograms are the Monte Carlo results. Each plot is labelled with the final redshift. The mass function calculated in this way develops an excess of haloes starting from redshift 0.5 when compared with the EPS prediction. This excess increases with z and becomes larger that an order of magnitude at $z = 7$.

3.3.1 The Lacey & Cole original recipe

Let us consider the distribution function $f(M, z|M_0, z_0)$ given by eq.(3.1). It represents the probability that a new branch of mass $S(M)$ starts at redshift $\delta(z)$ from an initial trunk $S(M_0)$.

Warning: This conditional probability is a single trajectory probability and cannot be used to simultaneously extract the whole family of progenitors. To do this we need the joint probability function for a set of progenitors which cannot be analytically evaluated (Somerville & Kolatt 1999). We can use eq.(3.1) to extract only one progenitor.

Eq. (3.1) can be rewritten to evaluate the probability for a mass step ΔS in a "small" interval of time $\Delta\delta$:

$$K(\Delta\delta, \Delta S)\ \Delta\delta = \frac{\Delta\delta}{\sqrt{2\pi\Delta S}}\ \exp\left[-\frac{(\Delta\delta)^2}{2\Delta S}\right]\ \Delta\delta\ . \tag{3.6}$$

If the time step is "sufficiently" small the previous equation represents the probability that the halo $M(S)$ splits in a halo $M(S+\Delta S)$ and one of mass $M(S) - M(S+\Delta S)$.

Comment: We do not allow any process that involve more than a binary splitting in the time interval $\Delta\delta$: the two progenitors limit. This condition can be assured modeling the time step. Somerville & Kolatt (1999) suggest that a good choice for $\Delta\delta$ is:

$$\Delta\delta \sim \sqrt{\frac{dS}{dM}\Delta M_c} \tag{3.7}$$

with $\Delta M_c \ll M_0$.

Starting from the main trunk $S(M_0)$ at z_0 we pick up a new mass $M(S+\Delta S)$ at a time $\delta(z_0) - \Delta\delta$ (consequently we find a second progenitor $M^{2p} = M(S) - M(S+\Delta S)$). The time increment is fixed while the the mass (de-)increment is extracted from Gaussian distribution with zero mean and unit variance as suggested by eq.(3.6) applying the change of variable $x \equiv \Delta\delta/(2\sqrt{\Delta S})$. We fix a mass threshold $M_{\rm th}$ which is the minimum mass of the progenitor at redshift z_{max}. If both the progenitors have mass greater than $M_{\rm th}$ a new branch is created with mass M^{2p}. On the contrary if the mass of one of the two progenitors is below the threshold, the parent halo (de-)accreted mass in the time interval $\Delta\delta$. We then work backward in time following all the branches created, until z_{max} and we identify all the progenitors of the primary halo at that time.

With this simple procedure we reconstruct the mass function of the progenitors which should be in perfect agreement with the EPS expression. Unfortunately, as other authors have already noticed (Somerville & Kolatt 1999, Sheth & Lemson 1999), the method we have just explained over-predicts the number of haloes and gives results which are not at all comparable with the EPS prediction. In Fig. (3.1) we show the conditional mass function for a halo of mass $M_0 = 500 M_{\rm th}$ at different redshifts, the lack of agreement with the analytical prediction is evident especially at high redshift.

3.3.2 The second progenitor problem

The reason way the LC original recipe does not reproduce the EPS analytical predictions is connected with the way LC identify the two progenitors. One progenitor is chosen according to the distribution function, while the second is assumed to be the residual mass M^{2p}. This procedure is obviously violating a fundamental rule: both progenitors must be chosen from the same distribution function. The second progenitor cannot be M^{2p} but it would be contained inside the residual mass.

There are different ways to overcome this problem. Somerville & Kolatt (1999) and Sheth & Lemson (1999) propose two similar algorithms to account for the accretion on the residual mass. They assume that M^{2p} is a reservoir of mass fragments which could contain or not other progenitors. The two methods differs in the way they re-extract the new haloes according to the proper distribution probability. LC propose a more refined approach (Cole et al. 2000) based on a modified EPS Monte Carlo algorithm, which explicitly accounts for the accreted mass and does not quantise the progenitor halo mass: they Monte Carlo generate the probability of splitting and evaluate the fraction of mass in fragments smaller then $M_{\rm th}$ which contributes to the accretion. The authors test their models with numerical N-body simulations in order to check the ability of their Merger Tree recipe to reproduce the numerical experiments (Somerville et al. 2000, Sheth & Lemson 1999, Lacey & Cole in prep).

3.3.3 The MCMT recipe

We present now a different EPS based procedure to generate merger trees. The main rules of this Monte Carlo recipe are similar to that of the LC one: binary splitting, progenitors extracted from the mass-weighted distribution and if we extract a progenitor smaller than the threshold mass we count it as accreted mass. The main difference between this Monte Carlo merger tree (hereafter MCMT) and the LC one is that we randomly extract the time interval instead of the mass increment.

Let us then consider a fixed mass increment ΔS. We can use eq. (3.6) to pick up a time variation $\Delta\delta$. If the time step is positive then we have a first upcrossing event, or better a binary splitting event. As for the LC procedure, we have two progenitors M and M^{2p}. We notice that also in this case only one of them is extracted from the distribution. If both the progenitors are greater then the mass threshold we have a merging event, otherwise we have an accretion event. The code works backward in time as described in the previous Section and it identifies all the progenitors at $z_{\rm max}$.

A small mass increment assures the two progenitor limit. Our choice is to use ΔS equal to a small fraction of the mass threshold. We tested the code varying the mass increment and we noticed that if we reduce too much ΔS the code becomes extremely slow, but its accuracy does not sensibly increase. We believe that a good compromise between the code speed and its accuracy is to assume $\Delta S = S(M_{\rm th}/10)$.

The first and obvious test is to compare the conditional mass function obtained from the MCMT with the EPS analytical predictions. In Fig. (3.2) we present the result of this comparison. As for the LC recipe, the MCMT develops an excess of

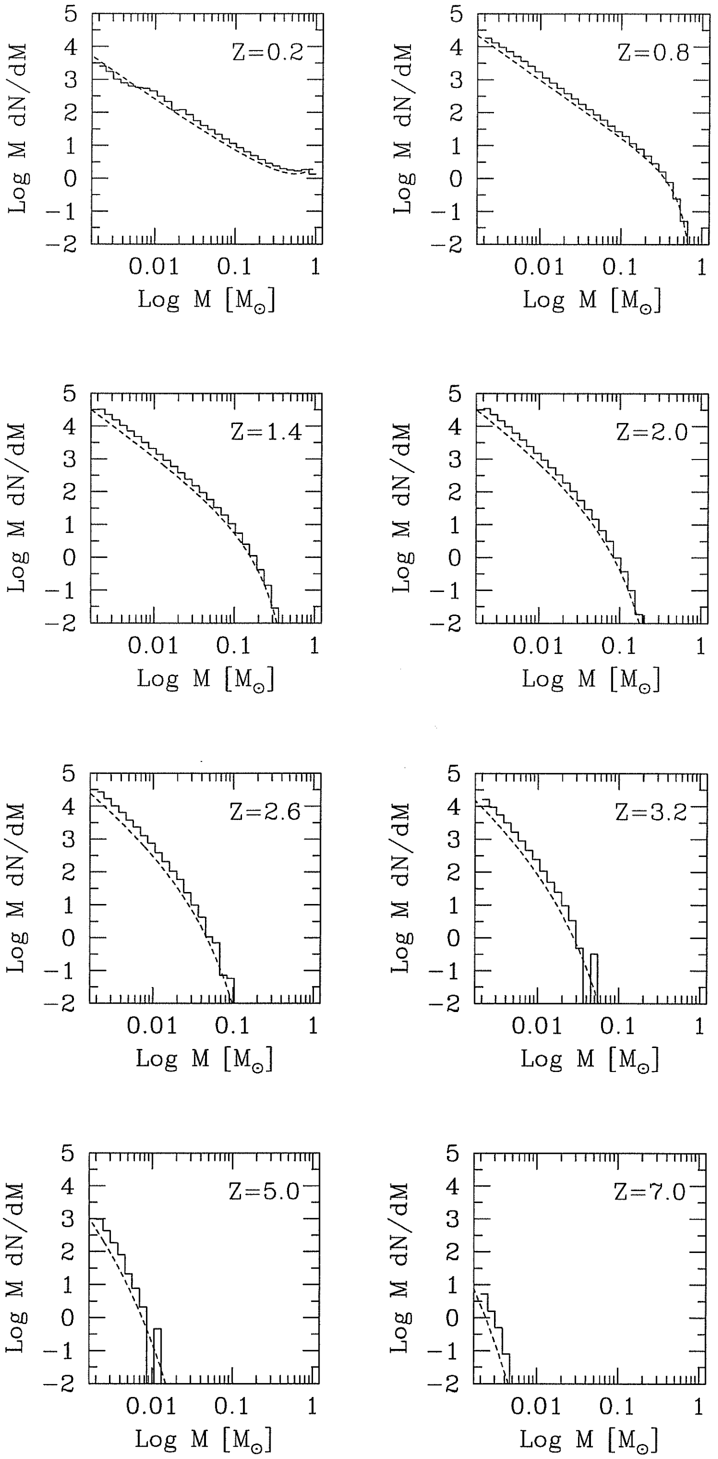

Figure 3.2: Conditional mass function for a parent of mass $M_0 = 500\ M_{\rm th}$ calculated with 1000 realizations of the MCMT without accretion. The dashed lines are the EPS predictions while the histograms are the Monte Carlo results. Each plot is labelled with the final redshift. The mass function calculated in this way developed an excess of haloes starting from redshift 0.5 when compared with the EPS prediction.

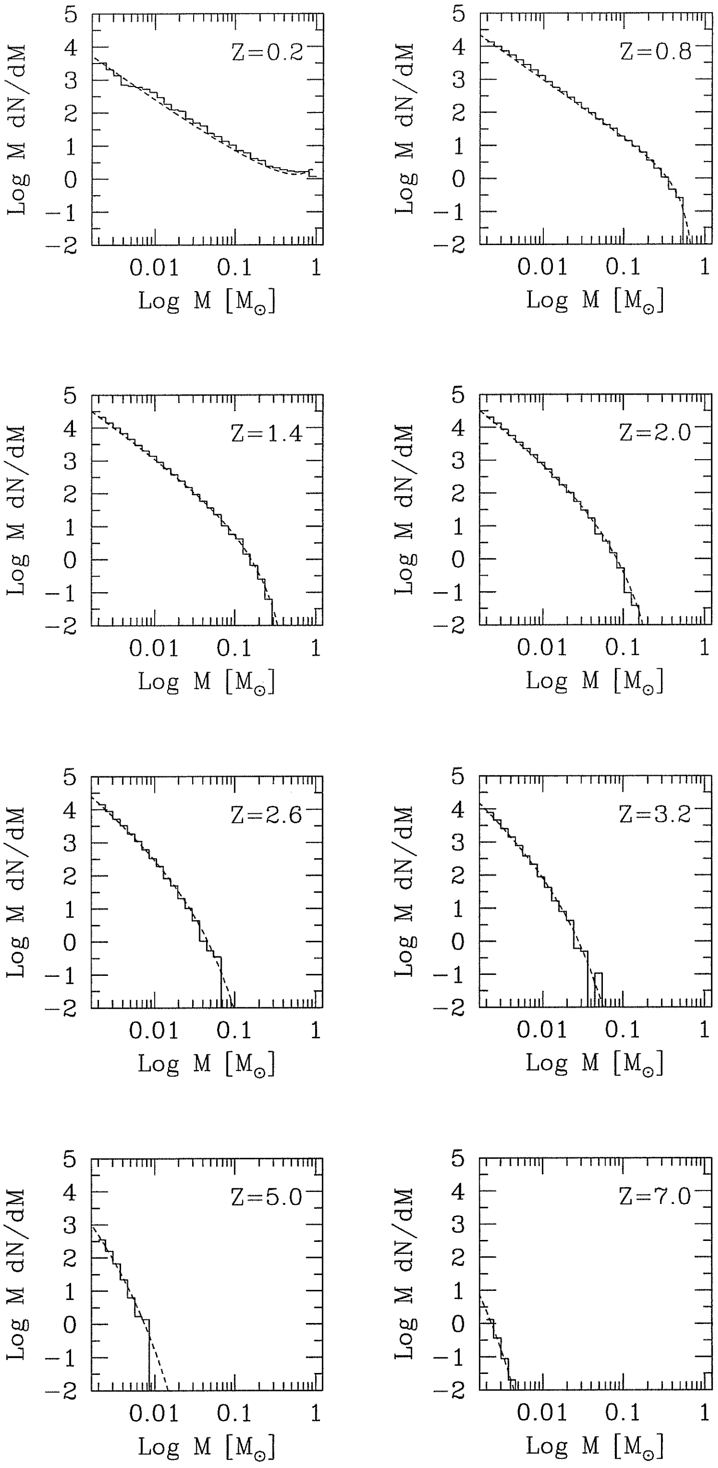

Figure 3.3: Conditional mass function for a parent of mass $M_0 = 500\ M_{\rm th}$ calculated with 1000 realizations of the MCMT with accretion. The dashed lines are the EPS predictions while the histograms are the Monte Carlo results. Each plot is labelled with the final redshifts. The accretion parameter is set to $\xi_a = 0.18$. We notice that up to high redshift MCMT reproduces with great accuracy the EPS prediction.

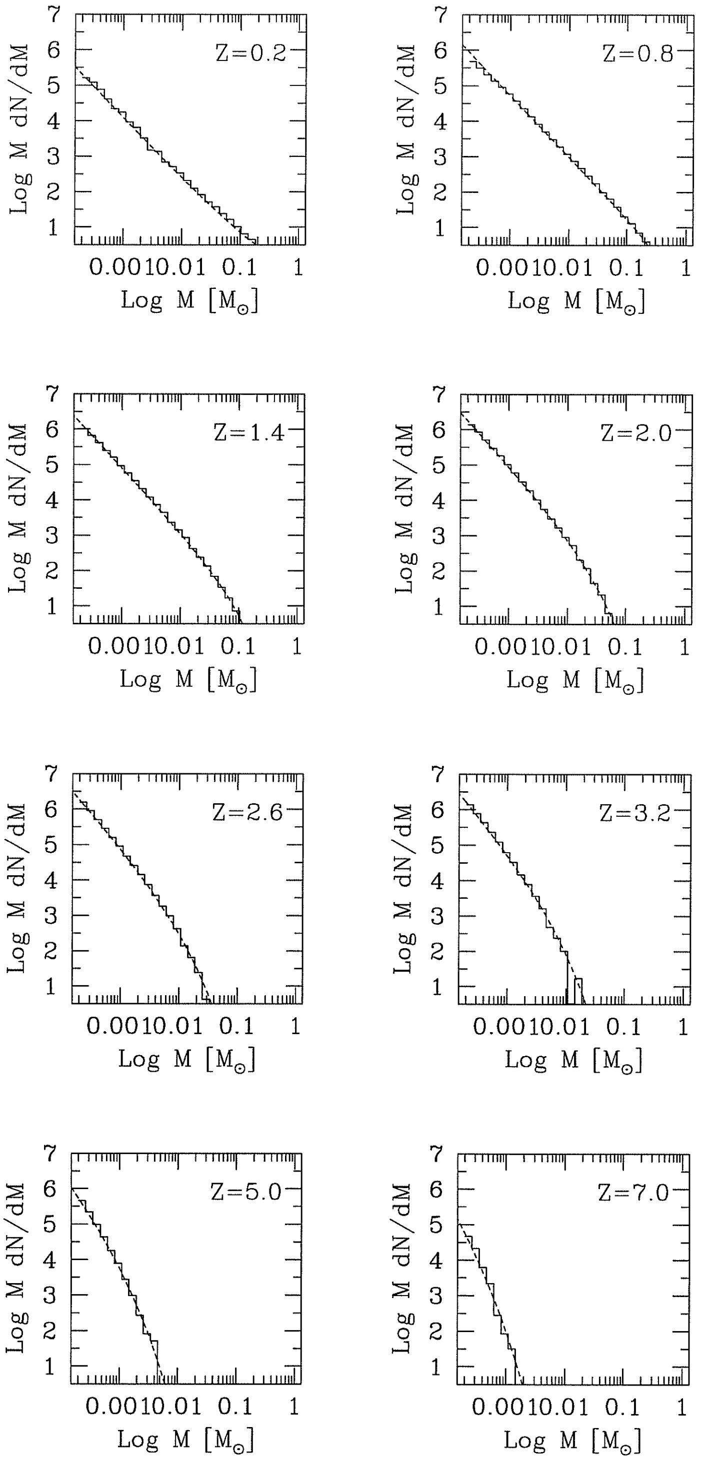

Figure 3.4: Conditional mass function for a parent of mass $M_0 = 5000\ M_{\rm th}$ calculated with 1000 realizations of the MCMT with accretion (accretion parameter $\xi_a = 0.18$). The dashed lines are the EPS predictions while the histograms are the Monte Carlo results. Each plot is labelled with the final redshift.

haloes compared to the EPS prediction, but in this case the discrepancy is smaller and more constant in redshift. We believe that this behaviour is due to the way we extract the progenitors. Choosing the mass increment is a more certain way to assure the binary limit, partially curing the discrepancy present in the LC method. But we still miss a fundamental ingredient: the accreted mass on the second progenitor. In fact, M^{2p} is the residual mass that could or not contain the second progenitor.

We start from the hypothesis that the non-extracted progenitor is a fraction of the residual mass. Then we introduce an *accretion parameter* which reduces M^{2p} of a fixed fraction $\xi_{\rm a}$. We select then two progenitors M and $M^{2p} = M^{2p}(1 - \xi_{\rm a})$ and some diffuse mass $\xi_{\rm a} M^{2p}$ which accretes on the halo in the time interval $\Delta\delta$.

To find the proper value of $\xi_{\rm a}$ we must compare the MCMT conditional mass function with the EPS prediction. In principle we cannot be sure that the same accretion parameter works for different parent masses and different redshifts. Then we perform different runs of MCMT with different value of $\xi_{\rm a}$ and we find that our best choice for the accretion parameter $\xi_{\rm a} = 0.18$ is valid for all the tested configurations.

To give a taste of the ability of MCMT, we present the comparison of our merger tree with *fixed accretion* with the EPS predictions for a parent halo of $M_0 = 500\, M_{\rm th}$ and $M_0 = 5000\, M_{\rm th}$. The results are shown in Fig. (3.3) and Fig. (3.4). The discrepancy between the two curves is below the 10 per cent.

We also check the average fraction of mass of the primary halo contained in the progenitors. This quantity is a direct measure of the ability of MCMT in describing the accretion events, and it is a more robust test of the fixed accretion approximation. We compare the analytical prediction with the MCMT results averaged over 1000 realizations, the results are shown in Fig. (3.5). MCMT seems to reproduce accurately also this quantity.

Finally we test the code for different cosmological models, always with similar results.

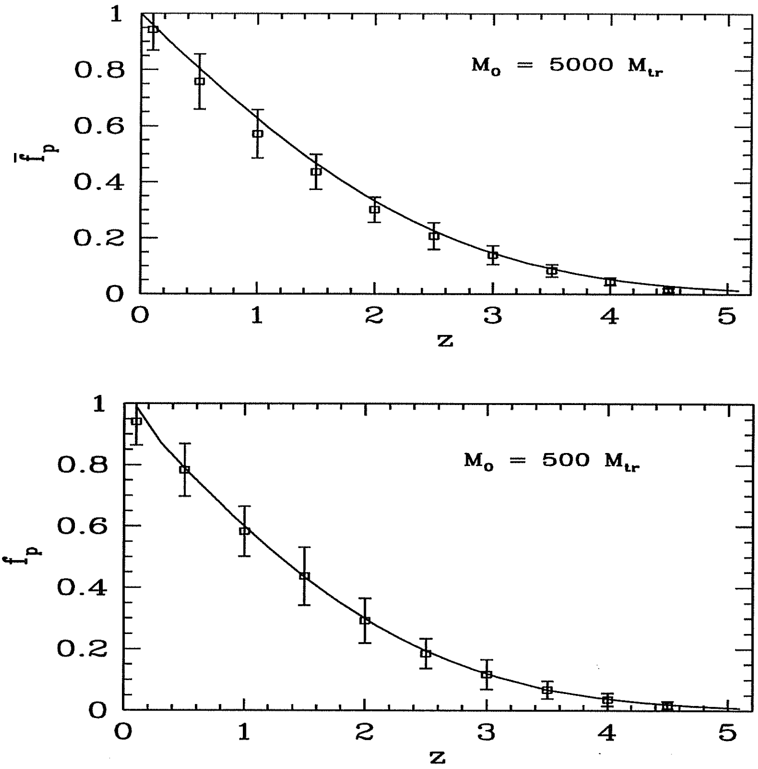

Figure 3.5: The average mass fraction of the primary halo contained in progenitors with mass $M > M_{\rm th}$ as a function of redshift. The solid lines are the EPS predictions while the points with error bars are the MCMT results.

3.4 Testing the MCMT

Since a merger tree code should be used as a valid alternative to the N-body simulations, it is useful to investigate whether the distributions predicted by MCMT are in agreement with that measured in the numerical experiments.

Here, we present a comparison between N-body simulations and MCMT: we study the ability of our code in predicting the statistics of the merging histories, with a special attention on the distribution of the progenitors mass and number.

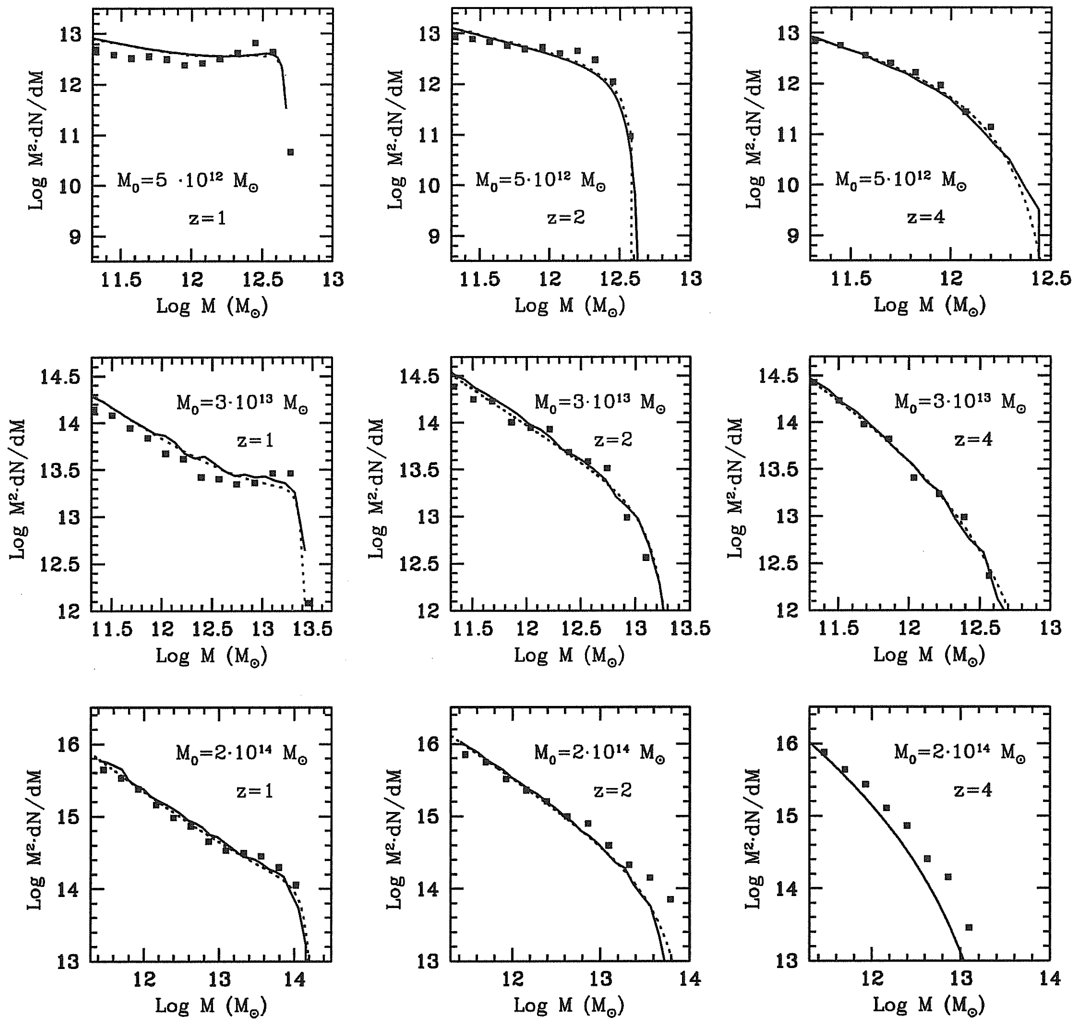

Figure 3.6: Conditional mass functions for parent haloes identified at $z = 0$. The mass threshold is fixed at $M_{\rm th} = 7.6 \times 10^{10}$ $\mathrm{M}_{\odot}$, the redshift increases from left to right and covers the values: $z = 1,\ 2,\ 4$. The mass of the parent halo increases from top to bottom, the adopted values are: $M_0 = 5. \times 10^{12} M_{\odot}$, $3. \times 10^{13} M_{\odot}$ and $2.0 \times 10^{14} M_{\odot}$. The points represent the simulation data while the solid lines are the prediction of MCMT; the dashed lines are the analytical predictions of the EPS formalism.

3.4.1 The N-body simulations

We use a ΛCDM simulation run with the Hydra code (Couchman, Thomas & Pearce 1993) on a box of 100 h^{-1} Mpc with 256^3 particles. The simulation was done by Tom Theuns in cooperation with Silicon Graphics/Cray Research utilising the Origin 2000 super computer at DAMTP, Cambridge. The cosmological parameters are $\Omega_0 = 0.3$, $\Omega_\Lambda = 0.7$ $h = 0.65$ and $\sigma_8 = 0.9$ and we set $M_{\rm th} = 7.6 \times 10^{10}\ M_\odot$ that corresponds to 30 particles.

The haloes are identified using a standard friend-of-friend (FOF) algorithm with linking length 0.2 times the inter-particle distance. Note that, following the suggestion by Jenkins et al. (2001), we do not change linking length with the cosmology. We adopt 30 particles as the minimum mass of the haloes when we analyse the conditional mass function.

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 makes the progenitor identification process more ambiguous. We then adopt two simple rules:

1. 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 percent of cases);

2. 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.

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

3.4.2 The comparison

The conditional mass function

In Fig. (3.6) we compare the conditional mass functions obtained from the MCMT and the simulation, we also plot the analytical prediction from the EPS formalism, eq. (3.2). The bottom panels of Fig. (3.6) show the results for a cluster-sized parent of $M_0 = 2 \times 10^{14}$ $\mathrm{M}_\odot$, the case of haloes corresponding to small groups ($M_0 = 3 \times 10^{13}$ $\mathrm{M}_\odot$) and galaxies ($M_0 = 5 \times 10^{12}$ $\mathrm{M}_\odot$) are presented in the mid and upper panels. 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 MCMT (the solid lines in the plots) shows a very good agreement when compared with the EPS analytical result. On the other hand, the figure shows a discrepancy between the N-body results and the EPS predictions already pointed out by other authors for the mass function of haloes (Gelb & Bertschinger 1994; Governato et al. 1999; Somerville et al. 2000, Jenkins et al. 2001; Bode et al. 2001): the EPS predictions 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.

The distribution of the largest progenitor

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 before. The histograms on Fig. (3.7) show the distribution of the mass of the larger progenitor normalised to the parent mass, M_1/M_0, predicted by MCMT. The symbols connected with lines denote the corresponding simulation results. For this particular statistic the agreement between the numerical experiment and MCMT 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 progenitor first becomes greater than

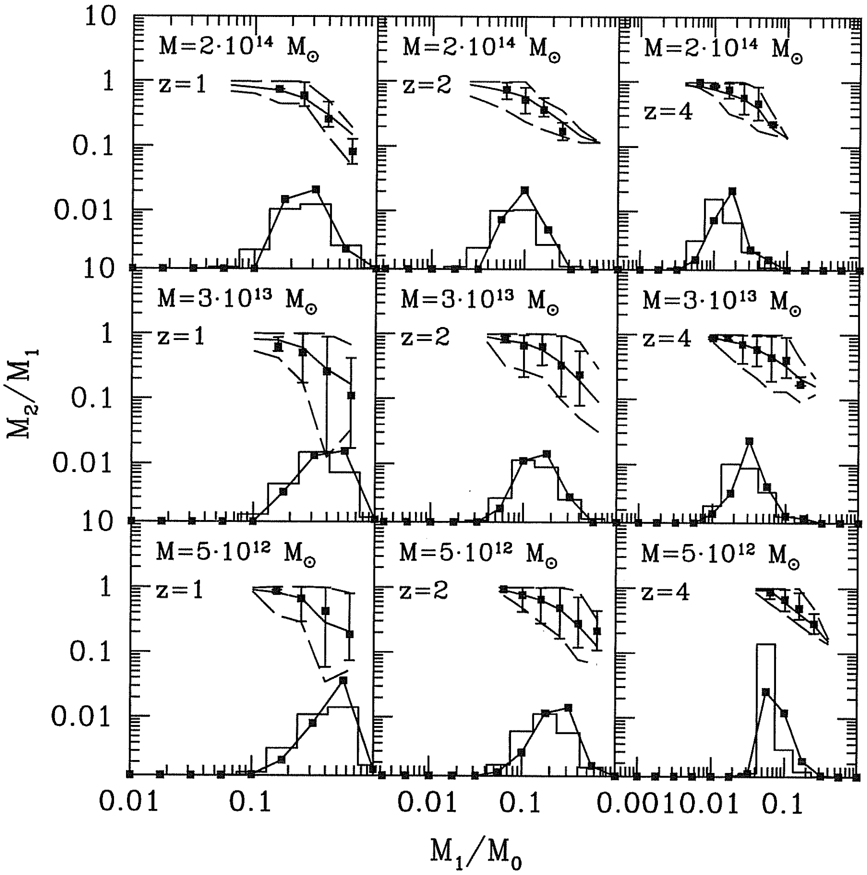

Figure 3.7: The distribution of the mass of the largest progenitor M_1 for the ΛCDM case with mass threshold $M_{\rm th} = 2.3 \times 10^{11} M_\odot$ (30 particles). The histograms are the MCMT predictions and the points connected with solid lines are the simulations. The quantity plotted on the upper 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 is the PINOCCHIO result and the dashed lines show its 1σ variance. The points with error bar are the simulation data.

$M_0/2$. 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. Then MCMT seems to reproduce with good accuracy also the halo formation times. 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 Fig. (3.7), we show the distribution of M_2/M_1 (the ratio of the second largest progenitor and largest ones) vs M_1/M_0. The points are the mean value of the distribution and the error bars are the the corresponding 1σ variance, both measured in the simulations. The solid lines and the dashed lines are the same quantities predicted by MCMT. The agreement is good.

Progenitors in numbers

Finally we analyse the statistical properties of the distribution of the number of progenitors of a halo of mass M_0.

In Fig. (3.8) 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 normalisation) the integral of the conditional mass 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 MCMT for different parent masses and redshifts. The filled symbols connected with lines are the distribution extracted from the simulations. We note that MCMT overestimates the distributions for all redshifts and masses.

The ability of MCMT in predicting the distribution of the number of progenitors can be quantified by comparing the first and second moments, μ_1 and μ_2, measured in the simulations with their values predicted by MCMT. In Fig. (3.9) 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 MCMT prediction 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 eq. (3.2).

Warning: For arbitrary initial conditions the EPS formalism cannot analytically evaluate the higher moments of the distribution.

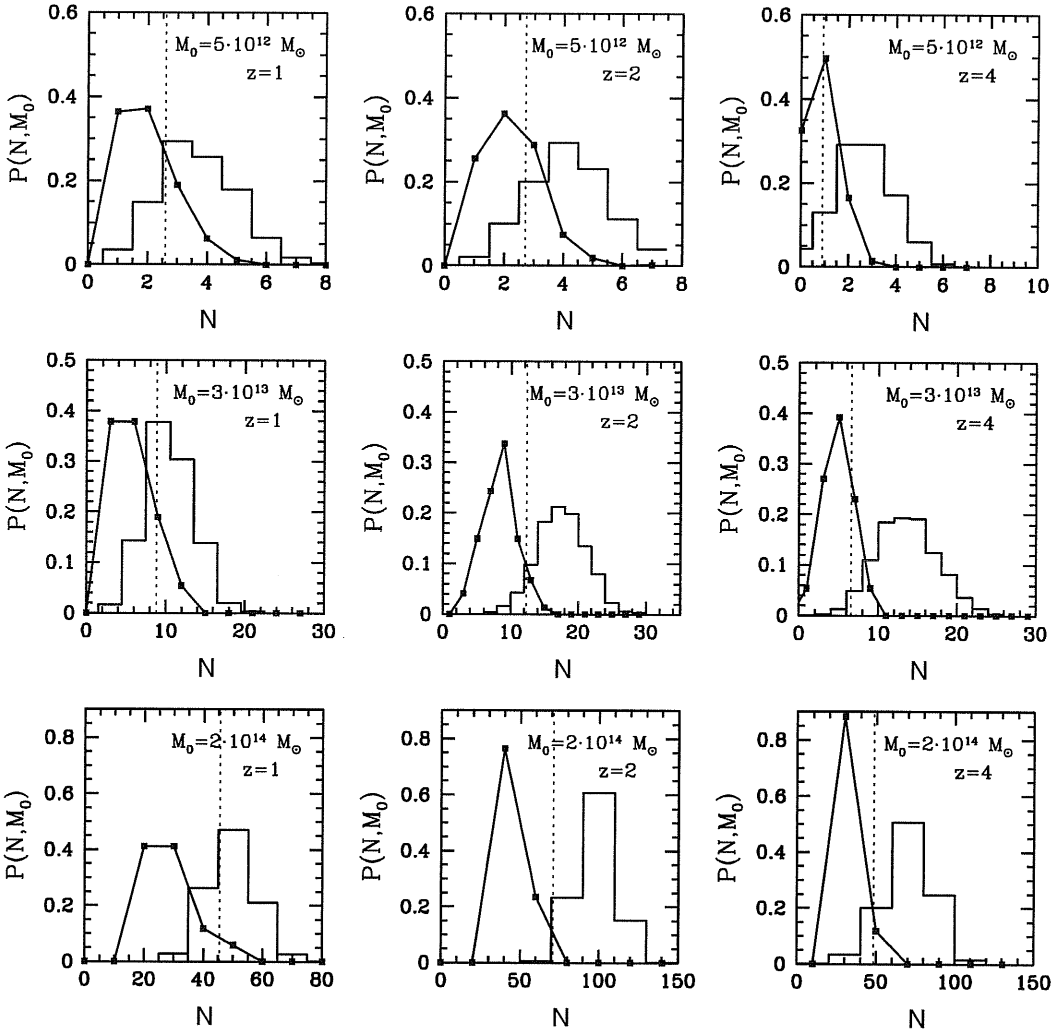

Figure 3.8: Probability that an halo M_0 at $z = 0$ has N progenitors for the ΛCDM case. The threshold mass is $M_{\rm th} = 2.3 \times 10^{11} M_\odot$ (30 particles). The points connected with solid lines represent the simulation data while the histograms are the prediction of MCMT. The vertical dotted lines are the average values analytically calculated from EPS formalism.

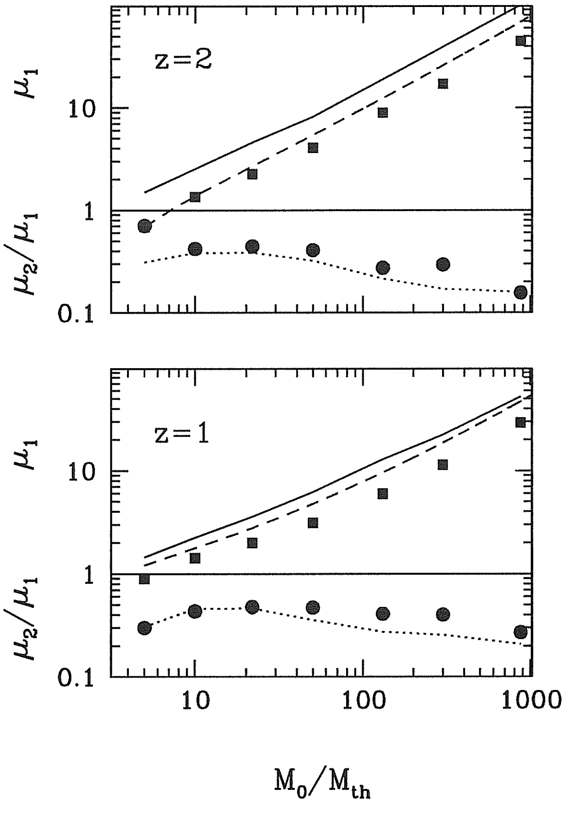

Figure 3.9: The first two moments of the distribution of the number of progenitors $P(N, M_0)$ as a function of the parent mass M_0 at redshift 1 and 2. Lines are the MCMT results and symbols are the simulation data. The threshold mass is $M_{th} = 2.3 \times 10^{11} M_\odot$ and we plot the mean (squares 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.

3.5 Conclusions

We have presented a semi–analytical procedure to generate merger histories of virialized haloes. This procedure, based on the EPS formalism, can be used to build up a complete description of the formation path of a given halo and, in principle, it can represent a valid alternative to the use of N-body simulations.

We can summarise the main features of the MCMT algorithm as follows:

1. A parent forms from a cascade of binary merging events;
2. We introduce a resolution mass $M_{\rm th}$: a mass fragment smaller that $M_{\rm th}$ is treated as accreted mass.
3. During a time step $\Delta\delta$ we extract one progenitor from the mass-weighted distribution function and we evaluate the residual mass. A fixed fraction of the residual mass is treated as diffuse mass that accretes onto the halo.

Warning The accreted mass is not necessarily smaller then the threshold one. During a merging event an arbitrary number of mass fragments with mass $M < M_{\rm th}$ can accrete onto the halo.

This algorithm represents an improvement with respect to the LC original recipe which presents a discrepancy with respect to the analytical EPS predictions. On the contrary, the MCMT code reproduces with good accuracy the EPS conditional mass function at all redshifts.

We have also tested the statistical predictions of the MCMT against a set of large high-resolution N-body simulations for a ΛCDM cosmology. The comparison reveals an evident discrepancy between the semi-analytical and the numerical results. The MCMT develops an excess of low mass progenitors and a defect of high mass ones, which affects not only the conditional mass function, but also the other statistics like the distribution of the number of progenitors.

This behaviour is the crucial weakness of any EPS-based merger tree (see also Somerville et al. 2000), it is, in fact, intrinsic in the EPS formalism itself. We could cure this weakness by introducing the ellipsoidal collapse as suggested by Sheth & Tormen (2002).

The corrected EPS formalism reproduces with good accuracy the N-body data, but still lacks the information regarding the position, the spin and the internal structure of haloes. We develop a different approach to follow the formation and evolution of DM haloes in hierarchical cosmological models which combines the advantages of the EPS semi-analytical procedure with a refined description of the halo dynamics. The next Chapters will be devoted to the description of this procedure.

Chapter 4

PINOCCHIO: the algorithm

Once upon a time there was a piece of wood.
It was not an expensive piece of wood. Far from it.
Just a common block of firewood, one of those thick, solid logs
that are put on the fire in winter to make
cold rooms cozy and warm.

C. Collodi

We present a new algorithm to compute the formation and evolution of DM haloes in a given linear density field. Since this method describes the hierarchical build-up of objects that have undergone OC, we refer to it as PINOCCHIO: PINpointing Orbit-Crossed Collapsed HIerachical Objects (Monaco et al. 2002; Monaco, Theuns & Taffoni 2002).

PINOCCHIO consists of a two step procedure. In the first step we use LPT in the context of the extended PS approach, to provide predictions for the collapse of fluid elements in a given numerical realization of a linear density field. Mass elements are assumed as collapsed at orbit crossing. In the second step the collapsed points are grouped into haloes using a "fragmentation" algorithm that mimics the hierarchical growth of structures through accretion and mergers.

This Chapter is organized as follow. In §4.1 we present the algorithm to evalu-

	$N_{\rm part}$	$L_{\rm box}$ (Mpc/h)	Ω_0	Ω_Λ	h	Γ	$\sigma_8(z=0)$	$M_{\rm part}$ (M$_\odot$)
SCDM	360^3	500	1.0	0.0	0.5	0.5	1.0	1.49×10^{12}
ΛCDM	256^3	100	0.3	0.7	0.65	0.195	0.9	7.64×10^{9}
ΛCDM128	128^3	100	0.3	0.7	0.65	0.195	0.9	6.11×10^{10}

Table 4.1: Simulations used for the analysis.

ate the collapse time of fluid elements, while §4.2 deals with the "fragmentation" algorithm. The performance of the PINOCCHIO code is analysed in §4.3.

4.1 The collapse algorithm

In the PINOCCHIO formalism, LPT is used to predict the collapse time (in the orbit crossing-sense [1]) of fluid elements, starting from a numerical realization of a linear density field.

Then, before using OC as collapse prediction, it is necessary to decide whether LPT (and ELL in particular) is accurate enough to reproduce the OC-collapsed regions, and how these are related to the relaxed haloes. This can be done by applying LPT to the initial conditions of a large N-body simulation, and comparing the LPT OC regions to those computed by the simulation.

4.1.1 The simulations

For this and further comparisons we use two collisionless simulations. The first, a standard CDM model (SCDM), has been performed with the PKDGRAV code, and consists of 360^3 ($\sim 46 \times 10^6$) DM particles (Governato et al. 1999). This simulation was provided by Fabio Governato, Tom Quinn and Joachim Stadel. The second simulation has been performed with the Hydra code (Couchman, Thomas & Pearce 1995), and consists of 256^3 DM particles in a flat Universe with cosmological constant (ΛCDM) (see also §3.4.1). Tom Theuns provided this simulation in cooperation with Silicon Graphics/Cray Research utilising the Origin 2000 super computer at DAMTP, Cambridge.

---

[1] Mass elements are assumed to have collapsed after undergoing orbit crossing.

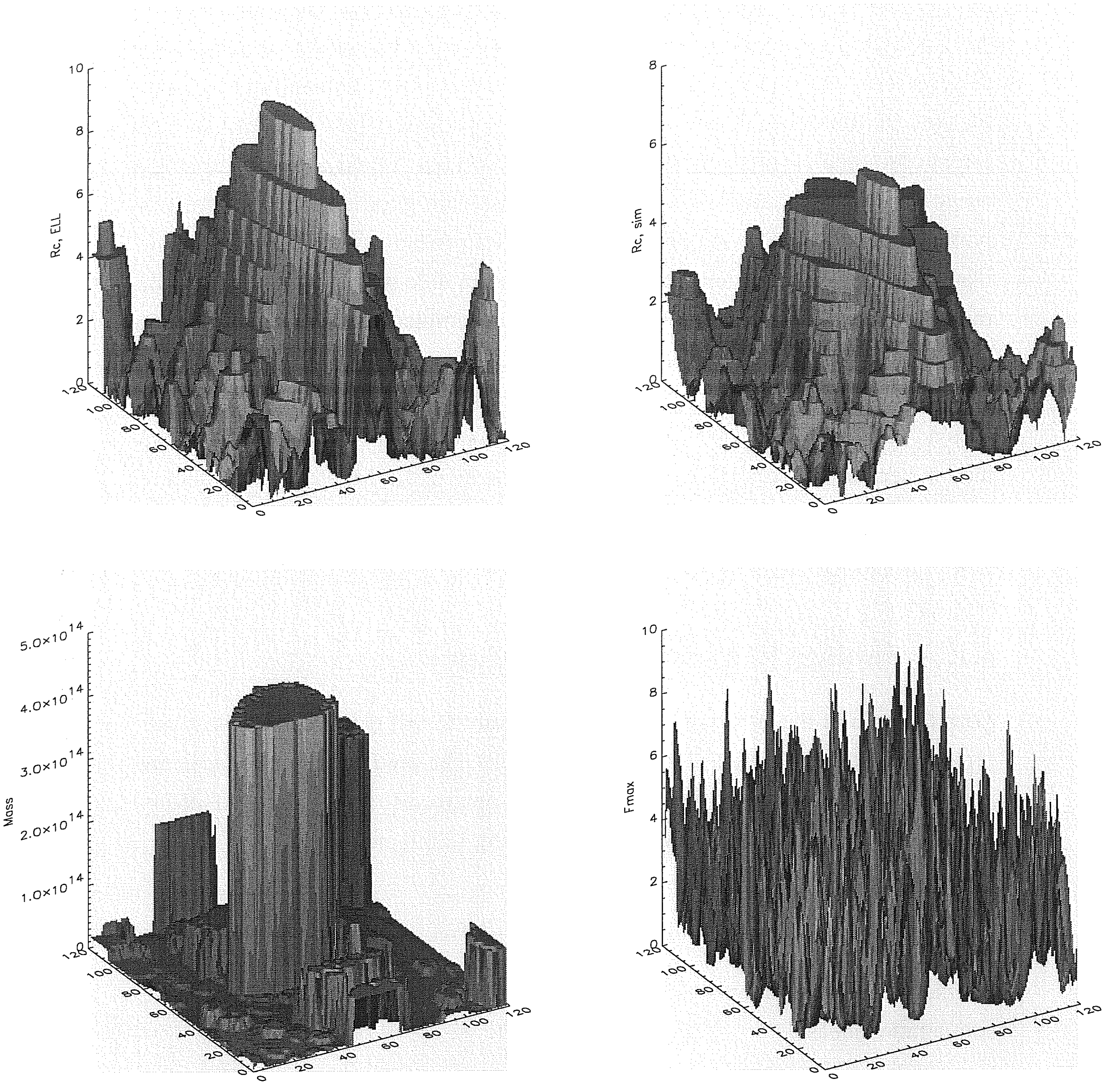

Figure 4.1: Upper panels: collapse radius fields R_c for a section of the Lagrangian space of the ΛCDM simulation at redshift $z = 0$. In the left panel we show the ELL prediction, and in the right panel the results from the simulation. Lower left panel: mass field for the same section; the mass field gives for each particle the mass of the halo it belongs to at $z = 0$. Un–grouped particles are assigned 0 mass. Lower right panel: inverse collapse time $F_{\max}$ for the same section.

In order to test for resolution effects, the latter simulation has been run with 128^3 particles, resampling the initial displacements on the coarser grid (we will refer to it as ΛCDM128). The main characteristics of the simulations are summarised in Table 4.1. These simulations allow us to test PINOCCHIO for different cosmologies, different resolutions, and different N-body codes, reaching a range of at least 5 orders of magnitude in mass with good statistics in terms of both numbers of haloes and numbers of particles per halo. The PKDGRAV simulation samples a very large volume, making it suitable for testing the high mass tail of the mass function. The Hydra simulation samples a much smaller volume but at higher resolution, so we can test the power-law part of the mass function at small masses. Note that in all the simulations the particles are initially placed on a regular cubic grid. We have compared our results with another ΛCDM simulation performed with PKDGRAV, with the same box (in Mpc/h) and number of particles as the SCDM one. The comparison confirms all the results given in this Chapter, but is not presented here.

4.1.2 Testing OC as definition of collapse

With the help of the N-body simulations we want to test if OC can be used as a proper definition of collapse for a mass element. The predictions of collapse are performed as follows. The linear contrast δ_l is obtained from the initial displacements of the simulation using the relation (see eq. [2.69] and eq. [2.70]):

$$S_{a,a}(\mathbf{q}, t_i) = -\delta_l(\mathbf{q}) b(t_i). \tag{4.1}$$

For the SCDM simulation the displacements are first resampled on a 256^3 grid for computational ease.

Comment: In this case, as well as in the whole collapse algorithm, differentiations are performed with Fast Fourier Transforms (FFTs). This procedure allows one to recover the linear contrast with minimum noise and no bias.

The linear contrast δ_l is then FFT-transformed and smoothed on many scales R with a Gaussian window function in Fourier space:

$$\tilde{W}(kR) = \exp\left(-k^2R^2/2\right). \tag{4.2}$$

The smoothing radii are equally spaced in $\log R$, except for the smallest smoothing radius which is set to 0 in order to recover all the variance at the grid scale. The largest smoothing radius is set such that the variance of the linear density contrast $\sigma(R_{\rm max}) = 1.686/6$, making the collapse of a halo at this smoothing scale approximately a 6 σ event. The smallest non-zero smoothing radius is set to a third of $R_{\rm max}$. Because of the stability of Gaussian smoothing, 25 smoothing radii in addition to $R = 0$ give adequate sampling for a 256^3 realisation (we use 15+1 smoothing radii for 128^3 grids). For each smoothing radius R the deformation tensor, $\varphi_{a,b}(\mathbf{q}, R)$, is obtained in the Fourier space from the FFT-transformed, smoothed linear density contrast $\tilde{\delta}_l(\mathbf{k}; R)$ as $\tilde{\varphi}_{a,b}(\mathbf{k}; R) = -k_a k_b/k^2\, \tilde{\delta}_l(\mathbf{k}; R)$, and then transformed back to real space, again with FFT. The ELL collapse times are computed for each grid point from the value of the deformation tensor as described in §2.7.3 and Appendix B of Monaco (1997a).

As in the EPS formalism, it is convenient to use the growing mode $b(t)$ as time variable, and in place of the collapse time t_c to record the growing mode at collapse, $b_c = b(t_c)$. With the procedure outlined above, a collapse time is computed for each grid vertex $\mathbf{q}$ and for each smoothing radius R, i.e. $b_c = b_c(\mathbf{q}; R)$. We define the inverse collapse time field F as:

$$F(\mathbf{q}; R) \equiv 1/b_c(\mathbf{q}; R). \tag{4.3}$$

In the case of linear theory $F = \delta_l/\delta_c$. The values of the F-field at a single point $\mathbf{q}$ correspond to the trajectories in the $F - R$ plane (or equivalently the $F - \sigma^2(R)$ plane) used in the excursion set approach to compute the mass function (see §2.6.4). In fact, as shown by Monaco (1997b), this quantity is obtained from the absorption rate of the $F(R)$ trajectories by a barrier put at a level F_c.

Warning: As the smoothing filter is Gaussian, these trajectories *are not* random walks as in the EPS formalism, and they are strongly correlated.

In general, the computation of the absorption rate requires no free parameter as long as the collapse condition does not. To solve the cloud-in-cloud problem, we record for each grid point the *largest* radius R_c at which the inverse collapse time overtakes F_c; the grid point is assumed to be collapsed at all smaller scales. We call $R_c(\mathbf{q})$, the *collapse radius* field (Monaco & Murante 1999). R_c depends on the height of the barrier as well as on time.

The R_c field for the simulations is obtained as follows. The displacement field $\mathrm{S}^{\mathrm{sim}}$ (i.e. the displacement of N-body particles from their initial position on the grid) is smoothed in the Lagrangian space $\mathbf{q}$ with the same set of smoothing radii. (Also here, we resample the large 360^3 simulation to a 256^3 grid using nearest grid point interpolation.) Each smoothed field is differentiated using FFTs along the three spatial directions and the Jacobian determinant, $\det(\delta^K_{ab} + S^{\mathrm{sim}}_{a,b})$, is computed for each grid vertex. For each grid point, we again record the largest smoothing radius $R^{\mathrm{sim}}_c(\mathbf{q})$ at which the Jacobian determinant first becomes negative (hence passing through 0).

The R_c field computed using LPT and obtained from the simulations at redshift $z = 0$ are compared in Fig. (4.1). The two fields are remarkably similar, exhibiting the same structure of broad peaks, with the difference that the peaks of the simulation are lower, as anticipated by Monaco (1999). In Fig. (4.2) we show a more quantitative point-by-point comparison between the two fields [2]. There is a reasonably tight correlation between the predicted and numerical collapse-radius fields, which confirms the power of LPT to predict the mildly non-linear evolution of perturbations; it is noteworthy that this comparison does not involve free parameters.

---

[2] For display purposes, some random noise has been added to the discrete values of R_c; in this way the values lie in squares instead of points

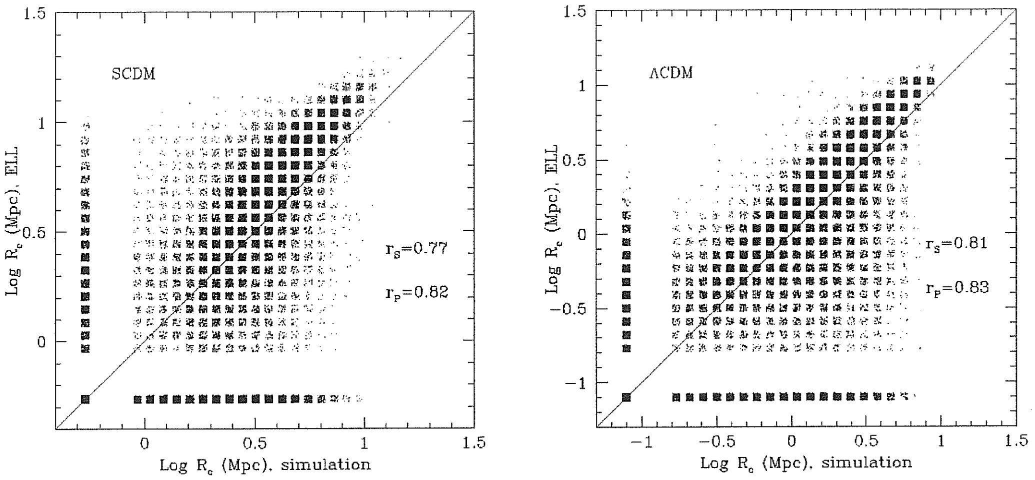

Figure 4.2: Comparison of the collapse radius fields R_c, as predicted by ELL, with the values found in the simulations, for a random sample of ~20000 points for the SCDM model (left panel), and the ΛCDM model (right panel). For clarity some random noise has been added to the discrete R_c values, so that they lie on squares instead of points.

Comment: The correlation is quantified by the Spearman rank correlation coefficient r_S and Pearson's linear correlation coefficient r_P, both reported in the panels of Fig. (4.2): a high value of r_S indicates the existence of a relation with moderate scatter, a high value of r_P indicates the existence of a good linear relation. The coefficients take the value ~ 0.8, confirming the correlation.

However, as also noted in Fig. (4.1), the relation between the two R_c fields is not unbiased: the simulated R_c field is lower than the ELL one, especially at large R-values. The cause of this behaviour can be understood as follows. LPT predicts that after OC particles do not remain bound to the caustic region but move away from it, in contrast to what happens in the simulations. Therefore, as in this analysis particles are not explicitly restricted to the pre-OC (single stream) regime, the displacements in the simulation are always smaller than those predicted by LPT. As a consequence, the collapse radius obtained by the smoothed displacements of the simulation is lower than that predicted by LPT. This bias disappears at small radii, which are however dominated by numerical noise.

The difference between the LPT and the simulation R_c fields can be also quantified by the cumulative distribution of the R_c fields as a function of R, or equivalently of the variance $\sigma^2(R)$. We will denote this function by $\Omega(<\sigma^2)$, since it is also the fraction of mass collapsed on a scale $\geq R$ where the rms is smaller than σ. This quantity is used in the PS approach to obtain the mass function (eq. [2.55]). The functions $\Omega(<\sigma^2)$ from ELL and the simulations are compared in Fig. (4.3). The LPT curves are by construction independent of time and cosmology, so that only the $z = 0$ LPT prediction is shown. In contrast, the Ω curves obtained from the simulations change with time. At late times, particles have crossed the structure they belong to many times and the numerical displacements differ more and more from the LPT ones. This is confirmed by the fact that the point of intersection between the $\Omega(<\sigma^2)$ obtained from LPT and simulation roughly scales as $b(t)^2$. Most notably, the difference between predictions and simulations tends to vanish for the highest redshifts; in this case the particles have not had time to cross the structures, and their trajectories are very similar to the LPT ones. In all cases we notice that the numerical $\Omega(<\sigma^2)$ functions become larger than the LPT ones at the smallest, unsmoothed scales, especially in the SCDM case and at higher redshift. This is most likely due to numerical noise present in the simulation, that enhances the level of non-linearity of the displacements, and in the SCDM case to the resampling from 360^3 to 256^3 grids.

For comparison, we show in Fig. (4.3) also $\Omega(<\sigma^2)$ from linear theory with $\delta_c = 1.686$, which falls short of both the ELL prediction and the simulations. We have verified that linear theory (with Gaussian smoothing!) misses the collapse of many mass points that belong to filaments or to low mass haloes. Decreasing δ_c to 1.5 improves the agreement only at the largest masses, but does not solve the problem at small masses. The Zel'dovich approximation severely under predicts $\Omega(<\sigma^2)$ at large masses, but approaches the ELL curve for lower mass (Monaco 1997a). Consequently, using either linear theory or the Zel'dovich approximation instead of ellipsoidal collapse, would significantly decrease the accuracy of PINOCCHIO

We have also computed the R_c field using full 3rd-order LPT. With respect

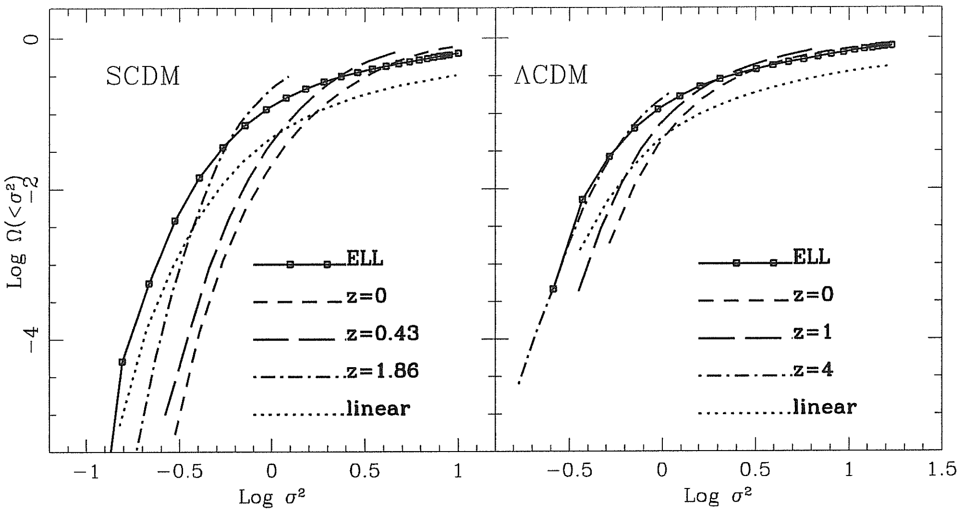

Figure 4.3: Cumulative distributions $\Omega(<\sigma^2)$ of the R_c fields for the SCDM and ΛCDM simulations.

to ELL, the fraction of collapsed points increases at small scales R, but decrease at large radii, because of the already mentioned inability of 3rd-order LPT to reproduce the spherical limit correctly (Monaco 1997a). We have verified that the correlation with the numerical R_c field is noisier, and that the additional small-scale contribution of collapsed matter consists mainly of particles in filaments. Moreover, the computation is much more demanding than the ELL case. We conclude that there is no advantage in using the full 3rd-order LPT solution.

4.1.3 R_c and the simulated haloes

Having demonstrated the ability of LPT in predicting collapse in the OC sense (without free parameters), we need to decide whether OC may be of any use to predict which mass elements are going to end up in relaxed haloes. In order to do so, we compute the 'mass field' from the simulation, which assigns to each grid vertex in the initial conditions, the mass of the halo that the corresponding particle ends-up in. Haloes have been identified in the simulation using a standard friends-of-friends (FOF) algorithm, with a linking length 0.2 times the mean inter particle distance.

Warning: The simulation haloes were identified using a standard FOF algorithm with linking length 0.2, irrespective of cosmology. In this way, haloes are defined above a fixed fraction of the *mean* density – as opposed to a fixed fraction of the *critical* density. Jenkins et al. (2001) showed that this makes the mass function almost universal with cosmology, and in addition it is similar to the definition used in PINOCCHIO.

The mass field is shown in Fig. (4.1) for the same slice of the ΛCDM simulation as the other panels. A FOF halo looks like a plateau, with the plateau's height giving the halo's mass. There is a broad agreement between the peaks in the R_c and mass fields, because massive (low mass) objects are generally associated with large (small) smoothing radii. Consequently, there certainly is some connection between orbit crossed regions and relaxed haloes. However, there are some important differences as well.

Not all the FOF points fall within the boundaries of the R_c contours. This fact was already addressed by Monaco & Murante (1999), and is expected because the OC criterion tends to miss those infalling particles that have not made their first crossing of the structure. For SCDM, the fraction of FOF particles not predicted to be OC-collapsed ranges from $\sim$10 per cent at large masses to $\sim$20 per cent at smaller masses; smaller values are obtained for ΛCDM, where the fraction of collapsed mass is higher. This has a modest impact on the results, and is hardly noticeable in Fig. (4.1).

More importantly, the reverse is true as well: many particles assigned non-vanishing or even high R_c values do not belong to a halo. These particles are in the moderately over dense filaments and sheets that connect the relaxed haloes. These structures, although indeed in the multi-stream regime, are in a relaxation state very different from that of the haloes. It is apparent that the removal of such sheets and filaments (hereafter referred to as filaments) is an important issue that needs to be addressed.

4.1.4 Computing the collapse time

Another feature apparent when comparing the mass and R_c fields (Fig. [4.1]) is that many FOF haloes may correspond to a single broad peak of R_c. This makes the time-dependent R_c field unsuitable for addressing the fragmentation of matter into haloes and filaments. It is more convenient to follow a procedure similar to the merging cell model (Rodrigues & Thomas 1996; Lanzoni et al. 2000), i.e. recording for each mass point the largest F-value it reaches, or, in other words, the highest redshift at which the point is predicted to collapses in the OC sense (for SCDM it is simply $F = (1+z_c)$, where z_c is the collapse redshift). This is another way to solve the so-called cloud-in-cloud problem (Bond et al. 1991): a point that collapses at some redshift is assumed to be collapsed at all lower redshifts. We therefore record the following quantity:

$$F_{\rm max}(\mathbf{q}) \equiv {\rm max}_R[F(\mathbf{q}; R)]. \tag{4.4}$$

Together with $F_{\rm max}$, we also store for each point the smoothing radius $R_{\rm max}$ at which $F = F_{\rm max}$, and the corresponding Zel'dovich velocity $\mathbf{v}_{\rm max}$ computed at the time $b(t) = 1/F_{\rm max}$ appropriate for the smoothing radius $R_{\rm max}$.

In contrast to R_c, the inverse collapse time $F_{\rm max}$ evidently does not depend on time, while it does depend on the smoothing radius. The excursion set of those points where $F_{\rm max}$ is greater than some level F_c gives the mass that has collapsed before the time t_c that corresponds to F_c, at the highest resolution on the grid (i.e. without smoothing, $R = 0$). The lower right panel of Fig. (4.1) plots the $F_{\rm max}$ field for the same section as the other panels. Within each large object identified in the mass field, F_c has many small peaks that correspond to objects forming at higher redshifts. These peaks are modulated by modes on a larger scale that follow the excursions of the R_c field. Those large scale modulations are ultimately responsible for the later merging of these small peaks into the massive object identified at late times. In this way, PINOCCHIO combines the information on the progenitors to reconstruct the merger history of objects, as described in detail in the next Section.

4.2 The Fragmentation algorithm: identification and evolution of haloes.

In the PS and excursion set approaches the mass of the objects that form at a scale R is simply estimated as

$$M \simeq \frac{4\pi}{3}\bar{\rho}R^3. \tag{4.5}$$

A more detailed treatment of the complex processes that determine the shape of the Lagrangian region which collapses into a single halo is required to get an improved description of the formation of the objects, and thus an improved agreement with simulations at the object-by-object level. In PINOCCHIO, this is done by generating realisations of the density field on a regular grid, computing the $F_{\rm max}$ field as explained in the previous Section, and then 'fragmenting' the collapsed medium into haloes and filaments by considering the fate of each particle separately.

The fragmentation code mimics the two main processes of hierarchical clustering: the accretion of mass onto haloes and the merging of haloes. The particles of the realisation are considered in order of descending $F_{\rm max}$-value, i.e. in chronological order of collapse. At a given time the particles that have already collapsed will be either assigned to a specific halo, or associated with filaments. Because of the continuity of the transformation between Lagrangian and Eulerian coordinates (eq.[2.65]) a particle must touch a halo in the Lagrangian space if it will accrete on it.

Warning: Here it is assumed that a particle that accretes onto a halo never escapes back into the field. Such stripping does rarely happen in simulations, so we neglect it.

Thus a collapsing particle can accrete only onto those haloes that are 'touched' by it, i.e. that already contain one of its 6 nearest neighbours in the Lagrangian space of initial conditions (we call these particles *Lagrangian neighbours*). To decide whether the particle does accrete onto a touching halo, we displace it to the Eulerian space according to its $\mathrm{v}_{\rm max}$ velocity. The halo is displaced to its Eulerian position at the time of accretion, using the average velocity of all its constituent particles.

Comment: The velocity of a halo is an average over velocities calculated at different smoothing radii. A better estimate (but expensive in terms of computer memory) would be to average the unsmoothed velocities over the particles of the halo. Fortunately,the stability of the velocity to smoothing makes the two estimates very similar, once the average is performed over many particles.

In the following we express sizes and distances in terms of the grid spacing. The size $R_{\rm N}$ of a halo of N particles is taken to be:

$$R_{\rm N} = N^{1/3}. \tag{4.6}$$

The collapsing particle is assumed to accrete onto the halo, if the Eulerian (comoving) distance d between particle and halo is smaller than a fraction of the halo's size $R_{\rm N}$,

$$d < f_{\rm a} \times R_{\rm N}. \tag{4.7}$$

The free parameter $f_{\rm a}$, which is smaller than one, controls the over density that the halo reaches in the Eulerian space, $1 + \delta_{\rm halo} \sim 3/4\pi f_{\rm a}^3$. Therefore, this criterion selects haloes at a given over density, making it similar to the usual FOF or similar selection criteria.

When a collapsing particle touches two (or more) haloes in the Lagrangian space, then we use the following criterion to decide whether the two haloes should merge. We compute the Eulerian distance d between the two haloes at the suspected merger time using the halo velocities described above. The haloes are deemed to merge when d is smaller than a fraction of the Lagrangian radius of the larger halo:

$$d < f_{\rm m} \times \max(R_{\rm N1}, R_{\rm N2}). \tag{4.8}$$

This condition amounts to requiring that the centre of mass of the smaller halo, say halo 2, is within a distance $f_{\rm m} R_{\rm N1}$ of the centre of mass of the larger halo 1.

We note that PINOCCHIO is not restricted to binary mergers. In principle, a particle has 6 Lagrangian neighbours so up to 6 haloes may merge at the same time. In practice binary mergers are the most frequent, but ternary mergers also occur, while mergers of four haloes or more are rare.

In more detail, the fragmentation code works as follows. We keep track of halo (or filament) assignment for all particles. For each collapsing particle we consider the halo assignment of all Lagrangian neighbours; touching haloes are those to which a Lagrangian neighbour has been assigned. The following cases are considered:

(1) *Seed haloes:* If none of the neighbours have collapsed, then the particle is a local maximum of $F_{\rm max}$. This particle is a seed for a new halo of unit mass, created at the particle's position.

(2) *Accretion:* If the particle touches only one halo, then the accretion condition is checked. If it is satisfied, then the particle is added to the halo, otherwise it is marked as belonging to a filament. The particles that only touch filaments are marked as filaments as well.

(3) *Merging:* If the particle touches more than one halo, then the merging condition is checked for all the touching halo pairs, and the pairs that satisfy the conditions are merged together. The accretion condition for the particle is checked for all the touching haloes both before and after merging (when necessary). If the particle can accrete to both haloes, but the haloes do not merge, then we assign it to that halo for which $d/R_{\rm N}$ is the smaller. Occasionally, particles fail to accrete even though the haloes merge.

(4) *Filaments:* With these rules for accretion and merging, some collapsing particles do not accrete onto a halo at their collapse time. Since these particles tend to occur in the mildly overdense regions that connect the haloes (visible as a filamentary network between haloes in simulations), we assign them to a 'filaments' group. When a particle is accreted onto a halo, all filament particles that neighbour it are accreted as well. This is done in order to mimic the accretion of filaments onto the haloes. Notice that up to 5 filament particles can flow into a halo at each accretion event.

This fragmentation code runs extremely quickly, in a time almost linearly proportional to the number of particles. At late times, slightly more time is spent in updating the halo assignment lists in case of mergers, but this does not slow down the code much.

In high density regions where most of the matter has collapsed, it can happen that pairs of haloes that are able to merge are not touched by newly collapsing particles for a long time. This problem can be solved by keeping track of all the pairs of touching haloes that have not merged yet, and checking the merging condition explicitly at some time intervals. Such a check slows the code down significantly, and has only a moderate impact on the results when the fraction of collapsed mass at the grid scale is large. Similarly, the accretion of filament particles onto haloes can be checked at some given time intervals, but again, the impact on the results is modest but the increase in computer time may be substantial.

4.2.1 The fragmentation parameters

While the dynamical estimate of collapse time does not introduce any free parameter, the fragmentation process does.

Comment: The same happens in the simulation, where any halo-finding algorithm has at least one free parameter, such as the linking length for FOF haloes. This is because the definition of what constitutes a DM halo is somewhat arbitrary, and hence also the corresponding mass function is not unique (Monaco 1999). Fortunately, different clump-finding algorithms usually give similar results, so that this ambiguity is in general not a real problem.

The accretion and merging conditions given in eq.(4.7) and eq. (4.8) introduce the first two parameters. Unfortunately those conditions work well when the haloes contain sufficiently many particles, while for smaller haloes, the limiting distance $f_a R_N$ or $f_m R_N$ may be comparable to the grid spacing. In this case, the Zel'dovich velocity v_{max} needs to be very accurate in order that accretion or merging take place, and this may lead PINOCCHIO to underestimate the number of very low mass objects. The simplest solution to this problem is to add a

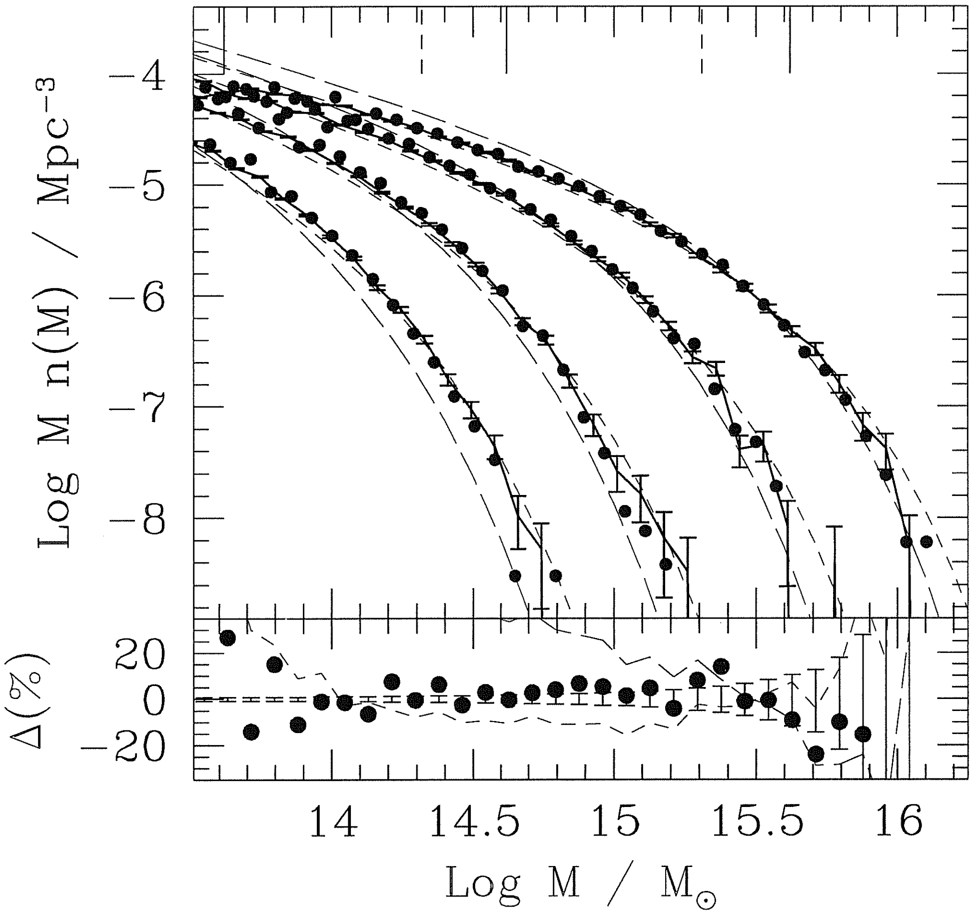

Figure 4.4: Comparison of mass function $Mn(M)$ in a SCDM model ($\Omega_m = 1$). Top panel: simulated mass function for FOF selected haloes (Full lines with Poissonian error bars), PINOCCHIO mass function (filled circles), the fit by Sheth and Tormen (short-dashed lines) (eq. [2.64]) and PS function (long-dashed lines) (eq. [2.62]) , at redshifts $z = 0$, 0.43, 1.13 and 1.86 (higher redshift curves are off-set by 0.1 dex both vertically and horizontally for improved clarity). Vertical lines show limits corresponding to simulation haloes with 10, 50, 100, 500 and 1000 particles (256^3 re-sampling). Bottom panel: Difference between simulated mass function and PINOCCHIO (filled dots), Sheth and Tormen fit (short-dashed line) and PS (long-dashed line) at $z = 0$.

constant $f_{\rm r}$ to the right hand side of eq. (4.7) and eq. (4.8), of order of the grid spacing. This brings the number of parameters to three: $f_{\rm a}$, $f_{\rm m}$ and $f_{\rm r}$.

The best fit parameters for PINOCCHIO will be chosen so as to reproduce the mass function of the FOF haloes of the simulations (see Table [4.1]), with linking length equal to 0.2 times the inter-particle distance, at many redshifts [3] (see Fig. [4.6] and Fig. [4.4]).

If we use the SCDM simulation to calibrate the fragmentation parameters we find that the optimal values are $f_{\rm a} = 0.18$, $f_{\rm m} = 0.35$ and $f_{\rm r} = 0.70$. As shown in Fig. (4.4), with this particular choice, PINOCCHIO reproduces the mass function $Mn(M)$ to better than $\gtrsim$ 10 per cent at all redshifts, in the mass range in which haloes have at least ~30 particles and Poisson error bars are small. In the bottom panel of Fig. (4.4), we plot the residuals with respect to the $z = 0$ FOF mass function. The PS mass function, which over (under) predicts the number of low (high) mass objects, is shown for comparison as well.

However, when applied to the ΛCDM simulation, this choice produces a systematic excess of low mass objects at high redshift, of order $\gtrsim$ 20 per cent at $z = 4$ for objects of 30 particles. This excess is barely noticeable in the SCDM simulation at $z = 1.13$ (Fig. [4.4]). This systematic effect is due to the fact that the accuracy of LPT in estimating the velocities is not constant in time: it depends on the degree of non-linearity reached, worsening at later times. It can be measured by comparing the Zel'dovich displacements with those from the simulation, for particles that are just experiencing OC collapse, according to the $F_{\rm max}$ field.

In Fig. (4.5) we show that the error in the displacement increases as the field becomes more non-linear. The rate of increase is very similar for the two cosmological models plotted. The errors in the displacements are much smaller than the displacements themselves, demonstrating the power of the Zel'dovich approximation. While the average displacement grows as $b(t)$, its error grows as $b(t)^{1.7}$.

The fact that displacements are computed more accurately at earlier times has two important consequences. First, the accuracy of the reconstruction of particle position will degrade with time, as we illustrate in §4.1.2. Second, for a given

---

[3] We have checked with one SCDM output that the differences in the haloes as defined by the HOP (Eisenstein & Hut 1998) and SO (Lacey & Cole 1994) algorithms are much smaller than the accuracy with which we are able to recover the FOF haloes.

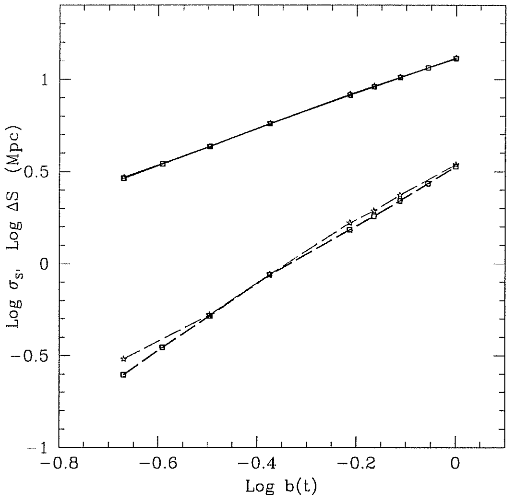

Figure 4.5: Error in the estimate of the Zel'dovich displacements for particles that have just undergone orbit crossing, as function of the growing mode b. Continuous lines are the average displacement of the collapsing particles, dashed lines are the error in the estimate of these displacements, as computed by comparing with the simulation results. Thick lines are obtained from the ΛCDM simulation, thin lines from the ΛCDM128 simulation.

set of parameters $f_{\rm a}$, $f_{\rm m}$ and $f_{\rm r}$, objects will tend to accrete mass more easily at higher redshifts than at later times. The reason is that, in order to accrete a particle onto a halo at late times, we should make these parameters sufficiently generous so that the particle falls within d of the halo according to eq. (4.7), even though its position is not accurately determined. This may lead to too much accretion at earlier times, when the positions are more accurate.

It is possible to improve PINOCCHIO to correct for this numerical problem. What is relevant in the fragmentation code is not the absolute displacement of a particle, but the displacement relative to that of the halo. The distance between a collapsing particle and the centre of mass of a group is $d \sim S_{a,b} \times R_{\rm N}$. Considering that $S_{a,b} \propto b$, its variance scales as the variance σ of the linear density and the relative error on $S_{a,b}$ grows $\propto b(t)^{0.7}$. We can estimate the uncertainty on d, given the errors in reconstructing positions as:

$$\delta d = f_{\rm s}\sigma(R_{\rm N})R_{\rm N}b(t)[\sigma(R_{\rm N})b(t)]^{0.7}. \qquad (4.9)$$

Here, f_s is another free parameter. We only introduce this extra parameter in the accretion condition, since the results do not improve when we apply a similar correction to the merger condition. The accretion and merging conditions are then respectively:

$$d < f_\mathrm{a} \times R_\mathrm{N} + f_{ra} + \delta d \ , \tag{4.10}$$

$$d < f_\mathrm{m} \times \max(R_\mathrm{N1}, R_\mathrm{N2}) + f_{rm}. \tag{4.11}$$

We note that the resolution parameter f_r is now different for accretion and for merging.

4.2.2 Fixing the parameter values

Our algorithm contains now five parameters. Their values have been determined by generating many realisations of Gaussian fields for different cosmological models, box sizes and resolutions (including the initial conditions of the SCDM, ΛCDM and ΛCDM128 simulations used here), and reproducing for each realisation the corresponding mass function. In the cases where we do not have the N-Body simulation results, we fit the PINOCCHIO mass function to the analytical mass function of Sheth & Tormen (1999).

The best fit is easily achieved, as the effects of small variations of only one parameter are rather simple. In particular, f_m determines the overall slope of the mass function, f_rm the slope at low masses, f_a the normalisation, f_ra the abundance of low mass haloes and f_s the abundance of low mass haloes at low redshifts.

The best fit values are $f_\mathrm{m} = 0.35$ and $f_\mathrm{rm} = 0.7$. The parameters for accretion are found to be correlated,

$$f_\mathrm{ra} = 0.40 - 3.5\,(f_\mathrm{a} - 0.22)\,. \tag{4.12}$$

In addition, f_ra correlates with the degree of non-linearity, as quantified by $\Sigma \equiv \sigma(R = 0)/l_\mathrm{grid}$. Here, $\sigma(R = 0)$ is the variance at the level of the grid and l_grid the grid spacing. Σ is sensitive to both the degree of non-linearity reached and the level of accuracy of the Zel'dovich displacements. The best fit for f_a is

$$f_\mathrm{a} = 0.22 + (\log \Sigma - 0.36) * 0.11\,. \tag{4.13}$$

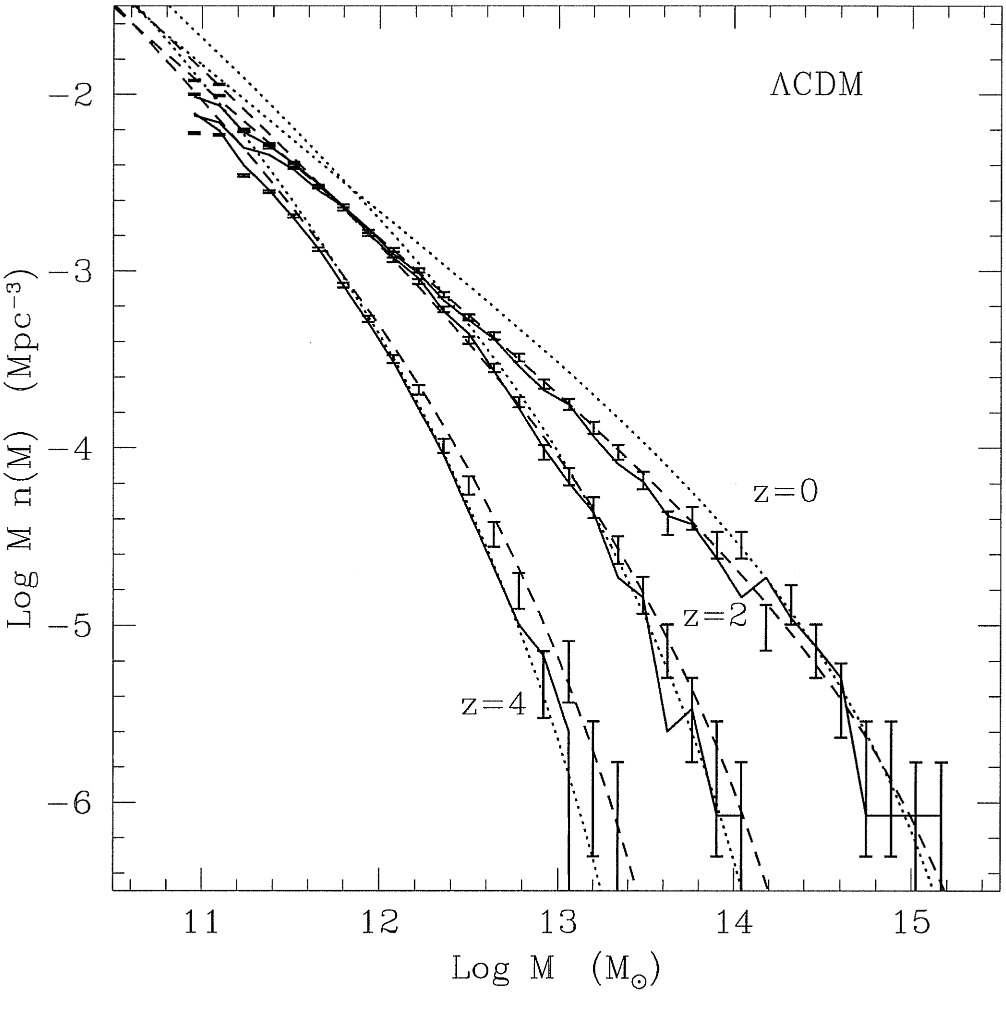

Figure 4.6: Mass functions for the ΛCDM model at different redshifts indicated in the panel. Error bars denote Poissonian errors for the simulated mass function, continuous lines are the PINOCCHIO predictions, dotted and dashed lines are the PS and ST predictions, respectively.

We also demand that $0.22 \leq f_{\rm a} \leq 0.26$. The best fit $f_{\rm s} = 0.06$. (In all cases a change in the last significant digit gives differences in the mass function appreciable at the 5 per cent level).

Unfortunately, these parameters are sensitive to the algorithm used to generate initial conditions, in particular they depend on how the small scale power close to the Nyquist frequency is quenched. For the ΛCDM models, we used the initial conditions generator distributed with HYDRA (Couchman et al. 1995), where power below the Nyquist frequency (on a grid with unit grid spacing, taken to be $k_e = 0.8\pi$), is quenched exponentially $\propto \exp[-(k/k_e)^{16}]$. The previous set of parameters apply for this type of initial condition generator. In contrast, the initial conditions for the 360^3 SCDM simulation were generated on a 180^3 grid, without an additional cut-off of small scale power. The corresponding PINOCCHIO parameters are $f_{\rm a} = 0.19$, $f_{\rm ra} = 0.60$ and $f_{\rm s} = 0.04$.

Other degeneracies are possible among these parameters. We have verified through extensive analysis that the object-by-object agreement is rather insensitive to the precise values, once the mass function fits well, the object-by-object agreement is good too.

4.3 Performance

We have applied PINOCCHIO to the initial conditions of the simulations presented in Table (4.1). Those simulations allow us to test the code for different cosmologies, different resolutions and different N-body codes.

The speed

First, we notice that PINOCCHIO is fast. Resampling the initial conditions onto a 256^3 grid, the first stage of computing orbit-crossing requires $\sim$ 6 hours of CPU time, the second step of identifying the haloes takes just a few minutes. (Timings refer to a Pentium III 450MHz personal computer. Memory requirement in this case amounts to $\sim$ 512Mbytes of RAM.) These timings should be contrasted with the several hundreds of hours on supercomputers required to perform the original simulations.

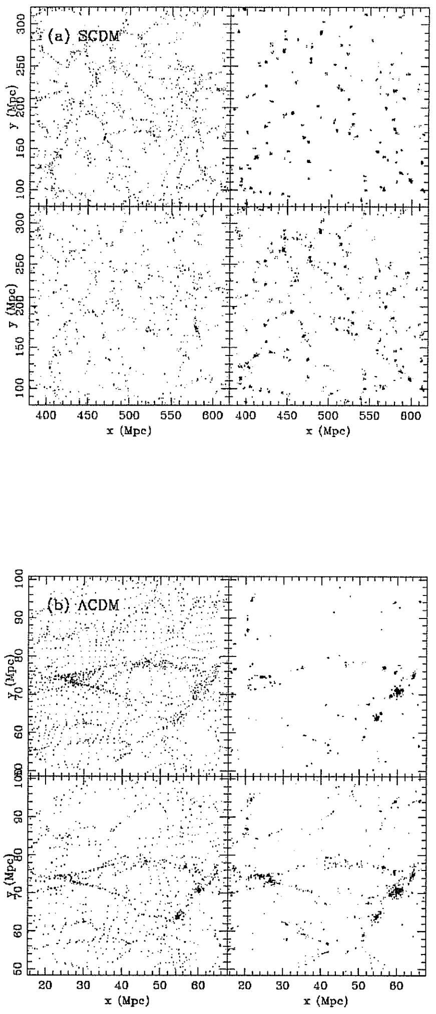

Figure 4.7: Final positions of particles at $z = 0$ from a slice of the initial conditions for the SCDM model (plot a) and the ΛCDM model (plot b). In each plot, the left panels show those particles that are in filaments (i.e. that have undergone OC but are not assigned to a halo), the right panels show particles that are assigned to haloes. Upper panels are obtained from the simulations, lower panels refer to PINOCCHIO.The large visual difference between the two cosmologies is mostly due to the very different box size used.

One way to understand the large speed-up between an N-body simulation and PINOCCHIO is that most of the CPU time used in the N-body simulation is spent integrating the orbits for particles *already inside a halo.* These particles undergo large accelerations as they orbit inside the halo, and hence may require thousands of time-steps in order for their orbits to be integrated accurately. PINOCCHIO, on the other hand, computes the particle's orbit *before* it enters any high density region and completely ignores particles once they are inside a halo.

Comment: A N-body simulation could have been run at a coarser resolution, e.g. with a larger softening (the one used was 20 kpc) or timestep, resulting in a lower running time for an accuracy still comparable to or better than that of PINOCCHIO. However, the softening cannot be increased much without starting to lose the small haloes and consequently changing the mergers. On the other hand, to avoid particles flying out of the objects, we must take care of not increasing the timestep too much. In this way it is difficult to gain more than a factor of a few in the running time, which remains much larger than the PINOCCHIO one.

Obviously, all information on the internal structure of the halo is lost in the process, but it is well known that several millions of particles are required to get the internal structure correctly. (See the controversy about the slope of halo profiles as determined using high-resolution collisionless simulations, e.g. Ghingha et al. 2000 and §7.3).

The mass function

The comparisons of PINOCCHIO and FOF mass function for the SCDM simulation presented in Fig. (4.4), is obtained using the limited set of three free parameters. The results with the full five-parameter set are very similar and are not shown here. In Fig. (4.6), we compare the mass function computed using PINOCCHIO and the ΛCDM N-body simulation. The FOF haloes were identified as explained above. For reference, we also plotted the PS and Sheth & Tormen (1999; hereafter ST) mass functions. The choice of parameters reported before produces a

PINOCCHIO mass function which falls to within $\gtrsim$5 per cent of the simulated one from $z = 5$ to $z = 0$, for all mass bins with more than 30-50 particles per halo and for which the Poisson error bars are small. The only residual systematic is a modest, $\sim$10-20 per cent underestimate at the highest-mass bins and highest redshift. An accuracy of better than 10 per cent on the mass function for a given realisation is perfectly adequate for most applications, as it is usually smaller than the typical sample variance and of the intrinsic accuracy of $\sim 20 - 30$ per cent with which the mass function of N-body simulations is defined. Because PINOCCHIO is calculated for the *same* initial conditions as the simulation, Poisson error bars are not the correct errors to use for this comparison (notice that the Poisson error bars of the PINOCCHIO mass function are obviously very similar to those of the numerical one). We show them both for comparison with PS and ST and to understand which mass bins are affected by small number statistics. Taking the ST mass function (or the analytic fit of Jenkins et al. 2001) as a *bona fide* estimate, we have checked the validity of PINOCCHIO in reproducing the mass function of haloes in a wide variety of cosmologies and box sizes. The fit of the mass function is found to be still good even for halo masses as small as 10^5 $\mathrm{M}_\odot$ (ΛCDM cosmology), at a redshift high enough to avoid that the whole box goes non-linear.

Strictly speaking the agreement between the mass functions of PINOCCHIO and that measured in the numerical experiment is not a real prediction on the N-body results, as the fit is achieved by tuning the free parameters discussed in §4.2.1. However, the very existence of a limited set of parameters that allows to achieve such a good agreement in different cases (SCDM and ΛCDM, PKDGRAV and Hydra, small and large boxes) is a very important result. As shown also in Fig. (4.6), PINOCCHIO improves the fit with respect to PS, giving an accuracy very similar to the ST fit. Jenkins et al. (2001) showed that the ST fit underestimates the knee of the FOF mass function by $\sim$10-20 per cent[4]; we have verified that when this difference is evident the best fit PINOCCHIO mass function is more similar to the numerical one and to the Jenkins et al. (2001) fit than to the ST mass function. This is evident in Fig. (4.4) (where the residuals of the $z = 0$

---

[4] Sheth & Tormen (2001) show that a modest tuning of their parameters can remove this disagreement.

mass functions are shown), but is hardly noticeable in Fig. (4.6), where Poisson errorbars are larger. The comparison with the ΛCDM simulation shows that the fit is very good also down to the low mass tail $M \sim 10^{11}\ M_{\odot}$ or $M/M_* \sim 10^{-2}$ (M_* denotes the characteristic mass of the PS mass function, such that $\sigma^2(M_*) = \delta_c^2$; see, e.g., Monaco 1998).

In the PS and excursion set approaches, the mass function is 'universal' when expressed in terms of the variable $\Omega(<\sigma^2)$ (eq. [2.55]), which in this case gives the fraction of mass collapsed into objects larger than $M(\sigma^2)$ (with the mass given by eq. [4.5]). The mass functions obtained from a large set of numerical simulations is indeed found to be universal to within $\gtrsim$30 per cent (Jenkins et al. 2001). The PINOCCHIO mass function is not universal by construction, yet we find it to be nearly universal once the resolution effects described in the previous Section are taken into account.

However, the mass function of the Governato et al. (1999) SCDM simulation used here shows an excess of massive haloes at high redshift. This was already noticed by Governato et al. (1999), and quantified as a drift of the δ_c parameter from $\sim$1.5 at high redshift to $\sim$1.6 at $z = 0$. This trend is not confirmed by other simulations (Jenkins et al. 2001), nor by our ΛCDM simulation presented here. We find that PINOCCHIO reproduces the weak trend of Governato et al. (1999) in the SCDM simulation, but also the lack of such a trend in the ΛCDM one. We conclude therefore that this effect is likely to be linked to the initial conditions generator, which is different for the two realisations. Recall that the PINOCCHIO mass functions refer to the *same* initial conditions as were use to perform the simulations.

Haloes and filaments

The ability of PINOCCHIO to distinguish OC particles that collapse into haloes versus those that remain in filaments, is shown in Fig. (4.7). In this figure we plot the final position of the particles, at redshift $z = 0$, for a section of the initial conditions of the SCDM and ΛCDM simulations. Left panels show only the filament particles, defined as those which are in OC according to R_c but do not belong to any halo. Right panels show only those particles that are in haloes.

Upper panels show the result from the simulation, lower panels the PINOCCHIO predictions. Clearly, PINOCCHIO is able to distinguish accurately haloes from filaments, even though some filament particles are interpreted as halo particles and vice versa. When compared with Fig. (6) and Fig. (7) of Bond et al. (1991), Fig. (4.7) shows the marked improvement of PINOCCHIO with respect to the extended PS approach. We want to stress that filaments are important in their own right. For example, most of the Lyman-α absorption lines seen in the spectra of distant quasars are produced in filaments (e.g. Theuns et al. 1998), so it will be useful to be able to generate catalogues of haloes *and* filaments.

In the next Chapters we will focus on a detailed comparison with simulations in order to test the ability of PINOCCHIO in reproducing the numerical experiments. We will show that the match between our code and the N-body data is very good both for statistical quantities and at the object-by-object level.

Chapter 5

PINOCCHIO: the predictions

> *"I think I'll call him PINOCCHIO. This name will make his fortune.*
> *I knew a whole family of Pinocchi once, Pinocchio the father,*
> *Pinocchia the mother, and Pinocchi the children,*
> *and they were all lucky".*
>
> C. Collodi

In the previous Chapter we presented a detailed description of PINOCCHIO, a fast and perturbative approach to generate catalogues of DM haloes in hierarchical cosmologies. We also showed a preliminary comparison to simulations, demonstrating that PINOCCHIO can accurately reproduce the mass function of DM haloes measured in large N-body simulations.

However, the excellent performance of PINOCCHIO is not only confined to the determination of the mass function, it extends successfully to the object-by-object comparison. In fact, PINOCCHIO mimics the hierarchical growth of structures through accretion and mergers, and produces catalogues of haloes with known mass, position, and velocity. It provides a significant improvement over the EPS approach and its recent revisions (Sheth, Mo & Tormen 2001), which are known to be valid only in a statistical sense (Bond et al. 1991; White 1996).

In this Chapter we compared the results of PINOCCHIO with that of two N-

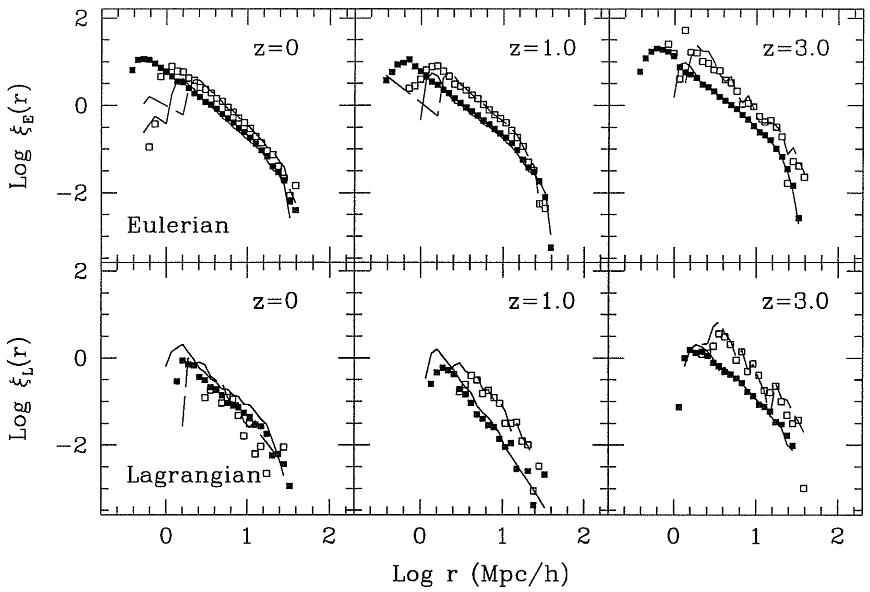

Figure 5.1: Eulerian correlation function (upper panels) and Lagrangian correlation function (lower panels) for the ΛCDM models at three redshifts indicated in the panel, and for two mass ranges. Symbols refer to simulation results, lines to PINOCCHIO predictions. Filled squares and continuous lines: correlation function for low mass haloes (mass M from 6.3×10^{11} to 3×10^{12} $M_{\odot}$), open squares and dashed lines: correlation function for massive haloes ($M > 3 \times 10^{12}$ $M_{\odot}$).

body simulations to test the ability of the code to reproduce the correlation function (§5.1.1) and also the mass, position and velocity of the DM haloes (§5.1.2). In §5.2 we show that PINOCCHIO can also predict the angular momentum of the haloes.

5.1 Detailed comparison to simulations.

The main characteristic of the simulations we adopt are summarized in Table 4.1.

To perform a detailed comparison we have to apply PINOCCHIO to the same realization of the initial density field used by the N-body code. In this way, we can compare the properties of *individual haloes*, not just the statistical quantities.

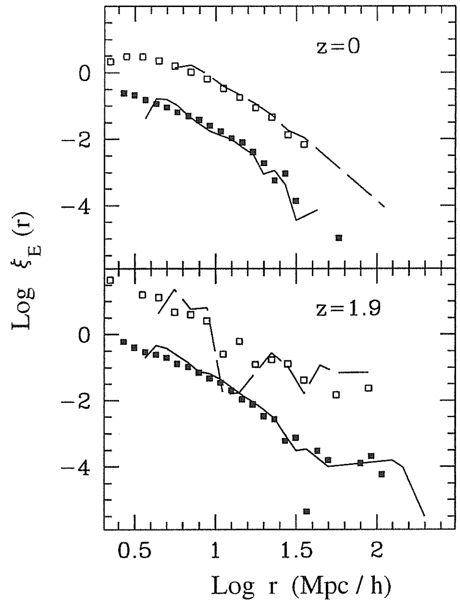

Figure 5.2: Eulerian correlation functions for SCDM models for the two redshifts indicated in the panels and for two mass ranges. Symbols refer simulation results, lines to PINOCCHIO predictions. Mass ranges are $10^{14} \leq \log M/M_{\odot} \leq 10^{14.5}$ (filled squares and full lines), and $\log M/M_{\odot} \geq 10^{14.5}$ (open squares and dashed lines respectively). Lower mass curves have been off-set vertically by 1 dex for clarity.

5.1.1 Statistical comparison

In Fig. (5.1) we show the correlation function of haloes as a function of mass, both in Eulerian and in Lagrangian space for the ΛCDM simulation. The correlation function has been computed using a standard pair counting algorithm. The agreement between PINOCCHIO and simulations is very good down to scales of a few grid cells, i.e. ~1-2 comoving Mpc/h (larger for rarer objects), below which the PINOCCHIO correlation functions become negative. This is in agreement with what found in Fig. (5.1.1) for the SCDM simulation where we show the results in the Eulerian space for two different redshifts.

The differences between PINOCCHIO and the N-body data are of order ~10-20 per cent in amplitude and $\lesssim$10 per cent in terms of scale at which a fixed amplitude is reached. This means that both the correlation length r_0, at which $\xi(r_0) = 1$, and the length at which $\xi = 0$ are reproduced with an accuracy of better than 10 per cent. This is an improvement with respect to the ST formalism, where the accuracy is of order $\sim$20 per cent (Colberg et al. 2001). More importantly, the trends of increased correlation for the more massive haloes, or for haloes of a given mass with increasing redshift, are both well reproduced.

Warning: The correlation functions in the Lagrangian space are noisier, and are reproduced with somewhat larger error, especially at $z = 0$ where they are slightly overestimated; however this error does not seem to propagate to the Eulerian correlation functions.

The two-point correlation function gives only a low-order statistics of the spatial distribution of a set of objects. To probe the accuracy of the PINOCCHIO results at higher orders, we have performed a count-in-cell analysis of the halo distribution, which, at variance with the correlation function, also depends on the phases of the space distribution of the haloes. This is shown in Fig. (5.3) for galactic-sized ($10^{12}M_\odot \leq M \leq 10^{13}M_\odot$) and group-sized ($M \geq 10^{13}M_\odot$) haloes of the ΛCDM realisation, and cell sizes of 2, 5 and 10 Mpc (corresponding to 1.3, 3.25 and 6.5 Mpc/h). The count-in-cells curves are well reproduced by PINOCCHIO, although their skewness is slightly underestimated, especially for larger cells and smaller masses. In particular, the void probability P_0 of finding no haloes in the cell is reproduced with a minimal accuracy of a few percent when it takes values in excess of 0.6.

5.1.2 Point-by-point and object-by-object comparison

The PINOCCHIO approach is also able to predict halo properties that correspond in detail to those obtained from simulations.

Agreement at the ‘point-by-point level’ requires that each particle is predicted to reside in the correct halo with the correct mass. Whether this agreement holds can be checked by comparing the mass fields already defined in §4 (an example of which is shown in Fig. [4.1c]).

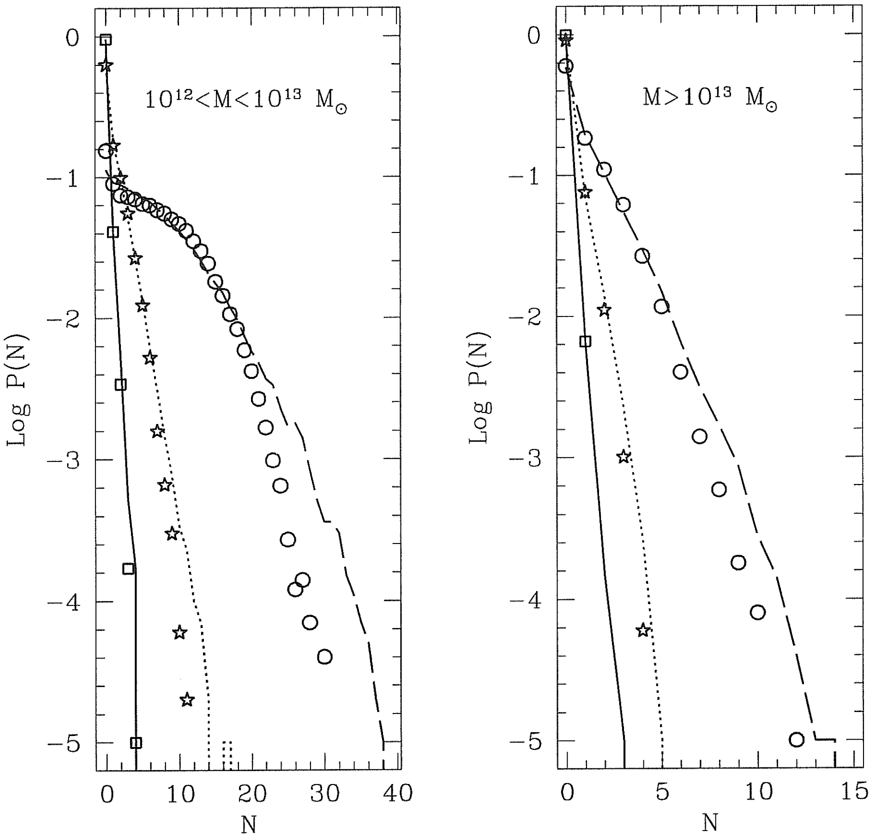

Figure 5.3: Count-in-cell analysis of the halo catalogues at $z = 0$. Left and right panels show results for the mass ranges indicated. Symbols refer to simulation results, lines to PINOCCHIO predictions. Continuous, dotted and dashed lines (or squares, stars and circles) refer to cell sizes of 2, 5 and 10 Mpc (1.3, 3.25 and 6.5 Mpc/h).

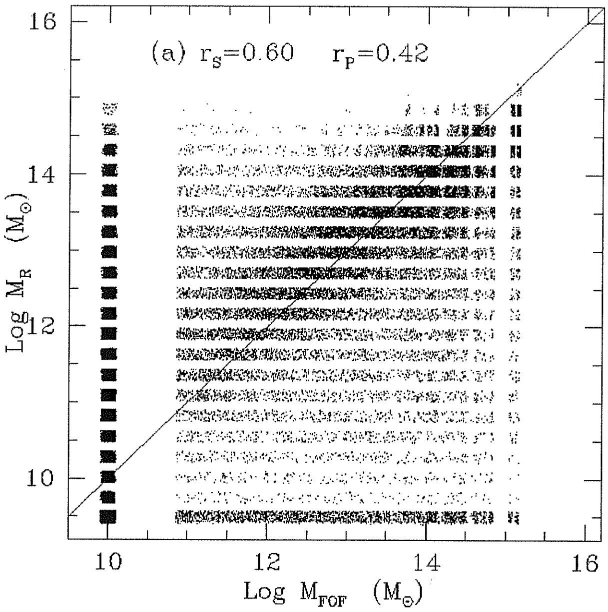

Figure 5.4: Comparison of the mass fields for FOF haloes identified in the ΛCDM simulation with that obtained from the R_c field using the PS mass-radius relation. For clarity some random noise has been added to the mass field values, especially to that obtained from the discrete R_c field.

In the PS approach, the mass of the halo to which a particle belongs is estimated as in eq. (4.5), with the $4\pi/3$ valid for top-hat smoothing (or sometimes left as a free parameter). In this case the mass field is simply related to the R_c field. A comparison between the mass fields obtained from the same R_c field of Fig. (4.2) (with arbitrary normalisation) and that of the simulation, $M_{\rm FOF}$, reveals only a poor correlation, as shown in Fig. (5.4) for a random sample of ~20000 particles extracted from the ΛCDM simulation. The tightness of the correlation is quantified by the r_S and r_P coefficients (see §4.1.2).

The point-by-point agreement is much better with PINOCCHIO (Fig. [5.5]), where the linear correlation coefficients r_P jumps from 0.42 to 0.69. The improvement of PINOCCHIO in the point-by-point comparison is not primarily due to the more accurate dynamical description of collapse. Rather it is due to the much more accurate description of the *shape* of the collapsing region, which is not restricted to the simple PS relation of eq. (4.5).

While the linear correlation coefficients improves significantly going from R_c to the PINOCCHIO mass field, the Spearman correlation coefficient r_S does not change much, since both panels contain a large number of outliers. However the presence of such outliers is not very important when the catalogue of objects is considered.

Comment: The "outliers" are particles that lie at the border of haloes, and are assigned to a halo by the simulation but not by PINOCCHIO, or vice versa. Such outliers are expected whenever the boundaries of haloes in the Lagrangian space are not perfectly recovered.

We next investigate the agreement of PINOCCHIO with the simulations at the object-by-object level, a coarser level of agreement but more relevant in practice. The degree of matching between halo catalogues is quantified as follows: for each halo of one catalogue, the haloes of the other catalogue that overlap for at least 30 per cent of the Lagrangian volume are considered. Among those, two haloes from different catalogues are 'cleanly assigned' to each other, when each overlaps the other more than any other halo . The fraction of haloes cleanly assigned is $f_{\rm cl}$, while the fraction of haloes that are split is $f_{\rm split}$. The remainder $1-f_{\rm cl}-f_{\rm split}$ is the fraction of haloes of one catalogue that does not overlap with anyone in the other catalogue. These fractions quantify the level to which two catalogues describe the same set of haloes. Another useful quantity is $f_{\rm ov}$, the average fraction that haloes overlap when they are cleanly assigned. All these estimators depend on whether PINOCCHIO is compared with simulations or vice-versa, but in general that difference is small as long as the comparison is good.

In Fig. (5.6) we show the values of these three indicators of the agreement between the two halo catalogues as a function of halo mass, for ΛCDM and the SCDM model. The agreement is very good at higher redshift with $\sim$80-90 per cent

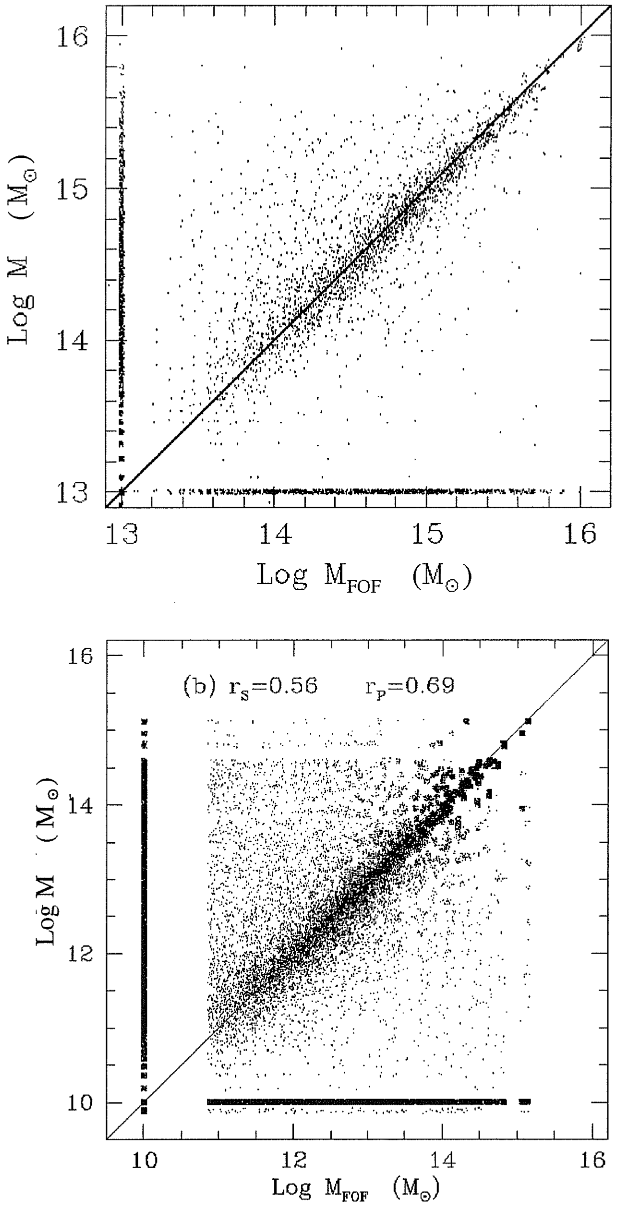

Figure 5.5: Comparison of the mass fields for FOF haloes identified in the simulations with that obtained from the R_c field using PINOCCHIO. The upper panel refers to the SCDM case and the lower panel to the ΛCDM. For clarity some random noise has been added to the mass field values.

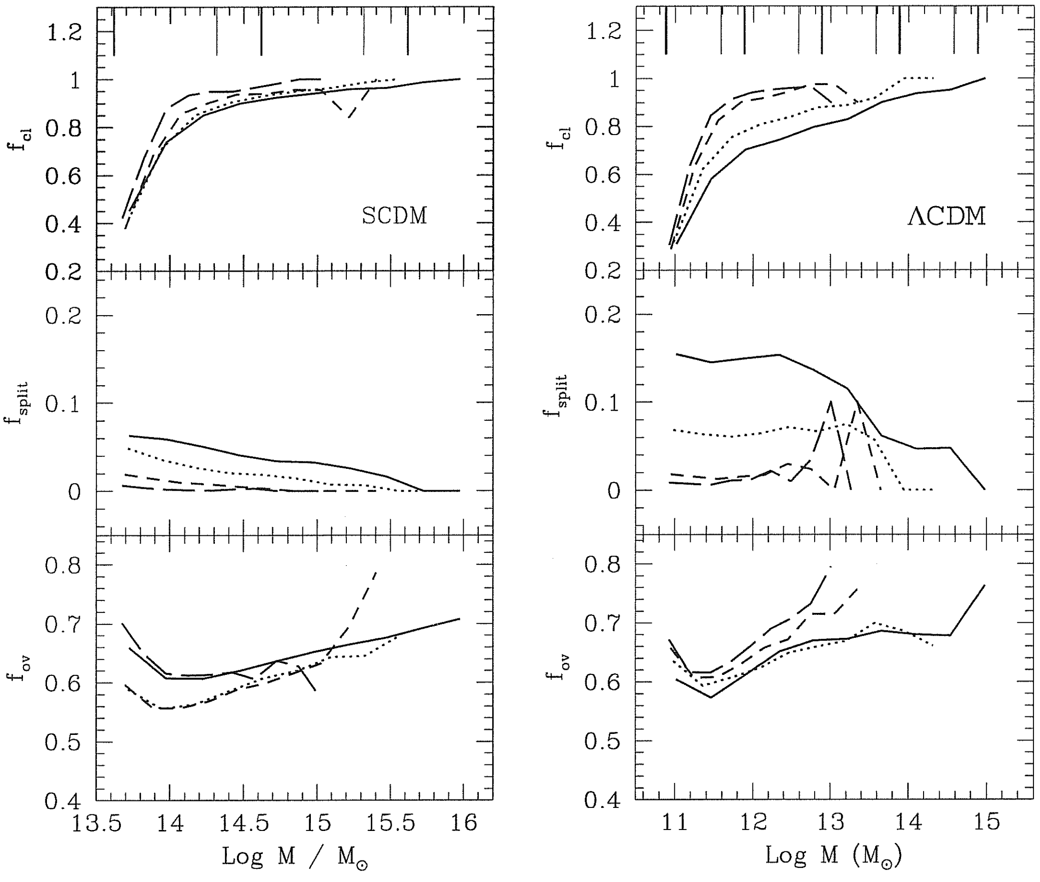

Figure 5.6: Comparison at the object-by-object level of haloes identified by PINOCCHIO and found in the ΛCDM (right panel) and SCDM (left panel) simulations, using a variety of statistics. Continuous, dotted, short-dashed and long-dashed lines refer respectively to redshifts $z = 0$, 1, 2 and 4. Top panel: fraction $f_{\rm cl}$ of cleanly assigned objects; middle panel: fraction $f_{\rm split}$ of non-cleanly assigned objects; bottom panel: average overlap $f_{\rm ov}$ for cleanly assigned objects. The vertical lines in the top panel indicate haloes with 10, 100, 10^3, 10^4 and 10^5 particles (heavy lines) or 50, 500, 5×10^3 and 5×10^4 particles (light lines).

of objects cleanly assigned when the haloes have at least 50 particles. The degree of splitting is $\lesssim$5 per cent, while the average overlap of cleanly-assigned objects $f_{\rm ov}$ ranges from 60 to 70 per cent nearly independent of mass and encouragingly larger than the 30 per cent lower limit. In the ΛCDM case, the agreement is slightly worse at lower redshift, with $f_{\rm cl}$ $\geq$70 per cent for haloes with at least 100 particles, and a $f_{\rm split}$$\sim$5–10 per cent. Within perturbative approaches there is obviously no advantage in going to higher resolution, as the accuracy of LPT worsens with the degree of non-linearity (see Fig. [4.5]). Anyway, the agreement is still very significant for the low redshift outputs, with a high fraction of cleanly assigned objects and a modest degree of splitting. In any case the results always improve with increasing number of particles. Monaco (1997a) estimated that LPT would break down when $\sim$50 per cent of the mass has undergone OC. Therefore, the agreement shown in Fig. (5.6) (and also in Fig. [4.7]) is better than expected.

In Fig. (5.7) we show the accuracy with which PINOCCHIO is able to estimate mass, Eulerian position and velocity of the cleanly assigned objects. In particular, we show both for SCDM and ΛCDM the scatter plots of the masses, and of velocity and position along one coordinate axis. For comparison, the scatter plot of the displacements of FOF haloes from the initial to the final positions are shown as well. Masses are recovered with an accuracy of $\sim$30 per cent for SCDM and $\sim$40 per cent for ΛCDM, nearly independent of mass. The average value is slightly biased, which results from our constraint in reproducing the mass function. Positions are recovered with a 1D accuracy of $\sim$1 Mpc, slightly depending on the box size and much smaller than the typical displacements, while velocities are recovered with a 1D accuracy of $\sim$150 or 100 km/s for SCDM or ΛCDM. In general, the velocities of the fastest moving haloes are underestimated. This could be fixed by extending the calculation of velocities to third order LPT, although a straightforward extension has been found not to work.

We stress that these comparisons are pure predictions of PINOCCHIO, in the sense that the free parameters of the method are constrained by the $z = 0$ mass function alone. The good agreement with the numerical simulations confirms that PINOCCHIO is a successful approximation to the gravitational collapse problem in a cosmological and hierarchical context.

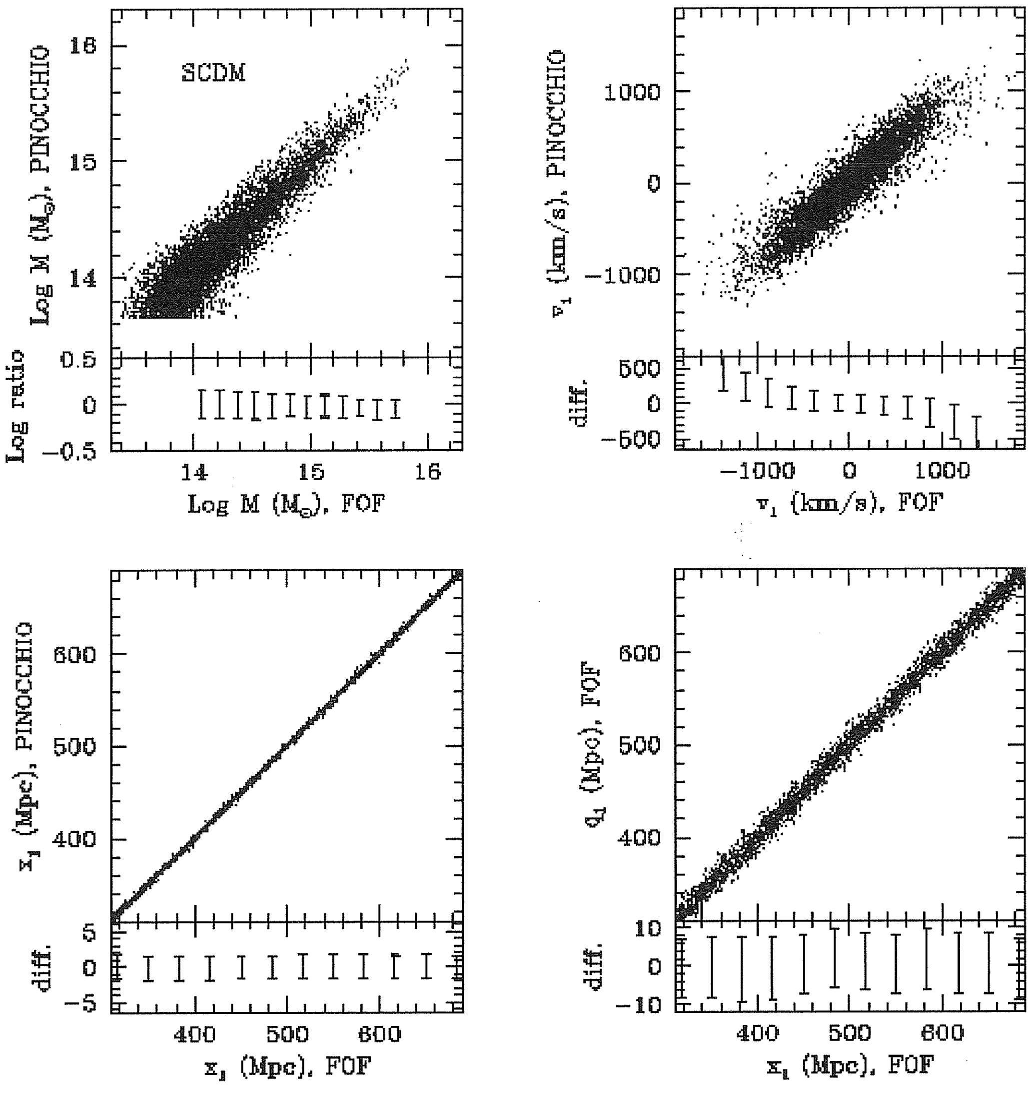

Figure 5.7: Difference in mass, position and velocity, $\log M$, x_1 and v_1 respectively, as estimated by PINOCCHIO and found from the simulation, for cleanly assigned haloes. The scatter around the mean is plotted below each panel. The lower right panels show for comparison the displacement of haloes according to the simulation. The first set of panels refer to the SCDM simulation, the second to the ΛCDM one

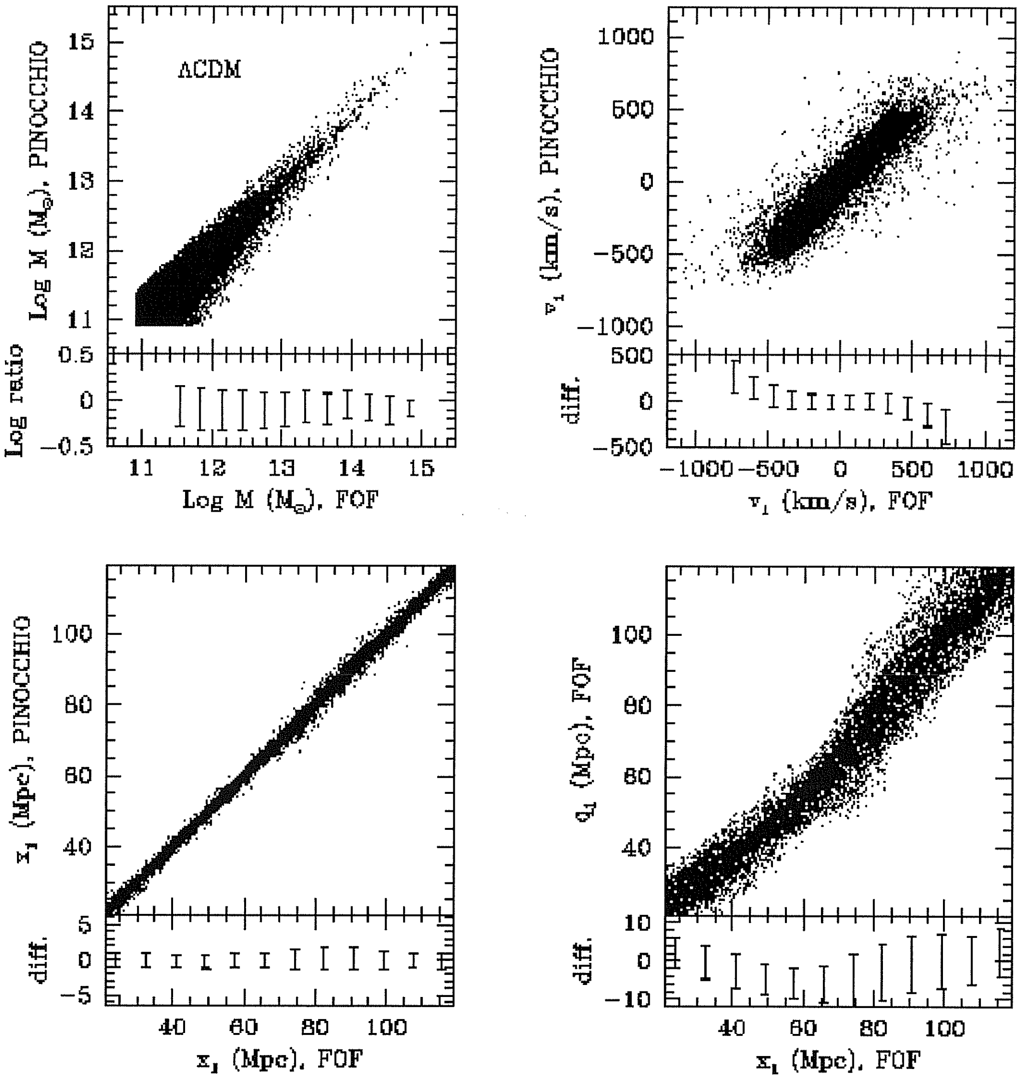

Figure 5.7: (continued)

5.1.3 Resolution effects

As discussed above, PINOCCHIO haloes resemble the FOF ones closely if they possess a minimum number of particles of around 30-100. Statistical quantities are well reproduced for haloes with at least 30-50 particles. These limits are comfortably similar to the minimum number of particles needed by a *simulation* to produce reliable haloes. To show this we plot in Fig. (5.8), for a random set of particles, the mass fields (i.e. the mass of the halo the particle belongs to) as determined by the 128^3 or 256^3 ΛCDM runs, both for the simulations and for PINOCCHIO. The result is shown at $z = 0$. There is considerable scatter between the masses of the haloes determined from simulations with different resolutions. This scatter is less than between PINOCCHIO and simulations, but not by much. This result is similar at higher redshifts. More details are given in §5.2.1, where it is shown that the match of the ΛCDM and ΛCDM128 halo catalogues shows a drop in the number of clean assignments for haloes smaller than $\sim$30 particles (Fig. [5.9a]), very similar to that shown in Fig. (5.6).

This result suggests that resolution affects PINOCCHIO in a similar way as it does numerical simulations. Better resolution leads to increased scatter in the identification of haloes, since the structures become more non-linear. For instance, we have verified that more massive haloes are reconstructed slightly *better* by the 128^3 PINOCCHIO run than by the 256^3 one. This is because at higher resolution, PINOCCHIO may decide to break-up a more massive halo in two. The degrading of the quality is modest and amounts to increased random noise which does not bias significantly the statistics of the haloes.

5.2 Angular momentum of the DM haloes

Haloes are thought to acquire their angular momentum from tidal torques exerted by the large-scale shear field while they are still in the mildly non-linear regime (Hoyle 1949; Peebles 1969; White 1984; Barnes & Efstathiou 1987; Heavens & Peacock 1988). In this hypothesis it is possible to estimate the angular momentum of haloes using the Zel'dovich (1970) approximation (Peacock & Heavens 1985; Catelan & Theuns 1996a) or higher-order LPT (Catelan & Theuns 1996b). The

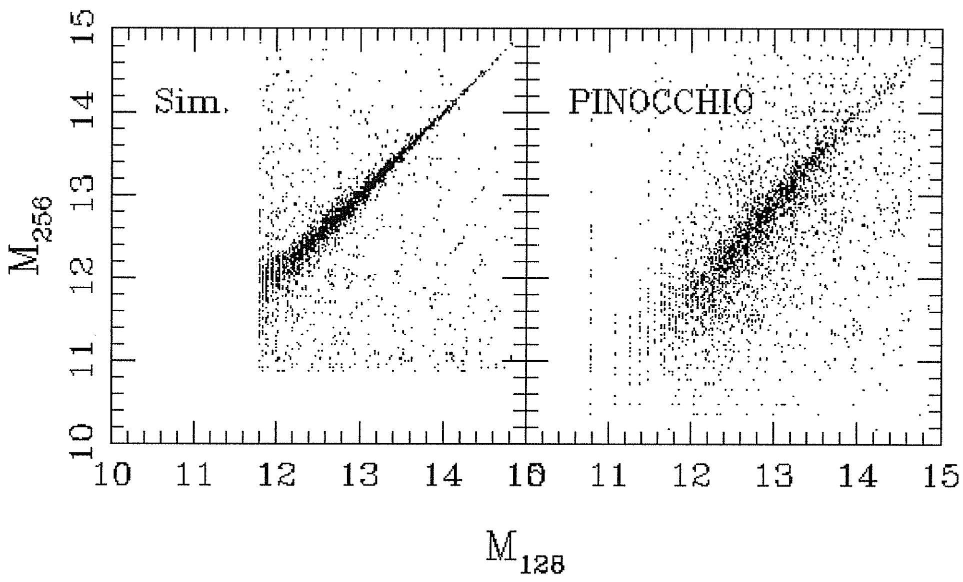

Figure 5.8: The effects of numerical resolution for simulations and for PINOCCHIO. Halo masses for a random set of particles from the 256^3 ΛCDM realisation as determined from the simulation and by PINOCCHIO (left and right panels respectively) are compared with the masses from the ΛCDM128 simulation.

biggest difficulty in this calculation is to identify the Lagrangian patch that is going to become a halo.

However, it was recently shown by Porciani, Hoffmann & Dekel (2001a,b) that the Zel'dovich approximation is unable to give very accurate predictions of the spin of haloes, as the highly non-linear interactions of neighbouring haloes tend to randomize their spins. Assuming to know exactly which particles are going to flow into a halo at $z = 0$ and using the Zeld'ovich approximation to compute the large-scale shear field, Porciani et al. (2001a) were able to recover the final angular momentum of the DM haloes with an average alignment angle (defined as the angle between true and reconstructed spins) of $\geq 40°$. Their analysis highlights the difficulty in predicting a higher-order quantity such as the spin of DM haloes.

5.2.1 Reliable estimate of the angular momentum of a DM halo

The calculation of the angular momentum from N-body simulations is also a subject of debate. Before entering into details of how to use PINOCCHIO to evaluate the spin, we give some details of the analysis we have performed on our reference simulations.

The main problem is to decide which is the minimum number of particles we have to use to reliably estimate the spin. This matter can be addressed by using our ΛCDM and ΛCDM128 simulations [1] . We consider the $z = 0$ outputs of the two simulations and match the halo catalogues in the same way it was done in the object-by-object comparison of PINOCCHIO and N-body catalogues. In practice, the 256^3 linking list is resampled to 128^3 by nearest grid assignment, i.e. simply by considering 1 particle over 8 and skipping the others.

Comment: This resampling is used only to match halo pairs, the halo properties are computed from the complete lists of particles. In the following we will assume the properties of the 256^3 groups as *bona fide* estimate, and will interpret the difference between 128^3 and 256^3 as the error on the lower-resolution groups.

---

[1] ΛCDM128 is run on the same initial conditions as ΛCDM, resampled on the coarser grid

Fig. (5.9) shows the fractions $f_{\rm cl}$, $f_{\rm split}$ and $f_{\rm ov}$ for the matching of the two catalogues, as a function of the mass of the halo according to the ΛCDM128 simulation. In this and subsequent panels the vertical line marks the groups of 100 particles (128^3). The matching of the two catalogues is excellent for groups larger than 100 particles, and still reasonable for groups as small as $\sim$30 particles. Mass estimates are pretty stable (Fig. [5.9b]), with an error of 30-40 per cent for the smallest groups, decreasing to the high-mass end.

Conversely, the error on the spin estimate turns out to be much larger. Fig. (5.9c) shows the fractional difference between halo spins as a function of mass (the rms difference is also shown), while Fig. (5.9d) shows the alignment angles of the spins (the rms of the mean is shown in this case, as in Fig. (5.12). The rms difference is still in excess of a a factor of two for haloes of 100 particles, and even larger for smaller haloes. Moreover, the spin directions of small haloes are very poorly correlated for haloes with less than 100 particles. We conclude that the lower limit for a correct order-of-magnitude estimate of the angular momentum of a halo is 100 particles, while a more precise estimate will require at least ten times more particles. We notice that this is at variance with other quantities, such as halo mass and velocity, that converge more rapidly. In the following we will restrict our analysis to groups larger than 100 particles.

5.2.2 The prediction of the angular momentum

With respect to the analysis of Porciani et al (2001a), the PINOCCHIO code presents the advantage of predicting with good accuracy the instant at which particles get into the halo, while the actual shape of the halo in the Lagrangian space is recovered with some noise, especially in the external borders that in fact contribute most to the angular momentum. We have verified that the direction of the largest axis of the inertia tensor of the haloes in the Lagrangian space is recovered within an alignment angle of $\sim$20°, while ellipticity and prolateness are correctly reproduced, although with much scatter.

The estimate of the angular momentum of haloes is easily performed within the fragmentation code, with negligible impact on its speed. When two haloes with angular momenta $\mathbf{L}_1$ and $\mathbf{L}_2$ merge, the spin $\mathbf{L}_{\rm merg}$ of the merger is estimated

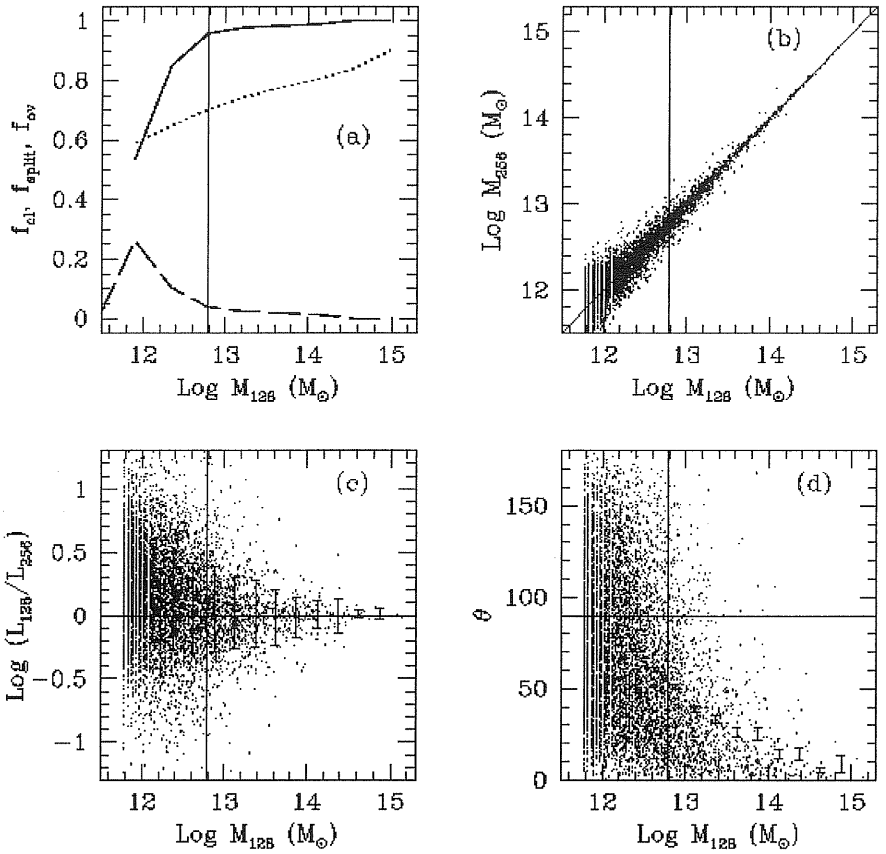

Figure 5.9: In all the panels the vertical line marks groups of 100 particles in the ΛCDM128 simulation. (a) Matching of the ΛCDM and ΛCDM128 halo catalogues. Continuous line: f_{cl}; dashed line: f_{split}; dotted line: f_{ov}. (b) Correlation of masses for the cleanly assigned objects.(c) Fractional difference of angular momenta for the cleanly assigned objects, as a function of mass. Error bars give the rms difference in bins of mass. (d) Alignment angle between the angular momenta of cleanly assigned objects. Error bars give the rms of the mean of the alignment angles in bins of mass.

as:

$$\mathbf{L}_{\rm merg} = \mathbf{L}_1 + \mathbf{L}_2 + \mathbf{L}_{\rm orb}, \tag{5.1}$$

where $\mathbf{L}_{\rm orb}$ is the orbital angular momentum of the two haloes:

$$\mathbf{L}_{\rm orb} = M_1(\Delta\mathbf{q}_1 \times \Delta\mathbf{v}_1) + M_2(\Delta\mathbf{q}_2 \times \Delta\mathbf{v}_2). \tag{5.2}$$

Here $\Delta\mathbf{q}_i \equiv \mathbf{q}_i - \mathbf{q}_{\rm cm}$, $\Delta\mathbf{v}_i \equiv \mathbf{v}_i - \mathbf{v}_{\rm cm}$, with $i = 1, 2$, $\mathbf{q}_{\rm cm}$ and $\mathbf{v}_{\rm cm}$ the position and velocity of the centre of mass. It is worth noticing that the use of Lagrangian coordinates $\mathbf{q}$ is justified by the parallelism of displacements and velocities. Following Catelan & Theuns (1997a), we stop the linear growth of velocities not at the time of merger $t_{\rm merge}$ but at the time $t_{\rm grow}$ defined as:

$$t_{\rm grow} = 0.5 t_{\rm merge} \tag{5.3}$$

where t is physical time. This is a suitable generalisation of the concept of 'detaching' of the perturbation from the Hubble flow. The case of accretion is treated as a merger with a 1-particle halo which carries zero spin.

The so-obtained angular momenta obey a mass–spin relation which is roughly consistent with that of the FOF groups. This is shown in the left panels of Fig. (5.10) for the ΛCDM simulation. Although qualitatively similar, the PINOCCHIO relation overestimates the FOF one by some factor which is larger for the smaller haloes. If the lower value of the spin is due to the higher degree of non-linear shuffling suffered by haloes because of tidal interaction with neighbours, this trend of having lower-mass haloes more randomized than higher-mass ones is in agreement with what suggested by Porciani et al. (2001a).

It is useful to improve this prediction, so as to obtain angular momenta for the haloes with accurate statistical properties. To this aim we decrease each component of the spin at random, following the simple rule:

$$L_i^{\rm new} = L_i \times [(1 - f_{\rm spin}) + f_{\rm spin} \times f_{\rm rand}], \tag{5.4}$$

where $f_{\rm spin} = f_0 + f_1(M/M_*(z))$ (forced to $0 \leq f_{\rm spin} \leq 1$) and $f_{\rm rand}$ is a random number ($0 < f_{\rm rand} < 1$). The two parameters f_0 and f_1 are fixed so as to reproduce at best the mass–spin relation of Fig. (5.10). Optimal values are $f_0 = 0.8$ and $f_1 = 0.15$. The right panels of Fig. (5.10) show the resulting mass–spin relations, which agrees fairly well with the FOF ones.

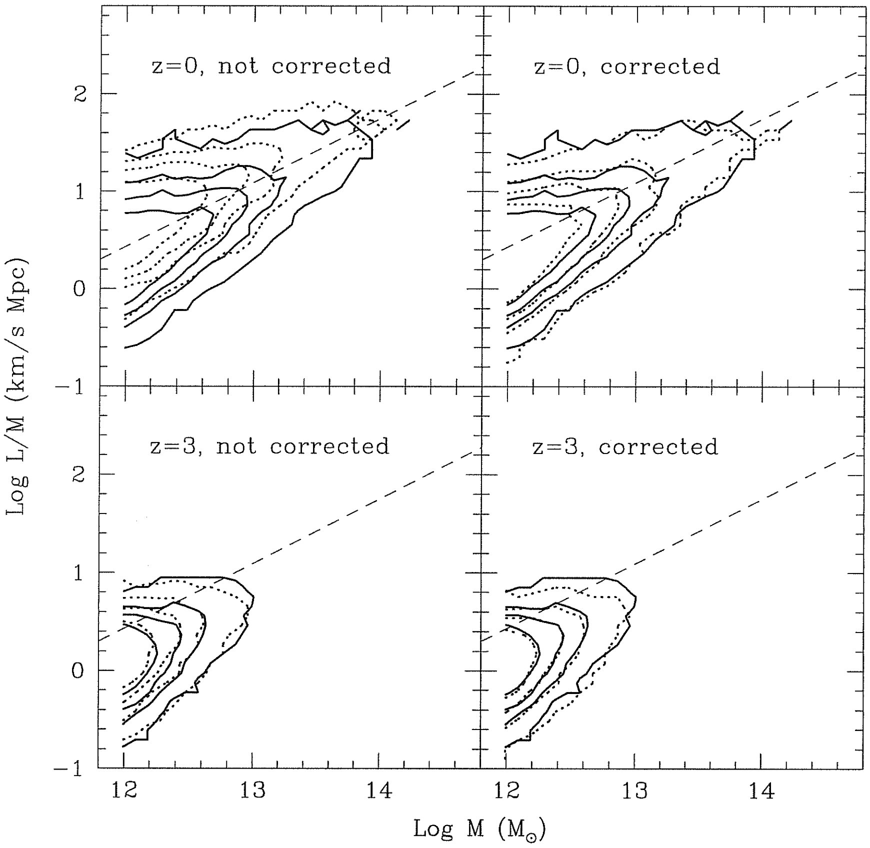

Figure 5.10: Mass–spin relation for dark-matter haloes. Contour lines trace the levels of 0.2, 1, 2 and 4 haloes per decade in $\log M$ ($M_\odot$) and $\log L/M$ (km/s Mpc, physical units). Continuous and dotted lines show the contours for FOF and PINOCCHIO haloes respectively. Dashed lines give the scaling $L \propto M^{5/3}$ (Catelan & Theuns 1996a). Upper panels: $z = 0$; lower panels: $z = 3$. Left panels: no correction; right panels: spin corrected as in eq. (5.4).

Apart from the mass–spin correlation shown in Fig. (5.10), the angular momentum is known to be nearly independent of other halo properties (Ueda et al. 1994; Cole & Lacey 1996; Nagashima & Gouda 1998; Lemson & Kauffman 1999; Bullock et al. 2001; Gardner 2001; Antonuccio-Delogu et al. 2001), with the exception of a weak dependence with the merger history of the haloes. The dependence of spin on the environment is still debated (Lemson & Kauffman 1999; Antonuccio-Delogu et al. 2001). Gardner (2001) has shown that haloes that have suffered a major merger tend to have higher spin. In Fig. (5.11) we show that this trend is successfully reproduced by PINOCCHIO haloes. Merged haloes at $z = 0$ have been selected by requiring that the second largest progenitor halo at $z = 0.25$ is larger than 0.3 times the final halo mass. To extract the mass–spin relation, we define the quantity $\lambda \equiv LogL - 1.5(LogM/M_*)$. As apparent in Fig. (5.11), the λ-distribution of the merged haloes is biased toward larger λ-values both for the simulation and for the PINOCCHIO haloes, although the trend may be slightly underestimated by PINOCCHIO.

The agreement at the object-by-object level is in line with the intrinsic limits of perturbative theories found by Porciani et al. (2001a). Fig. (5.12) shows the alignment angle θ for the spins of cleanly matched FOF and PINOCCHIO haloes, and their average values computed in bins of mass (error flags indicate the rms around the mean). While the left panel shows all haloes, the right panel is restricted to those pairs of haloes that overlap by more than 70 per cent. The average angle is significantly smaller than 90°, highlighting a significant correlation of PINOCCHIO and FOF spins. However, the alignment is at best as high as ~60°. This is mostly due to errors in the definition of the halo, as shown by the right panel, where the best reconstructed haloes with more than 1000 particles show an average alignment angle of ~30–40°, consistent with the intrinsic limit quoted by Porciani et al. (2001a).

To conclude, the prediction of angular momentum of haloes is severely hampered by the intrinsic limits of linear theory described by Porciani et al. (2001a) and further worsened by the error made by PINOCCHIO in assigning particles to haloes. The correct statistics is reproduced only by introducing two more ‘fudge’ free parameters, while the object-by-object agreement is poor although significant. However, even N-body simulations do not converge rapidly in estimating this quantity (see 5.2.1). Moreover, the important spin–merger correlation is recovered naturally. Although we do not claim this result as a big success, we notice that PINOCCHIO is, to our knowledge, the only perturbative algorithm able to predict the spin of haloes at the object-by-object level. Moreover, the prediction of spins comes at almost no additional computational cost, and the whole acquisition history of angular momentum can be followed for each halo. Thus, we regard the use of the angular momenta provided by PINOCCHIO as a viable alternative to drawing them at random from some distribution that fits N-body simulations (Cole et at. 2000; Vitvitsaka et al. 2001; Maller, Dekel & Somerville 2001).

5.3 Discussion and conclusions

PINOCCHIO is an approximation to the full non-linear gravitational problem of hierarchical structure formation in a cosmological setting, in contrast to the mostly statistical approaches such as the PS prescription. The good agreement in detail between PINOCCHIO and FOF haloes identified in simulations, explains the ability of the method to generate reliable halo càtalogues. It also demonstrates that the underlying dynamical approximations work well. With respect to the results of Monaco (1995; 1997a,b), PINOCCHIO addresses successfully the geometrical problem of the fragmentation of the collapsed medium into objects and filaments.

While a direct analytical rendering of the fragmentation prescription as used in PINOCCHIO seems very complex, because it requires knowledge of spatial correlations to high order, analytical progress might nevertheless be possible. For instance, Monaco & Murante (1998) proposed to generalise the mass-radius relation of PS, to allow a more general distribution of masses to form at a given smoothing radius. This was formulated in terms of a 'growing' curve for the objects, that gives the fraction of mass acquired by the object at a given smoothing radius. The mass function is then obtained by a deconvolution of the $\Omega(<\sigma^2)$ function (as obtained from ELL collapse, like in Fig. [4.3]) with the growing curve of the objects. This growing curve could be estimated from the results of PINOCCHIO, giving an improved analytical expression for the mass function. But in the case of Gaussian smoothing merging histories cannot be computed from the excursion set formalism, because the trajectories are strongly correlated (Peacock

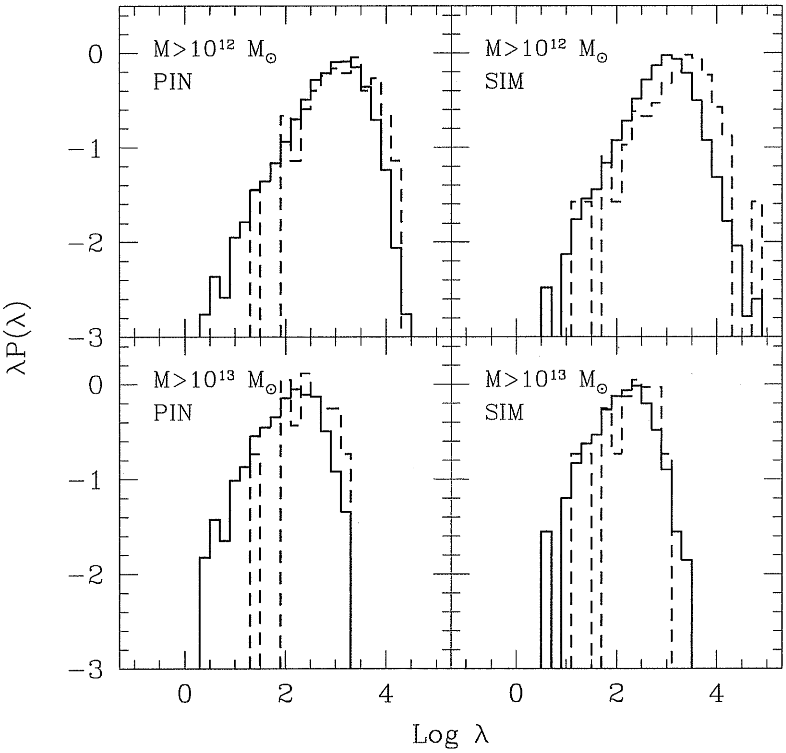

Figure 5.11: Correlation between spin and merging history. Continuous lines: all haloes; dashed lines: haloes that have suffered a major merger. Left panels: PINOCCHIO predictions; right panels: simulation. Upper panels: haloes with $M \geq 10^{12}\ M_\odot$; lower panels: haloes with $M \geq 10^{13}\ M_\odot$.

& Heavens 1990; Bond et al. 1991), so that the random walk formalism cannot be used. Moreover, it is impossible from such an approach to have full information on the spatial distribution of objects. So, such analytic extensions of PINOCCHIO would not be as powerful as the full analysis. Besides, analytic formalisms based on peaks (Manrique & Salvador Sole 1995; Hanami 1999) are manageable only when linear theory is used. We therefore regard methods like PINOCCHIO which are based on an actual realisation of the linear density field, as a good compromise between performing a simulation, and getting only statistical information from a EPS–like approximation.

Similar methods have been proposed in the literature, such as the peak-patch method of Bond & Myers (1996a), the block model of Cole & Kaiser (1988), and the merging cell model of Rodrigues & Thomas (1996) and Lanzoni et al. (2000). A qualitative comparison with peak-patch reveals a similar accuracy in reproducing the masses of the objects. From Fig. (10) of Bond & Myers (1996b) it is apparent that, in a context analogous to our SCDM simulation, masses are recovered with an accuracy of ∼0.2 dex, comparable or slightly worse than the one given in Fig. (5.7) for SCDM. Unfortunately, it is not clear from the Bond & Myers papers in which range of halo masses their method can be successfully applied. As linear theory under predicts the fraction of collapsed mass when the variance is large, a deficit of peaks corresponding to smaller masses is possible. The objects selected by peak-patch are constrained to be spherical in the Lagrangian space (they collapse like ellipsoids but start-off as spheres perturbed by the tidal field), while PINOCCHIO is not restricted in this sense and is able to reproduce the orientation of the objects in the Lagrangian space. Moreover, PINOCCHIO is not affected by the problem of peaks overlapping in the Lagrangian space. Finally, peak-patch has never been extended, to the best of our knowledge, to predict the merger histories of objects.

The merging cell model of Lanzoni et al. (2000) shares some properties with PINOCCHIO, in particular the fact that both codes build-up haloes through mergers and accretion. However, the non-linear ellipsoidal collapse of PINOCCHIO is an important improvement, as is the use of Gaussian filters instead of box car smoothing. The size of the merging objects tend to be quite large in the merging cell model, whereas PINOCCHIO allows accretion of single particles. We have

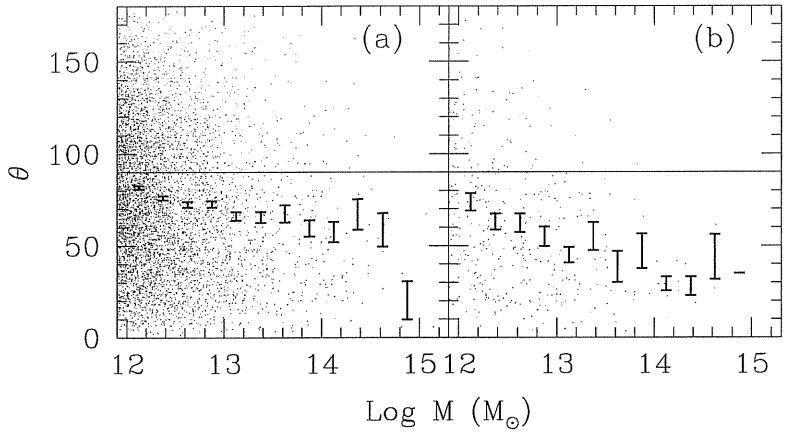

Figure 5.12: Alignment angle θ for pairs of cleanly assigned haloes at $z = 0$ as a function of mass. The errorbars denote averages in mass bins, errorbars give the rms of the mean. (a) all haloes, (b) haloes with $f_{\rm ov} > 0.7$.

been able to compare our results directly with those of Lanzoni et al. (2000). The haloes identified in the merging cell model do not accurately reproduce those from the simulations. This poorer level of agreement is partly due to the cubic shape of the cells and to the coarse resolution of the box car smoothing. As a consequence of these choices, massive haloes appear as big square boxes, and the mass function shows fluctuations of a factor of two which reflects the smoothing.

Chapter 6

PINOCCHIO: the hierarchical build-up of dark matter haloes

> *At this third lie, his nose became longer than ever,*
> *so long that he could not even turn around.*
>
> C. Collodi

The final goal of a numerical or semi-analytical code developed to study the structure formation process, is to trace the history of individual haloes. In §3, we presented a "history procedure" based on the EPS formalism which can be used to Monte Carlo generate synthetic catalogues of haloes. This semi-analytical approach is commonly applied to study in details merging histories, however, it suffers of various limitations, for example it does not directly provide spatial information for the haloes. A more complete description of the halo evolution is given by PINOCCHIO. In the previous Chapters we showed that this code produces catalogues of objects with known mass, position, and velocity. We need then to investigate if it is possible to use PINOCCHIO to trace the merger tree of individual haloes.

In this Chapter we describe how to evaluate the merging histories from PINOC-

CHIO (§6.1) and we compare its results both with N-body simulations and the MCMT presented in §3. In §6.2 we examine the ability of PINOCCHIO to reconstruct the main statistical properties of the merger trees, and in §6.3 we extend the analysis to the prediction of the correlation function.

6.1 Merger trees from PINOCCHIO

The merger histories of haloes are directly evaluated 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 we keep track at all times of the parent which incorporate it. Even though accretion is rigorously defined 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.

Comment: When using PINOCCHIO to evaluate the merging history of haloes we do not only follow the time evolution of the mass and number distribution of the progenitors, but (in contrast with the EPS–based merger tree codes) also of their distribution in space, their velocities and angular momenta.

6.2 Statistics of the progenitors

6.2.1 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 §4, for all the details we refer to Table (4.1).

Comment: The ΛCDM simulation is more suitable to reconstruct the merger tree to higher redshifts and lower masses thanks to the higher mass resolution, while the SCDM allows to test the merger trees for the more massive haloes.

To test the ability of PINOCCHIO in reproducing the conditional mass function, we adopt 10 particles as the minimum mass for the progenitor haloes. This is to check the agreement of the low mass tail of the distribution which is not reproduced by the EPS based codes (see §3.4.2). 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 described in §3.4.1.

One of the main features of PINOCCHIO is that we can 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 by far less expensive (in term of CPU time, disk space and human labour) in PINOCCHIO than in a simulation.

6.2.2 Progenitor mass function

The progenitor mass function $dN(M, z|M_0, z_0)/dM$ (see §3.1.1), 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 EPS formalism is given by eq. (3.2).

To identify the parent mass for both 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.

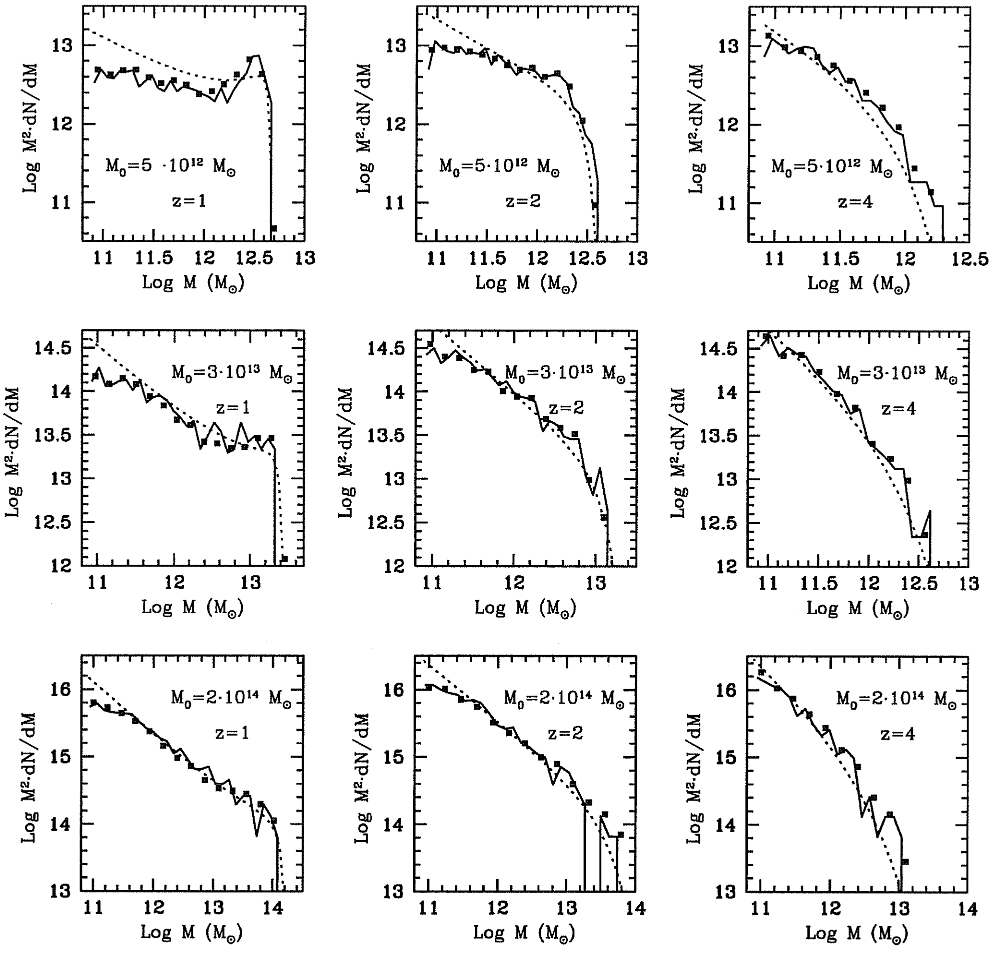

Figure 6.1: Conditional mass functions in the ΛCDM case for parent haloes identified at $z = 0$. The mass threshold is fixed at $M_{\rm th} = 7.6 \times 10^{10}$ M$_\odot$ (10 particles), the redshift increases from left to right and covers the values: $z = 1,\ 2,\ 4$. The mass of the parent halo increases from top to bottom, the adopted values are: $M_0 = 5. \times 10^{12} M_\odot$, $3. \times 10^{13} M_\odot$ and $2.0 \times 10^{14} M_\odot$. 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.

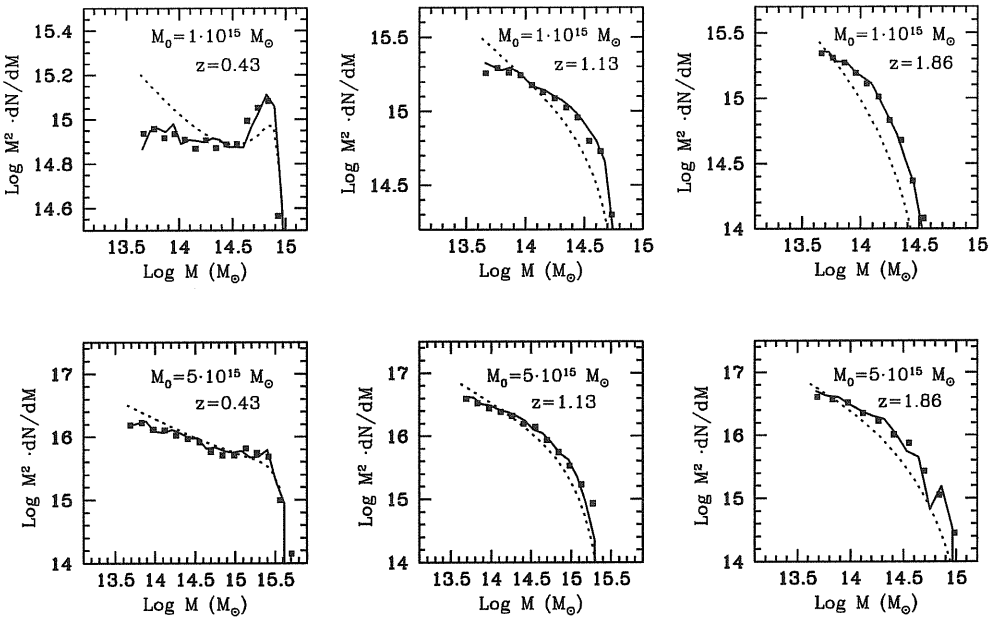

Figure 6.2: Same as in Fig. (6.1) but for the SCDM case. The mass threshold is $M_{\rm th} = 1.49 \times 10^{13} M_\odot$ (10 particles).

In Fig. (6.1) and Fig. (6.2) we compare the conditional mass functions obtained from PINOCCHIO and the simulations for the ΛCDM and the SCDM case, respectively. The bottom panels of Fig. (6.1) show the results for a cluster-sized parent of $M_0 = 2 \times 10^{14}$ M$_\odot$, the case of haloes corresponding to small groups ($M_0 = 3 \times 10^{13}$ M$_\odot$) and galaxies ($M_0 = 5 \times 10^{12}$ M$_\odot$) are presented in the mid and upper panels. In Fig. (6.2) we show the results for parents with mass comparable to massive clusters ($M_0 = 1 \times 10^{15}$ M$_\odot$ and $M_0 = 5 \times 10^{15}$ M$_\odot$). 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) presents a very good agreement when compared with the simulations. In Fig. (6.1) and Fig. (6.2) we show that the PINOCCHIO prediction fits the simulation data with similar accuracy for all the considered parent masses and redshifts and we identify a discrepancy between the two distribution which in general is

$\lesssim 20$ per cent.

A similar analysis was performed for the MCMT code (§3). In Fig. (3.6) we show the results obtained with MCMT for the same ΛCDM simulation, the same halo masses and redshifts.

Comparing the results of Fig. (6.1) and Fig. (3.6) we notice that PINOCCHIO reproduces the N-body data better than MCMT.

6.2.3 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 before. The histograms on Fig. (6.3) and Fig. (6.4) show the distribution of the mass of the larger progenitor normalised 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 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.

As already noticed in §3.4.2, the distribution of M_1/M_0 also provides a hint on the formation time of the parent. The good agreement of PINOCCHIO with the simulations can thus be also extended to the halo formation times. For instance Fig. (6.4) suggests that, in this SCDM cosmology, a halo of $1 \times 10^{15} M_{\odot}$ forms at $z \sim 0.43$ or later.

In the upper part of the plots of Fig. (6.3) and Fig. (6.4) the distribution of M_2/M_1 (the ratio of the second largest progenitor to the largest ones) given M_1/M_0 is shown. The points are the mean value of the distribution and the error bars are the 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.

These figures can be compared with Fig. (3.7), where we test the behaviour of the MCMT code. In this case PINOCCHIO does not improve with respect to our EPS–based merger tree code. In fact, the EPS formalism (and consequently the EPS-based semi-analytical codes) reproduces with good accuracy the more

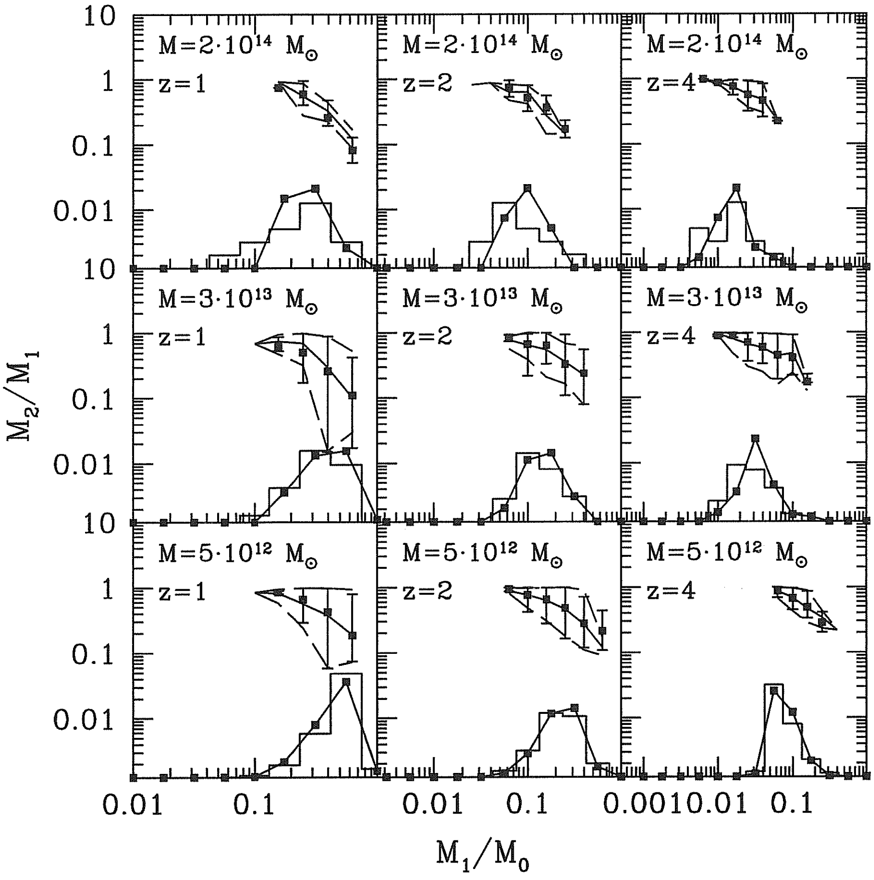

Figure 6.3: Distribution of the mass of the largest progenitor M_1 for the ΛCDM case with mass threshold $M_{\rm th} = 2.3 \times 10^{11} M_\odot$ (30 particles). The histograms are the PINOCCHIO predictions and the points connected with solid lines are the simulations'. The quantity plotted on the upper 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 is the PINOCCHIO result and the dashed lines show its 1σ variance. The points with error bar are the simulation data.

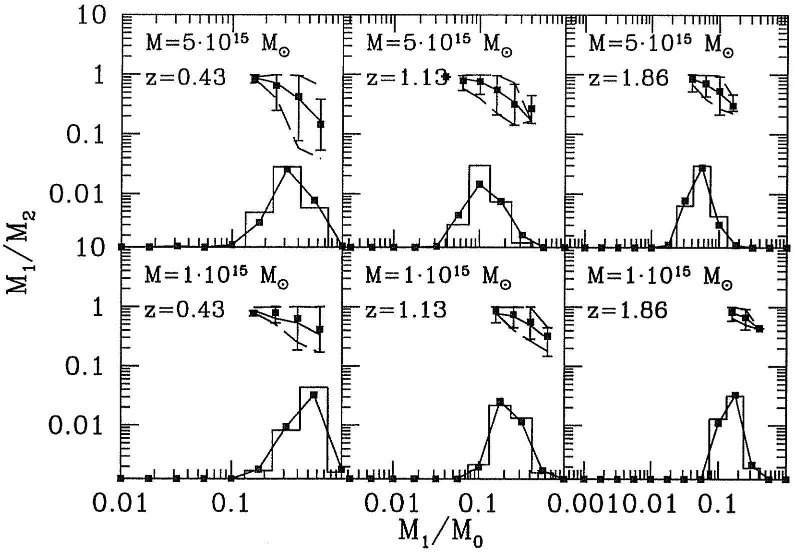

Figure 6.4: As Fig. (6.3) but for the SCDM case. The mass threshold is $M_{\rm th} = 1.3 \times 10^{14} M_\odot$ (30 particles).

massive progenitors, but fails in the low mass tail of the distribution (Lacey & Coles 1994, Somerville & Kolatt 1999, Sheth & Lemson 1999, Coles et al. 2000).

6.2.4 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 Fig. (6.5) and Fig. (6.6) 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 normalisation) the integral of the conditional mass 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 reproduces fairly well the distributions also for the more massive haloes and at all redshifts.

In Fig. (6.7) we show the average μ_1 and the rescaled variance μ_2/μ_1 as a func-

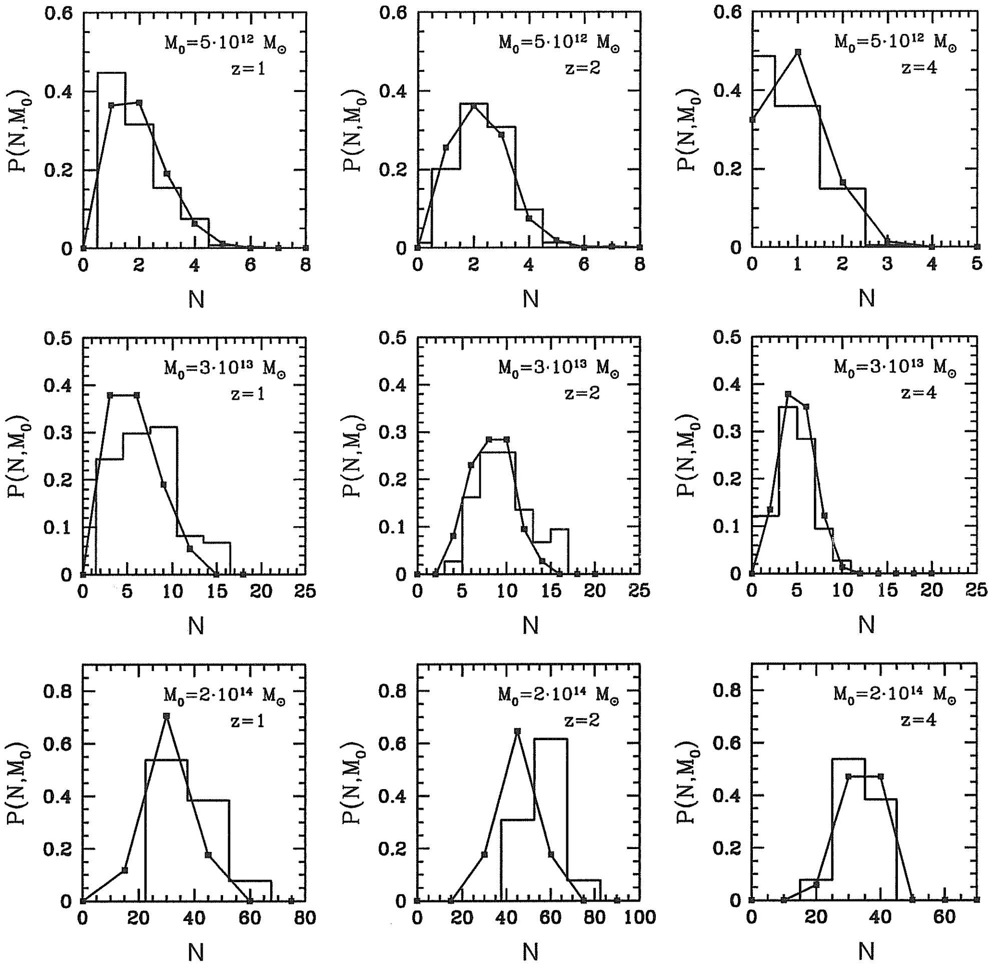

Figure 6.5: Probability that an halo M_0 at $z = 0$ has N progenitors for the ΛCDM case. The threshold mass is $M_{\rm th} = 2.3 \times 10^{11} M_\odot$ (30 particles). The points connected with solid lines represent the simulation data while the histograms are the prediction of PINOCCHIO.

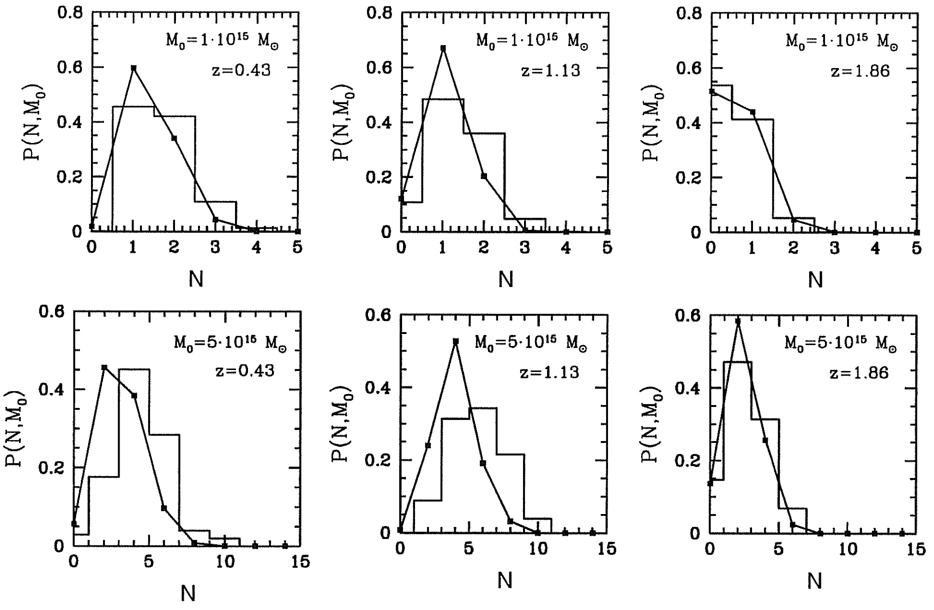

Figure 6.6: Same as in Fig. (6.5) for the SCDM case. The threshold mass is $M_{\rm th} = 1.3 \times 10^{14} M_\odot$ (30 particles).

tion of the parent halo mass for different redshifts. The lines are the PINOCCHIO prediction 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 eq. (3.2). The agreement between PINOCCHIO and the simulations varies from the 5 per cent of the ΛCDM to the 10 per cent of the SCDM case but it does not depend on the redshift.

Again PINOCCHIO is found to improve with respect to the EPS analytical prediction. In particular, at low redshift the EPS predictions underestimate the mean value by a factor that ranges from 20 to 30 per cent.

On the other hand, the MCMT is not able to reproduce the statistics of the number of progenitors: as shown in Fig. (3.8) and in Fig. (3.9) it always over predict the average number of haloes.

Our results can be also compared to those shown by Sheth & Lemson (1999b) and Somerville et al. (2000) for EPS–based merger trees and with Sheth & Tormen (2001) who elaborate an excursion set model based on ellipsoidal collapse.

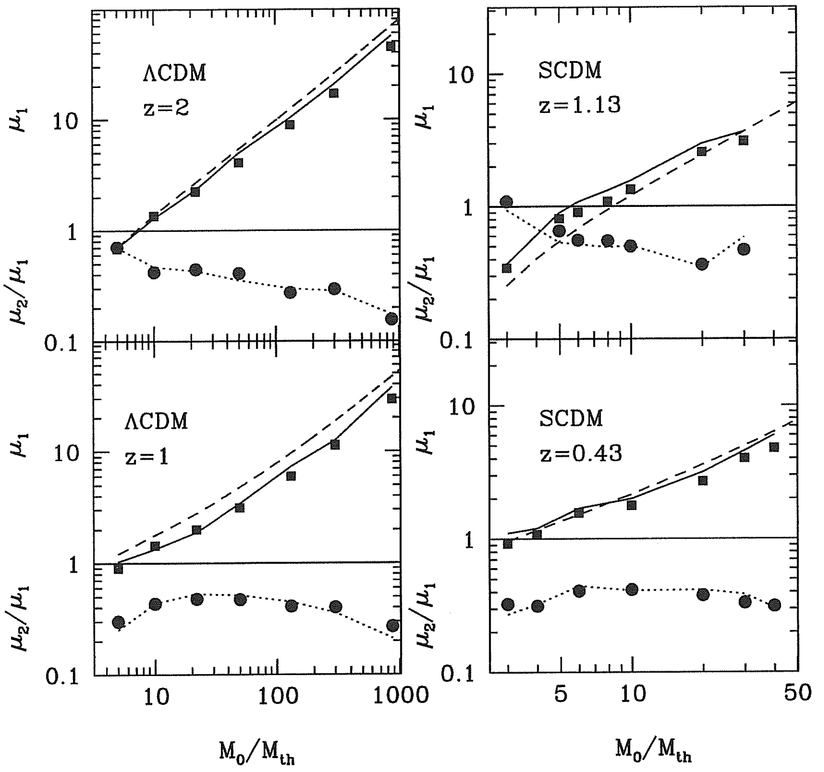

Figure 6.7: First two moments of the distribution of the number of progenitors $P(N, M_0)$ as a function of the parent mass M_0. Lines are the PINOCCHIO results and symbols are the simulation data. The left plots show the ΛCDM case at redshift $z = 1$ and 2. The threshold mass is $M_{\rm th} = 2.3 \times 10^{11} M_\odot$ (30 particles) and we plot the mean (squares and solid line) and the rescaled variance (circles and dotted line) up to $M_0 = 1000 M_{\rm th}$. The dashed line is the EPS analytical prediction for the mean. The right plots show the SCDM case at redshift z=0.43 and z=1.13. The threshold mass is $M_{\rm th} = 1.3 \times 10^{14} M_\odot$ (30 particles), and we plot the mean and the rescaled variance up to $M_0 = 50 M_{\rm th}$.

In general, PINOCCHIO reproduces the statistical properties of progenitor distributions with better accuracy then 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.

6.2.5 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 cleanly reconstructed. In §5.1.2 a pair of haloes coming 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. 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. (6.8) we show, for the parents that are cleanly identified, the fraction in number $f_{\rm p}$ 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 between 80 and 90 per cent. The fraction $f_{\rm p}$ is in general higher at higher redshift, when the object-by-object agreement between PINOCCHIO and the simulation is better. Haloes with larger mass tend to be reconstructed with worse accuracy, especially at $z = 1$. This is mainly due to the progenitors which carry a mass smaller then $\sim$20 per cent of M_0 (see right panels of Fig. [6.8]). The progenitors which carry a mass of less that $\sim$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, especially present in the larger haloes, are reconstructed in a noisier way.

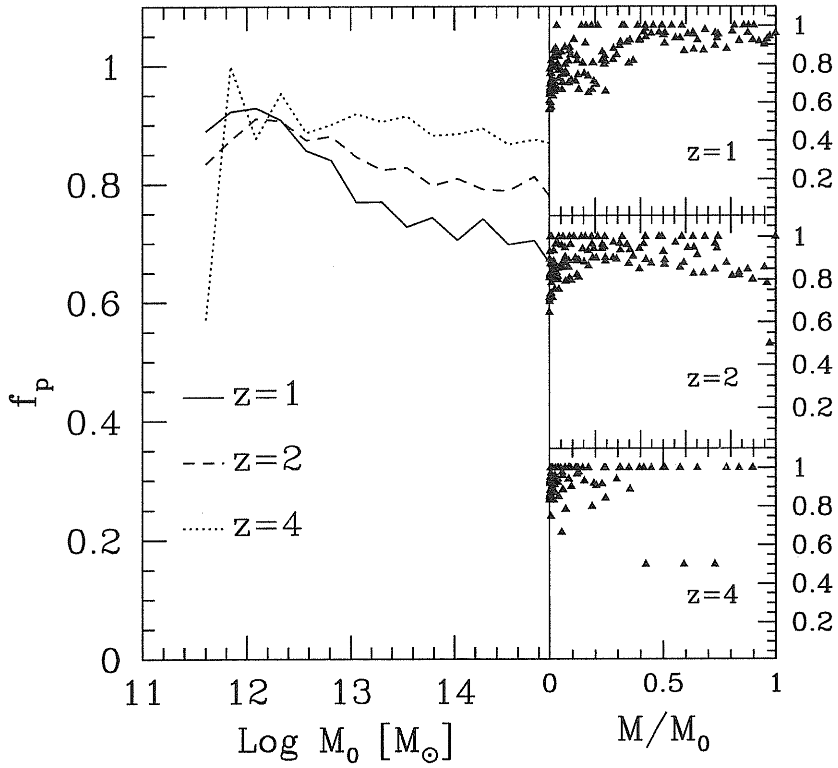

Figure 6.8: Fraction f_{p} of cleanly assigned progenitors for redshift $z = 1$, 2 and 4. The left hand side represents the fraction of cleanly assigned progenitors as function of the parent mass. The three lines correspond to the different redshifts. The tree plots on the right are the scatter plots of f_{p} as function of the progenitor mass M normalised to the parent mass M_0, for the parent haloes of mass 10^{11} $\mathrm{M}_{\odot} < M_0 < 10^{15}$ $\mathrm{M}_{\odot}$; the redshift increases from top to bottom.

6.3 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 $\gtrsim$40 per cent (Mo & White 1996; Jing 1998; Porciani, Catelan & Lacey 1999; Sheth, Mo & Tormen 2000; Colberg et al. 2001). In this approach it is not possible to know how the bias changes for haloes with different merger histories. This piece of information is precious to produce predictions on the bias of galaxies of different types, that typically have different merger histories.

As shown in the previous Chapter, PINOCCHIO haloes have the same correlation length r_0 as FOF haloes to within 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 the two-point correlation functions for their progenitors. In Fig. (6.9) the solid lines represent the two-point correlation function of progenitors, $\xi_{\mathrm{p}}(r)$, evaluated in PINOCCHIO compared with the same quantity measured in the simulation. The plots show that PINOCCHIO reproduces such correlation functions to within $\sim$20 per cent error.

We also compare this function with the average correlation function, $\xi_{\mathrm{h}}(r)$, at the same redshifts. It is apparent that PINOCCHIO reproduces correctly the larger clustering amplitude of haloes that flow into cluster sized one. The bias between the two halo populations is defined as: $b^2(r, z) = \xi_{\mathrm{p}}(r)/\xi_{\mathrm{h}}(r)$. We compare in the bottom row of plots in Fig. (6.9) the bias measured in the simulation with PINOCCHIO results. The bias is recovered to within $\sim$20 per cent and the scale dependence is correctly reproduced.

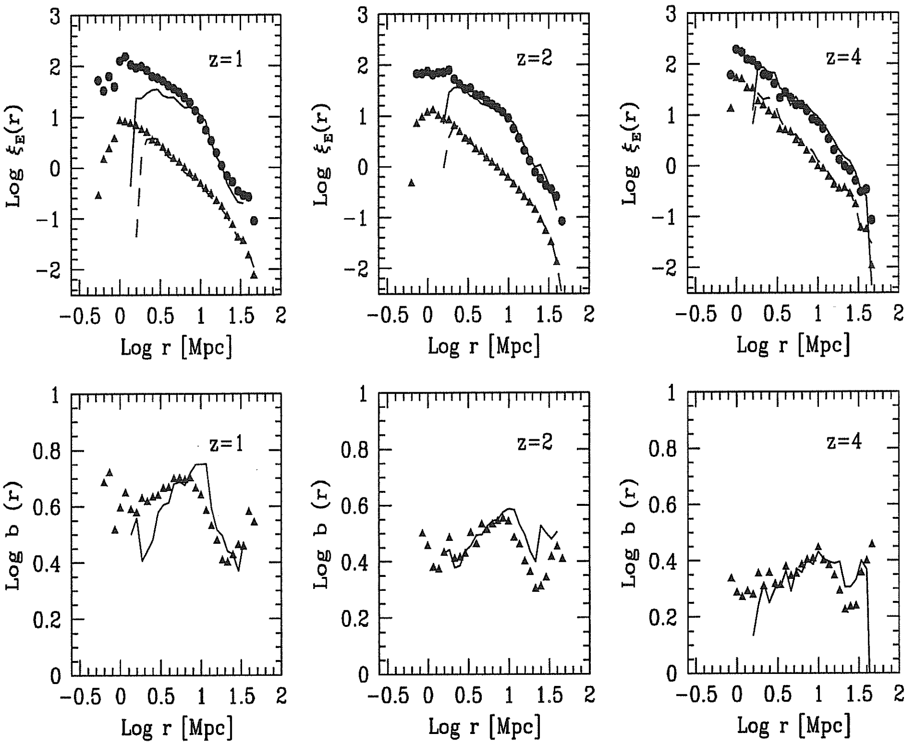

Figure 6.9: Top panels: correlation function for the progenitors of haloes with a mass greater than 10^{14} $M_{\odot}$ at z=0, ξ_{p} (circles and solid lines), and for all haloes larger than the threshold mass $M_{\mathrm{tr}} = 2.3\,10^{11} M_{\odot}$, ξ_{h} (triangles and dashed lines), at various redshift indicated in the panels. Symbols refer to FOF selected haloes obtained from the ΛCDM simulation, lines to the corresponding PINOCCHIO prediction. Bottom panels: bias $\xi_{\mathrm{p}}/\xi_{\mathrm{h}}$ for simulations (symbols) and PINOCCHIO (lines). PINOCCHIO is able to predict accurately the clustering of the selected halo types.

6.4 Conclusions

We have presented a novel procedure to study the merging history of haloes using the PINOCCHIO code and we have compared the results of PINOCCHIO with those of two large N-body simulations (ΛCDM and SCDM cosmologies) and with the MCMT code presented in §3.

The merger histories of the PINOCCHIO haloes resemble closely those found 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 10 particles). In particular statistical quantities like 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 $\sim$20 per cent typical error.

Those results should be compared with the EPS analytical predictions and the results of the EPS–based semi–analytical codes (Lacey & Cole 1993; Somerville & Kolatt 1999; Sheth & Lemson 1999a), in particular with MCMT presented in §3. The fit of the statistical quantities achieved by PINOCCHIO is much better than the estimates of MCMT, which show discrepancies up to a factor of 2 (see §3; Governato et al. 2000; Jenkins et al. 2001; Bode et al. 2001; Somerville et al. 1999; Sheth & Lemson 1999b, Cohn, Bagla & White, 2001).

In contrast to EPS formalism (Bond et al. 1991; White 1996) and consequently to MCMT, PINOCCHIO provides much more useful information on the haloes, such as positions, velocities, angular momenta. For example we can evaluate the correlation function of haloes that are progenitors of lower-redshift massive ones. This function is correctly reproduced to within an accuracy of $\sim$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.

Finally we noticed that the validity of PINOCCHIO extends to the object-by-object level: PINOCCHIO cleanly reproduces $\gtrsim$70 per cent of the progenitors when parent haloes are cleanly recognised themselves and this agreement slowly degrades with time.

Chapter 7

Dynamical evolution of DM satellites

The more I know, the less I understand.

G. Stewart

In the hierarchical clustering scenario a merging event between two or more haloes corresponds to the loss of identity of the single primitive units to form a new one. Moreover, after the encounter the haloes may retain their identity, and become substructures of the new system. This picture is confirmed by high-resolution N-body simulations (Moore, Katz & Lake 1996; Tormen 1997; Ghigna et al. 1998; Tormen, Diaferio & Syer 1998). Indeed, this is in line with the evidence of the presence of galaxies within galaxy groups or clusters.

The study of the various issues arising from hierarchical clustering scenarios, including the distribution of satellites around galactic systems, the impact of these satellites on the bulge of the main halo, the disk heating and the merging events between the halo and the satellites, is a crucial and still poorly understood step in the study of galaxy formation and evolution (see eg. Tóth & Ostriker 1992; Huang & Carlberg 1997; Velázquez & White 1999; Moore et al. 1999; Bullock, Kravtsov & Weinberg 2000)

High-resolution simulations are necessary to describe the evolution of satellites, at the cost of simulating one cluster at a time (Tormen 1997; Ghigna et al.

1998). It is then useful to relay again on the analytic modeling, which can be easily coupled with semi-analytical codes for structures formation as MCMT or PINOCCHIO. In this and next Chapter we introduce a semi-analytical procedure to study the dynamical evolution of satellites which synthesizes and generalizes the results of existing numerical studies.

This Chapter is organized as follow. In the first two Sections we discuss the theoretical scheme underlying our model for following the orbital decay. In §7.1 we review the Chandrasekhar (1943) theory of dynamical friction and in §7.2 we introduce an alternative approach developed by Colpi (1998; see also Colpi & Pallavicini 1998). As the dynamical friction force depends on the density profile of the main halo, §7.3 discusses the satellites evolution for two different density profiles: the singular isothermal sphere and the Navarro, Frank & White (1996) profile. In §7.4, we provide a useful fitting formula to estimate the decay time of rigid satellites orbiting in a Navarro, Frank & White (1996) profile. As the decay time of a satellite strongely depend on the initial obital energy and angular momentum (or better the initial orbital parameters), an important aspect is how to predict their value in a Cosmological environment. The last Section deals with this topic.

7.1 The orbit of satellites

In our study of the orbital evolution of satellites, we assume that the main halo is spherically symmetric. The gravitational force is then centrally directed. As a consequence the orbit can be determined using the planar polar coordinates $r(t)$ and $\theta(t)$, solving for the equation of motion (see e.g. Binney & Tremain 1987). The motion of a satellite is specified by the initial angular momentum J and orbital energy E per unit mass, or equivalently by the radius $r_{\rm c}(E)$ of the circular orbit having the same energy E, and by the circularity $\epsilon = J/J_{\rm c}$, where $J_{\rm c}$ is the angular momentum of the circular obit $J_{\rm c} \equiv V_{\rm c}(r_{\rm c}) \cdot r_{\rm c}(E)$ ($V_{\rm c}[r_{\rm c}]$ is the circular velocity at radius $r_{\rm c}[E]$).

We define a generalised orbital eccentricity:

$$e = \frac{r_{\rm apo} - r_{\rm per}}{r_{\rm apo} + r_{\rm per}} , \tag{7.1}$$

here $r_{\rm apo}$ and $r_{\rm per}$ are the roots of the orbit equation:

$$\frac{1}{r^2} + \frac{2[\phi(r) - E]}{J^2} = 0 \ , \tag{7.2}$$

and they are respectively the apocenter and the pericenter radii of the orbit. The function $\phi(r)$ is the gravitational potential the satellite is orbiting in.

Using the previous equation it is possible to evaluate a well defined relation between e and the orbital parameters, so that to each value of $r_{\rm c}(E)$ and ϵ it corresponds a different apoastron and periastron radius. Hereafter we define $x_c(E) \equiv r_c(E)/R_{\rm h}$, where $R_{\rm h}$ is the virial radius of the halo.

Eq. (7.2) identifies the stable orbit of a satellite subject to the force field $-\nabla\phi$. However, satellites are moving inside the diffuse background of DM particles of the main halo. For this reason, they decay toward the centre of the halo as the dynamical friction (hereafter DF) causes the loss of orbital energy and angular momentum. The DF braking force is a force of back-reaction resulting from the global distortion of the particle density field induced by the satellite.

To determine the satellite orbit, we calculate the trajectory of a point particle with the same total mass of the satellite $M_{\rm s}$, moving in the gravitational potential of the main halo identified by the density profile $\rho(r)$:

$$\ddot{\mathbf{r}} = -\nabla\phi + \mathbf{F}_{\rm df} \ . \tag{7.3}$$

Warning: The point particle approximation is sufficient as long as the scale of the satellite orbit is much larger than the satellite itself.

As a first order approximation, we can evaluate the drag force using the Chandrasekhar (1943) formalism. Under the hypothesis that a massive point particle is moving through an infinite, homogeneous background of much lighter particles with an isotropic Maxwellian velocity distribution of zero mean, the expression for the DF drag force is:

$$\mathbf{F}_{\rm df} = -4\pi \frac{\ln\Lambda \, G^2 \rho(r) M_{\rm s}}{V^3} \left[{\rm erf}(X) - \frac{2X}{\sqrt{\pi}} e^{-X^2} \right] \mathbf{V} \ , \tag{7.4}$$

where $X = V/(\sqrt{2}\sigma)$, $\mathbf{V}$ is the velocity of $M_{\rm s}$ and σ is the velocity dispersion of the halo. $\ln\Lambda$ is the so called Coulomb logarithm: we assume as a “standard” choice of Λ

$$\ln\Lambda = \ln\left(1 + \frac{M_{\rm h}}{M_{\rm s}}\right) . \tag{7.5}$$

(e.g White 1978; Binney & Tremain 1987; Lacey & Cole 1993)

Comment: There is some arbitrariness in the choice of Λ. In the original Chandrasekhar formulation, the Coulomb logarithm is the ratio between the maximum and minimum impact parameter of the background particles contributing to the wake. However, various authors discussed the possibility of modifying this "normalisation factor" in order to better describe the interaction of the satellite with the wake and to extend the formalism to a non–homogeneous distribution of background particles (see e.g. White 1976; Carr & Lacey 1987; Maoz 1993; Domínguez-Tenreiro & Gòmez-Flechoso 1998). Some others used Λ as a free parameter to be fitted using numerical simulations (Taylor & Babul 2001; hereafter TB).

This simple formalism evidences the two main characteristic of the DF:

- As the friction is the consequence of the particle overdensity behind the satellite, which is proportional to its mass, the DF force itself is proportional to $M_{\rm s}$.
- The drag force is a function of the **local** density $\rho(r)$ of the halo.

Numerous detailed studies of satellite dynamics (Duncan, Farouki & Shapiro 1983; Tremain & Weinberg, 1984; Weinberg 1986; Bontekoe & van Albada 1987; Cora, Muzzio & Vergne 1997; van den Bosch et al. 1999; Colpi & Pallavicini 1998; Colpi, Mayer & Governato 1999) have shown that the Chandrasekhar formulation is a good **local** approximation to the drag force on an extended satellite in a finite halo system, provided that the Coulomb logarithm is adjusted properly.

Apart form the arbitrariness in the choice of the normalisation, we identify other important weaknesses in this formalism that become crucial as the satellite mass increases. The Chandrasekhar formulation totally neglects the self-gravity of the particle wake behind $M_{\rm s}$. Thus, the force is always directed along $\mathbf{V}$ and eq. (7.4) neglects the attraction of the background particles among each other.

The great success of the Chandrasekhar formulation lies in the simple form of eq. (7.4), that allows easy and fast semi-analytical calculation of the orbital evolution. Moreover, it can be integrated to estimate the dynamical friction time–scale $\tau_{\rm df}$, i.e. the time the satellite needs to sink to the center of the main halo. For a rigid satellite $\tau_{\rm df,rig}$ represents the typical life–time.

7.2 The theory of Linear Response

In the previous Section we presented the Chandrasekhar's local approximation to the DF force. However, the global nature of the distortion suggests that a more refined approach must be considered. Here, we review the Theory of Linear Response (TLR), a relatively new approach to study DF in the non-uniform particle background of a spherical self-gravitating halo.

In TLR, the response of the main system to the satellite perturbation depends only on the properties of the underlying matter field in its unperturbed state: the (self)-correlations existing among the particles ultimately leads to energy dissipation.

Under the hypothesis that the N dark matter particles of mass m are in virial spherical equilibrium, the drag force $\mathbf{F}_{\rm df}$ on a satellite described as a point of mass $M_{\rm s}$ reads:

$$\begin{aligned}\mathbf{F}_{\rm df}(t) = \quad & GM_{\rm s}m\sum_{i=1}^{N}\int_{t_0}^{t} ds\int_{\Gamma} d\Gamma\left[\nabla_{\mathbf{p}_i(s)}f_0\cdot\frac{\mathbf{r}(s)-\mathbf{R}_i(s)}{|\mathbf{r}(s)-\mathbf{R}_i(s)|^3}\right] \\ & \times\left[GM_{\rm s}m\sum_{j}^{N}\frac{\mathbf{r}(t)-\mathbf{R}_j(t)}{|\mathbf{r}(t)-\mathbf{R}_j(t)|^3}\right], \qquad (7.6)\end{aligned}$$

where $d\Gamma$ is the elementary volume in the $6N$ dimensional phase space (Γ) of the particles in the halo, and f_0 the N-point equilibrium distribution function (hereafter we denote the total halo mass $M_{\rm h}$ as Nm).

Comment: The drag on $M_{\rm s}$ is a consequence of a memory effect that develops with time. It requires the knowledge of the dynamics of the N particles $[\mathbf{R}_i(s), \mathbf{p}_i(s)]$, as determined by the unperturbed Hamiltonian, over the whole interaction time from t_0 (when the perturbation is turned on) to the current time t. The distribution function f_0 incorporates the properties of the system in virial equilibrium.

Dark matter haloes in virial equilibrium can be regarded as an assembly of collisionless particles subject to a mean field potential ϕ_0 that can be computed solving simultaneously the Poisson and the Boltzmann equation. The distribution function can thus be written in terms of the one-particle phase space density $f^{1\rm p}(\mathbf{R}, \mathbf{p})$. Under this hypothesis, and due to the statistical independence of the particles, all cross correlation terms in the previous equation cancel identically in the limit of $N \gg 1$. Only the self-correlation properties of the collisionless background survive to yield:

$$\begin{aligned} \mathbf{F}_{\rm df}(t) = \quad & (GM_{\rm s})^2\, Nm^2 \int_{t_0}^{t} ds \int d^3\mathbf{R}\, d^3\mathbf{p} \\ \times \quad & \left\{ \nabla_{\mathbf{p}(s)} f^{1\rm p} \cdot \left[\frac{\mathbf{r}(s) - \mathbf{R}(s)}{|\mathbf{r}(s) - \mathbf{R}(s)|^3} - \int d^3\mathbf{R}' n_0(r') \frac{\mathbf{r}(s) - \mathbf{R}'}{|\mathbf{r}(s) - \mathbf{R}'|^3} \right] \right\} \\ \times \quad & \frac{\mathbf{r}(t) - \mathbf{R}(t)}{|\mathbf{r}(t) - \mathbf{R}(t)|^3} \end{aligned} \tag{7.7}$$

The new term appearing in brackets (involving the equilibrium background density $n_0[r]$) represents, at a given time s, the mean force acting on $M_{\rm s}$ resulting from the system as a whole. The recoil of the halo (due to linear momentum conservation) is a coherent shift of all the orbits of the background particles, giving origin to a global correlation among them.

Comment: TLR, as is formulated, can account naturally for the shift of the center of mass of the halo and allows the use of the one-particle distribution function $f^{1\rm p}$ for the system in virial equilibrium (see Colpi & Pallavicini 1998 and Colpi 1998). Thus, $\mathbf{F}_{\rm df}$ in the form of eq. (7.7) is the force as measured in the non-inertial reference frame comoving with the halo's center of mass.

In the context of the fluctuation-dissipation theorem, the braking force can be seen as an integral over time of the correlation function of a fluctuating component of the microscopic force:

$$\mathbf{F}^a_{\mathrm{df}}(t) \equiv \int_{t_0}^{t} ds\, K^a(t-s) = \int_{t_0}^{t} ds \int d^3\mathbf{R}\, d^3\mathbf{p} \nabla^b_{\mathbf{p}(s)} f^{1\mathrm{p}}\, T^{ba} \tag{7.8}$$

where the self-correlation tensor reads:

$$T^{ba} \equiv (GM_{\mathrm{s}})^2 N m^2 \left[\frac{R^b(s) - r^b(s)}{|\mathbf{r}(s) - \mathbf{R}(s)|^3} - \int d^3\mathbf{R}' n_0(r') \frac{R^b(s) - r'^b}{|\mathbf{r}(s) - \mathbf{R}'|^3} \right] \times \frac{R^a(t) - r^a(t)}{|\mathbf{r}(t) - \mathbf{R}(t)|^3}. \tag{7.9}$$

The correlation function $K^a(t-s)$ introduces a time scale τ^* characterising the rise time of the force $\mathbf{F}^a_{\mathrm{df}}(t)$; it is the scale over which the particles redistribute the satellite's orbital energy into the internal degrees of freedom of the system.

Within TLR it is also possible to interpret $\mathbf{F}_{\mathrm{df}}$ in terms of a global time–dependent density deformation, noting that eq. (7.7) can be written formally as:

$$\mathbf{F}_{\mathrm{df}} = -GM_{\mathrm{s}} N m \int d^3\mathbf{R}\, \Delta n(\mathbf{R}, t)\, \frac{\mathbf{r}(t) - \mathbf{R}}{|\mathbf{r}(t) - \mathbf{R}|^3}\,, \tag{7.10}$$

where the function $\Delta n(\mathbf{R}, t)$ maps the response, i.e., the time dependent changes in the density field $n_0(r) + \Delta n(\mathbf{R}, t)$ resulting from the superposition (memory) of disturbances created by the satellite over the entire evolution; the function $\Delta n(\mathbf{R}, t)$ can be derived comparing eq. (7.10) with eq. (7.7) (see also Colpi & Pallavicini 1998 for details).

Eq. (7.7) applies when the interaction potential between M_{s} and the particles is weak relative to the mean field potential ϕ_0 of the equilibrium system (when isolated). This is the reason way only the properties of the halo in virial equilibrium are requested to evaluate $\mathbf{F}_{\mathrm{df}}$. As a consequence, $\mathbf{F}_{\mathrm{df}}$ is accurate to second order in the coupling constant G. Higher orders terms would describe the self-gravity of the response, i.e., the modification in the self interaction potential due to the external perturbation driven by M_{s}. Eq. (7.7) can describe the sinking of satellites moving on arbitrary orbits, even outside the primary halo. Previous semi–analytical studies focused on purely circular orbits (Weinberg 1986) to explore the role of resonances and on almost radial orbits to explore the transient nature of the interaction (Séguin & Dupraz 1994).

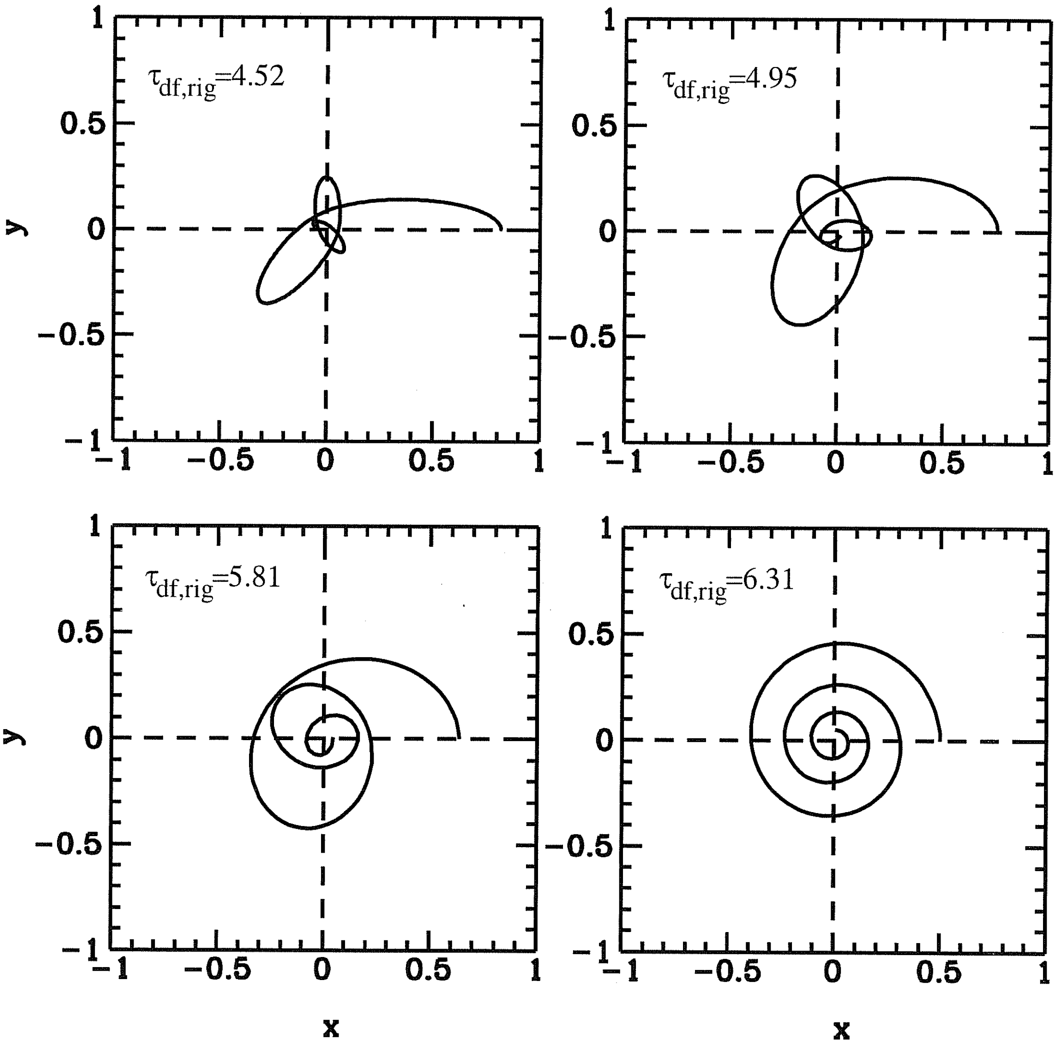

Figure 7.1: Collection of orbits in the plane (x, y) computed within TLR, for $x_c = 0.5$. The main halo has an isothermal profile. From top left to bottom right the initial circularity is $\epsilon = 0.31, 0.63, 0.91, 1.$ respectively. Length is in units ofR_h. Each plot is labelled with the corresponding dynamical friction time.

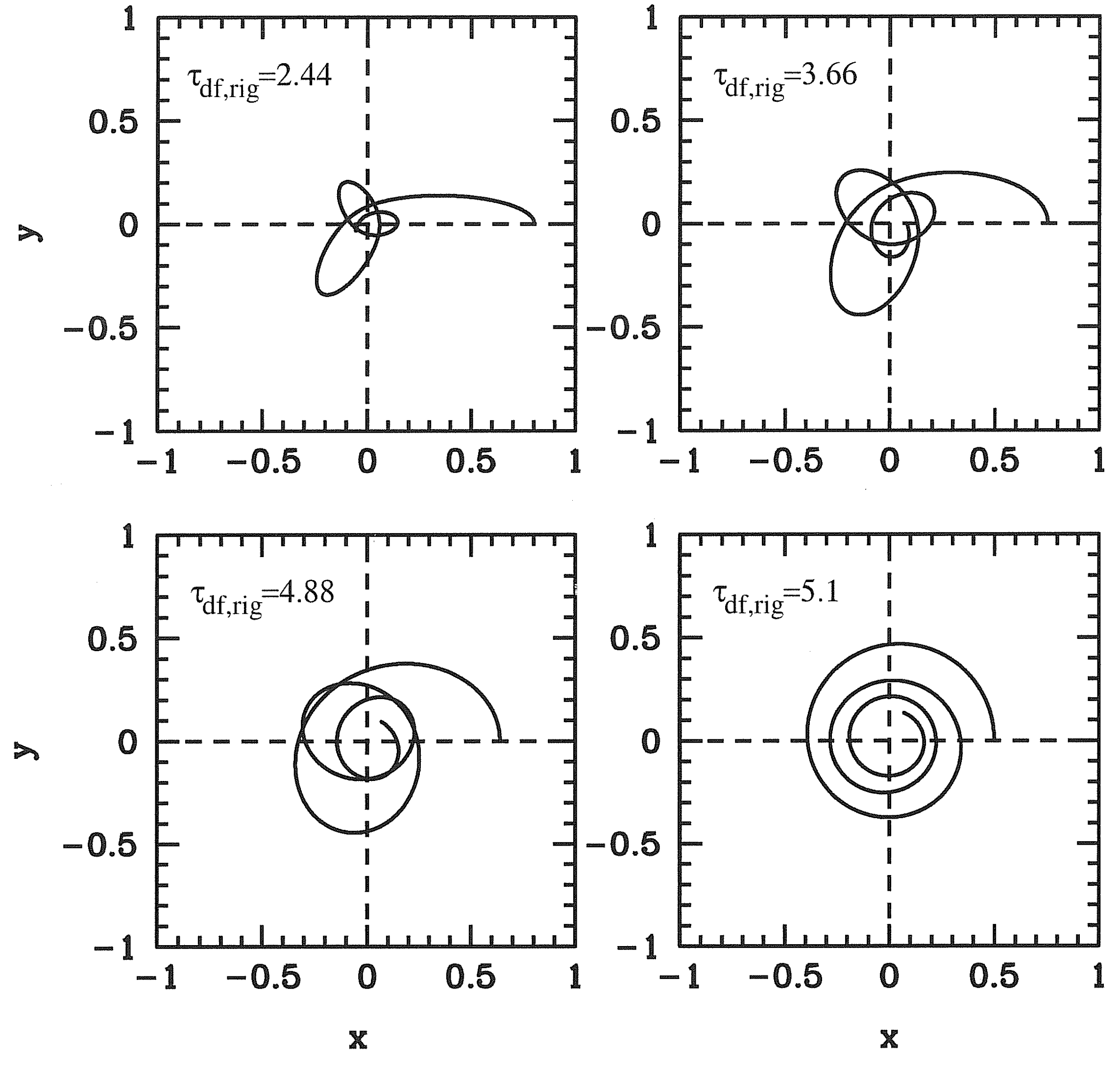

Figure 7.2: As Fig. (7.1) but the orbital evolution is calculated using the Chandrasekhar local approximation. The evolution is stopped at the same orbital angular momentum of the corresponding TLR orbit.

7.2.1 TLR: the Force of Back-reaction in a Spherically Symmetric Halo

In a nonuniform collisionless background the back-reaction force on M_s results from the combined action of a global tidal response related to the density gradients (absent in an infinite uniform medium) and from the development of an extended wake forming behind the satellite path that contributes mostly to its deceleration. The force acquires a component along $\mathbf{r}$ as the symmetry around $\mathbf{V}$ is lost, the underlying system being non homogeneous.

To estimate the drag in the domain where the satellite's velocity $\mathbf{V}$ (determined primarily by the mean field potential ϕ_0 of the unperturbed background) is comparable to the background velocity dispersion, the tidal and frictional contributions are separated, in fact they are two different aspects of the same process. Exploiting the time independence of the distribution function $f^{1\mathrm{p}}$ and of the phase-space volume $d^3\mathbf{R}\,d^3\mathbf{v}$ (hereafter $f^{1\mathrm{p}}$ will be considered as a function of $\mathbf{R}$ and $\mathbf{v} = \mathbf{p}/m$ and normalized accordingly), the drag force (eq. [7.7]) can be equivalently written as:

$$\begin{aligned}\mathbf{F}_{\mathrm{df}} = \quad & (GM_\mathrm{s})^2 Nm \int_{t_0}^{t} ds \int d^3\mathbf{R}\,d^3\mathbf{v}\ \nabla_\mathbf{v} f^{1\mathrm{p}}(\mathbf{R},\mathbf{v}) \\ \times \quad & \left[\nabla_{\mathbf{r}(s)}\psi(|\mathbf{r}(s)-\mathbf{R}|) - \int d^3\mathbf{R}'\, n_0(r')\nabla_{\mathbf{r}(s)}\psi(|\mathbf{r}(s)-\mathbf{R}'|)\right] \\ \times \quad & \nabla_{\mathbf{r}(t)}\psi(|\mathbf{r}(t)-\mathbf{R}(t-s)|) \end{aligned} \tag{7.11}$$

where ψ is proportional to the Newtonian gravitational potential

$$\psi(|\mathbf{r}-\mathbf{R}|) \equiv -\frac{1}{|\mathbf{r}-\mathbf{R}|}. \tag{7.12}$$

In eq. (7.11), $\mathbf{r}$ denotes the satellite position vector relative to the halo's center of mass, and is computed self-consistently following the actual dynamics of the satellite (that now acquires the reduced mass μ).

Finally we identify two important weaknesses of the TLR formalism:

1. Because of the difficulty of including particle dynamics as determined by the unperturbed Hamiltonian, Colpi (1998) approximated their motion as linear giving:

$$\mathbf{R}(t-s) = \mathbf{R} + (t-s)\mathbf{v}. \tag{7.13}$$

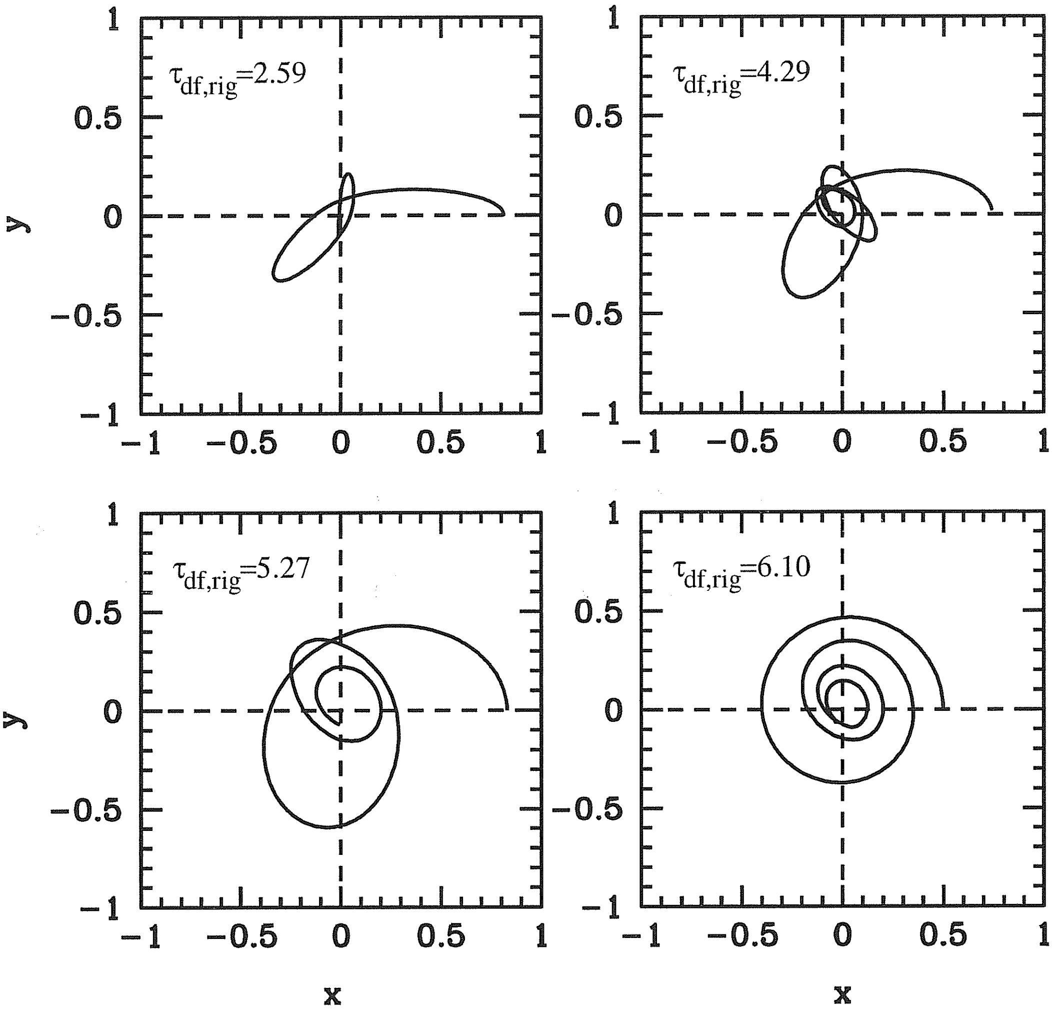

Figure 7.3: Collection of orbits in the plane (x, y) computed within TLR, for $x_c = 0.5$. The main halo has a NFW profile with $c_h = 10$. From top left to bottom right the initial circularity is $\epsilon = 0.34, 0.67, 0.92, 1.$ respectively. Length is in units ofR_h. Each plot is labelled with the corresponding dynamical friction time.

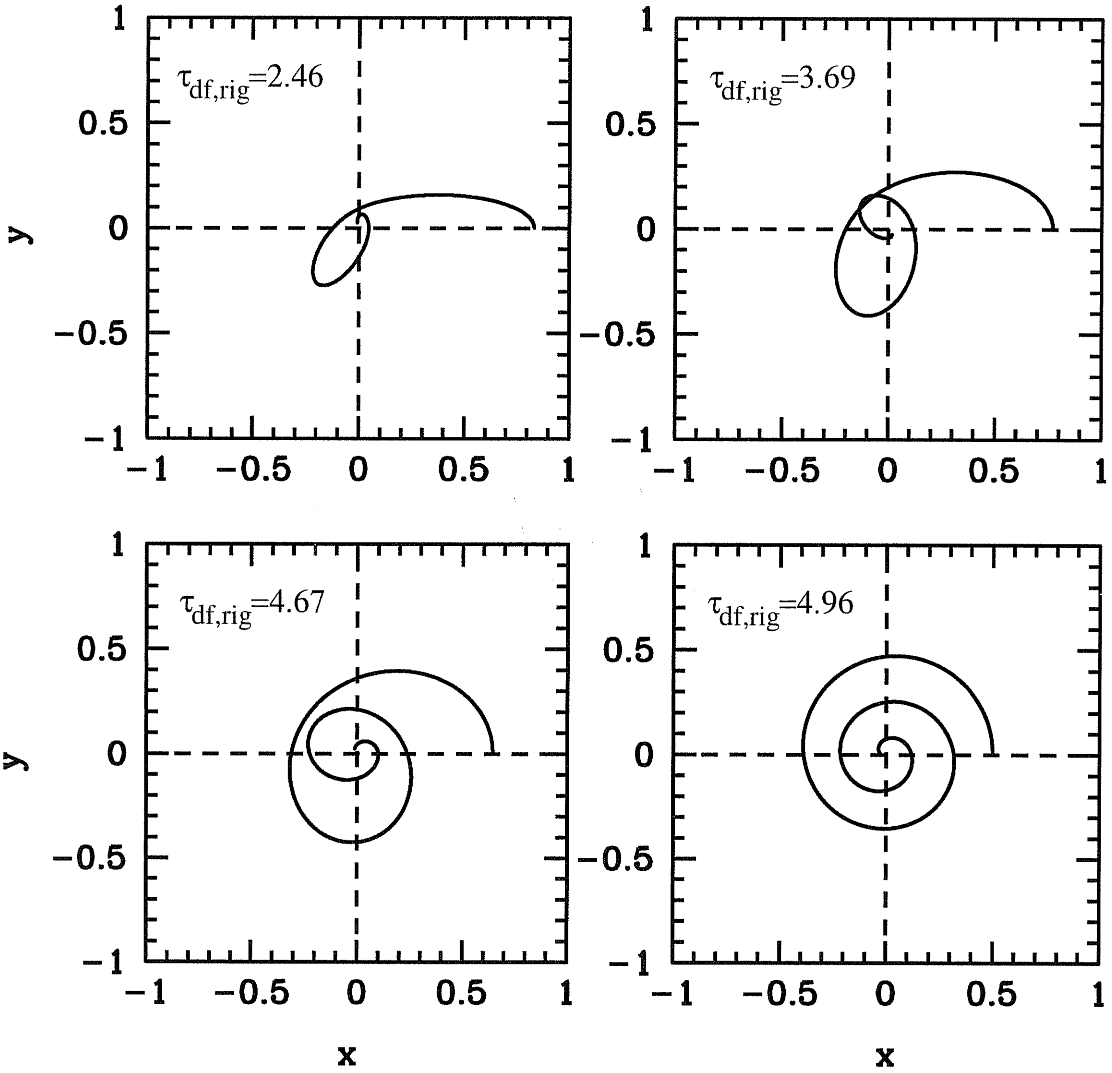

Figure 7.4: As Fig. (7.3) but the orbital evolution is calculated using the Chandrasekhar local approximation. The evolution is stopped at the same orbital angular momentum of the corresponding TLR orbit.

2. TLR does not describe short distance encounters as it is derived from a linear analysis expanded to first order in the perturbation. For a point like satellite moving in an infinite uniform medium, these encounters lead to a minimum impact parameter which is determined uniquely by V and the background velocity dispersion σ. Then, since the satellite has a finite size, in analogy with N-Body simulations, the short-distance two-body interaction ψ is smoothed introducing in the microscopic gravitational potential a softening length ϵ. Colpi, Mayer & Governato (1999; hereafter CMG) compared the model with N-Body simulations to test indirectly the validity of such an approximation finding an extremely good agreement.

Warning: The TLR neglects the acceleration of the particles, i.e., their "curvature", during the interaction of the satellite. CMG noticed that this simplification does not affect the accuracy of the method.

Finally, we notice that the drag force depends on the response of the particles and, in turn, on the characteristics of their equilibrium state which is related to the unperturbed density profile of the main halo.

7.3 The halo profile

As noticed in the previous Sections, an important role in the study of the orbital evolution of a satellite is played by the unperturbed background density profile.

N-body simulations have been extensively used to study the internal structure of DM haloes (Quinn et al. 1986; Frenk et al. 1988; Dubinsky & Carlberg 1991; Carlberg 1994; Navarro, Frenk & White 1996; 1997; Cole & Lacey 1996; Tormen, Bouchet & White 1997; Fukushige & Makino 1997; Brainerd et al. 1998; Ghigna et al. 1998; Tormen, Diaferio & Syer 1998; Huss, Jain & Steinmetz 1999; Okamoto & Habe 1999; Moore et al. 1999; Eke, Navarro & Steinmetz 2001); the shape of the inner slope of density profiles is hotly debated.

A systematic analysis performed on haloes of different mass identified at various redshifts, showed that haloes have a density profile characterised by a scale parameter $r_{\rm s}$. The so called NFW profile has the form $\rho(r) \propto 1/[r/r_s(1+r/r_s)^2]$ (Navarro, Frenk & White 1996; 1997).

More recently, Moore et al. (1999) suggested that a different profile, steeper in the internal regions, is more appropriate to fit simulated DM haloes. The authors proposed $\rho(r) \propto 1/[(r/r_s)^{1.5}(1+(r/r_s)^{1.5})]$ as the best fit to the N-body data. This result is confirmed by Ghigna et al. (2000). They used ultra-high resolution N-body simulations to study the internal properties of haloes. The authors suggested that, increasing the resolution in the N-body simulations, the profiles become closer to the Moore et al. (1999) fit.

A detailed discussion on the halo profile is beyond the aim of this thesis for an overview on this topic see Navarro, Frenk & White 1996, 1997, Moore et al. 1999, Ghigna et al. 2000 and Eke, Navarro & Steinmetz 2001. In the following we will use the singular isothermal profile, which allow simple analytical calculations and compare the results with those obtained with the NFW profile. All the calculations presented in this Chapter can be easily extended to the Moore et al. (1999) profile.

7.3.1 The singular isothermal sphere

If the halo profile is a singular isothermal sphere, then

$$\rho(r) = \frac{V_{\rm h}^2}{4\pi G r^2} \tag{7.14}$$

where $V_{\rm h}$ is the circular velocity which is constant with radius.

The mass profile of a spherically symmetric halo (i.e. the mass contained inside a sphere of radius r) can be obtained integrating eq. (7.14) over the spherical volume

$$M(r) = \frac{M_{\rm h}}{R_{\rm h}}\, r\ , \tag{7.15}$$

here $R_{\rm h}$ is the virial radius and $M_{\rm h}$ is the mass of the halo inside $R_{\rm h}$.

Using the Poisson equation it is possible to derive the gravitational potential

$$\phi(r) = V_{\rm h}^2 \ln\left(\frac{r}{R_{\rm h}}\right)\ . \tag{7.16}$$

To evaluate the decay time $T_{\rm rig}$ of a satellite moving on a circular orbit it is possible to analytically integrate eq. (7.3) to get

$$T_{\rm rig} = 1.17\, \frac{r_{\rm c}^2 V_{\rm h}}{G M_{\rm s} \ln(\Lambda)} \tag{7.17}$$

(Binney & Tremain 1987).

Lacey & Cole (1993) extended this formula incorporating the dependence of the sinking time on the initial circularity:

$$\tau_{\mathrm{df,rig}} = T_{\mathrm{rig}}\epsilon^{\alpha}. \tag{7.18}$$

They found $\alpha = 0.78$ with an accuracy better than the 3 per cent. More recently, van den Bosch et al. (1999) noticed that the decay time $\tau_{\mathrm{df,rig}}$ depends more weakly on the initial eccentricity and for the case $x_{\mathrm{c}} = 0.5$ they proposed α=0.53.

CMG used TLR to study the decay of a satellite on a singular isothermal sphere. Their results are consistent with that of van den Bosch et al. (1999). They noticed that α depends on the energy of the orbit and that for the cosmologically relevant orbits, the TLR approach (supported by a set of N-Body simulations), gives $\tau_{\mathrm{df,rig}} \propto \varepsilon^{0.4}$.

In Fig. (7.1) and Fig. (7.2) we compare the TLR and the local approximation for a collection of orbits and a satellite of $M_{\mathrm{s}} = 0.02M_{\mathrm{h}}$. Both the shape of the orbit and the DF time–scale slightly differs. We notice that the $\tau_{\mathrm{df,rig}}$ evaluated using the local approximation is in agreement with the Lacey & Cole (1993) estimate.

7.3.2 The NFW profile

Navarro, Frenk & White (1996) proposed a density profile characterized by a scale parameter r_{s}:

$$\rho(r) = \frac{M_{\mathrm{h}}}{4\pi R_{\mathrm{h}}^3}\frac{\delta_c}{(c_{\mathrm{h}}\, x)(1 + c_{\mathrm{h}}\, x)^2}\ , \tag{7.19}$$

where $x = r/R_{\mathrm{h}}$ is the dimensionless radius in units of the virial radius, $c_{\mathrm{h}} = r_{\mathrm{s}}/R_{\mathrm{h}}$ is the concentration parameter, and $\delta_{\mathrm{c}} = c_{\mathrm{h}}^3/[\ln(1 + c_{\mathrm{h}}) - c_{\mathrm{h}}/(1 + c_{\mathrm{h}})]$.

A family of haloes of given mass and size does not have a unique NFW profile; the concentration c_{h} plays the role of a free parameter that basically tells how much of the total mass is contained within a given inner radius.

Comment: The value of the concentration is not arbitrary. Eke, Navarro & Steinmetz (2001) (see also Navarro, Frenk & White 1996; 1997) noticed that the concentration reflects the mean density of the Universe at the (suitably defined) collapse time of the DM halo. For this reason small haloes, that form at earlier time, are more concentrated than more massive ones. If z_c is the collapse redshift of a DM halo $c_{\rm h} \propto (1+z_c)/(1+z_0)$ (see Eke, Navarro & Steinmetz 2001 for the details).

The mass profile can be written as:

$$M(r) = M_{\rm h} \frac{\ln(1+c_{\rm h}\,x) - c_{\rm h}\,x/(1+c_{\rm h}\,x)}{\ln(1+c_{\rm h}) - c_{\rm h}/(1+c_{\rm h})} \,, \tag{7.20}$$

and used to calculate the circular velocity profile, $V_{\rm c}^2(r) = GM(r)/r$, and the one-dimensional velocity dispersion $\sigma(r)$

$$\begin{aligned} \sigma^2(r) &= 75.53\; V_{\rm c}^2(2.15 R_{\rm h}/c_{\rm h})\; (c_{\rm h}x)(1+c_{\rm h}x)^2\; \mathcal{I}(c_{\rm h}x) \\ \mathcal{I}(x) &= \int_x^{\infty} \left[\frac{\ln(1+y)}{y^3(1+y)^2} - \frac{1}{y^2(1+y)^3} \right] dy \,, \end{aligned} \tag{7.21}$$

(Kolatt et al. 2000).

The gravitational potential of a NFW halo can be written as:

$$\phi(r) = -V_{\rm c}^2(r) + V_{\rm h}^2\, \frac{c_{\rm h}/(1+c_{\rm h}) - c_{\rm h}/(1+c_{\rm h}x)}{\ln(1+c_{\rm h}) - c_{\rm h}/(1+c_{\rm h})} \,, \tag{7.22}$$

here $V_{\rm h}$ is the value of the circular velocity at the virial radius

7.4 The Sinking of a Rigid Satellite in a NFW Profile

In this Section we explore the evolution of a rigid satellite of mass $M_{\rm s}$ orbiting inside a halo with NFW density profile. The halo is scaled to the Milky Way mass $M_{\rm h} = 10^{12} M_{\odot}$, has a tidal radius $R_{\rm h} = 200$ kpc and concentration $c_{\rm h} = 7$ or 14, within the spread of cosmological values (Eke, Navarro & Steinmetz 2001).

We carry out a detailed study of dynamical friction using TLR and N-body simulations to gain insight into the physical mechanisms that cause the satellite orbital decay.

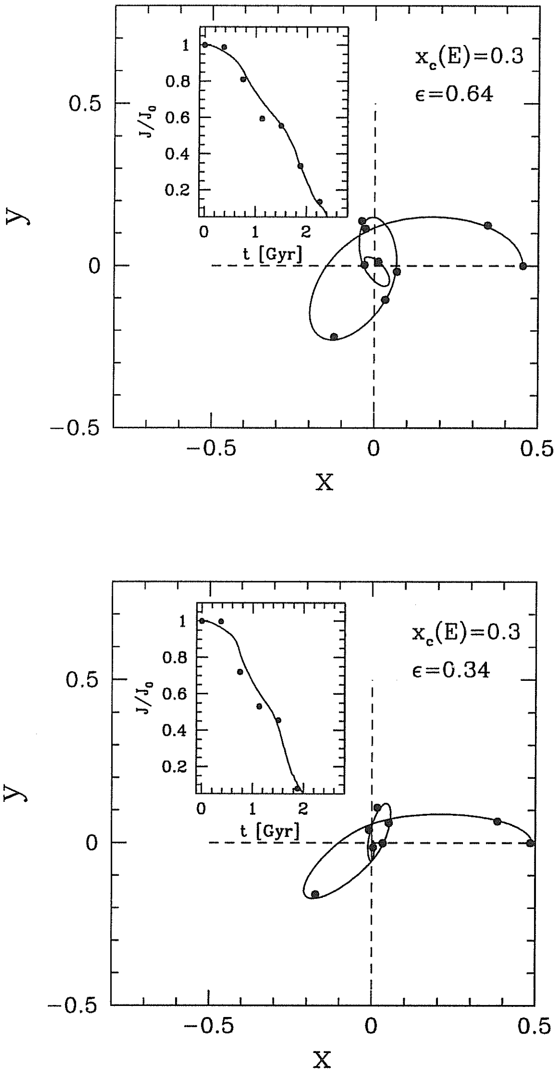

Figure 7.5: Dynamical evolution of $M_{\rm s}/M_{\rm h} = 0.02$ for $c_{\rm h} = 7$. The initial orbital parameters are $x_{\rm c}(E) = 0.3$ and $\epsilon = 0.64$ (top panel), $\epsilon = 0.34$ (bottom panel). Solid lines are the TLR results, points are the N-body data. We also plot the orbital angular momentum normalized to the initial one as a function of time. TLR reproduces with good accuracy both the path of the satellite and the loss of orbital angular momentum.

7.4.1 The simulations

To test the ability of TLR to describe the orbital decay of a rigid satellite, we compare TLR predictions with a set of N-body simulations. We use numerical simulations performed with PKDGRAV, a high-performance parallel binary treecode developed by the HPCC group in Seattle (Dikaiakos & Stadel 1996; Stadel, Wadsley & Quinn, in preparation). PKDGRAV has multistepping capabilities which makes it ideal for following accurately and efficiently a rapidly varying density field like that typical of simulations with tidal interactions (see Mayer et al. 2001). In the N-Body simulations the primary halo has 10^5 particles: it was first evolved in isolation for 10 Gyr and the stability of the density profile was verified.

The simulations have been carried out by Lucio Mayer at the CINECA Supercomputing Center (Bologna) and on a dual-processor ALPHA workstation at the University of Washington.

In Fig. (7.5) we plot the orbital path of a satellite $M_{\rm s}/M_{\rm h} = 0.02$ for different initial orbital parameters. The agreement between the numerical experiments and the TLR prediction is very accurate: TLR reproduces the orbital evolution of the satellite with an error lower that 5 per cent. This agreement also extends to the loss of orbital angular momentum.

Comment: The N-body simulations and TLR agree in a number of details on the evolution, the most remarkable being the temporary rise of the angular momentum observed during the final stages of the decay when $M_{\rm s}/M_{\rm h} = 0.02$. This is a manifestation of the fact that in the background medium, no longer uniform, the satellite moves inside or close to its distorted wake that, near pericenter, induces a positive torque.

7.4.2 The decay time in NFW profile

As TLR reproduces with excellent accuracy the N-body simulations, we use this semi–analytical approach to study the orbital decay of a rigid satellite.

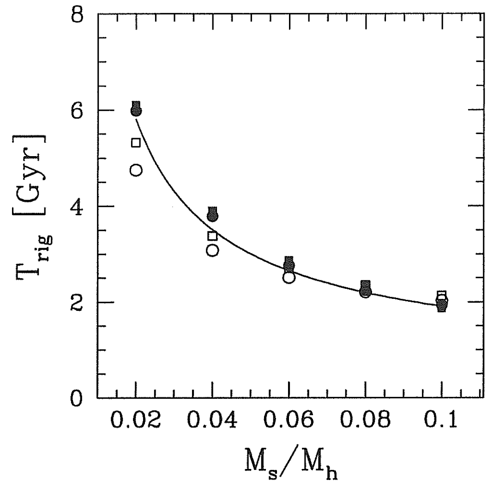

Figure 7.6: Dynamical friction time–scale T_{rig} versus $M_{\mathrm{s}}/M_{\mathrm{h}}$ for a satellite in a Milky Way like halo with $x_{\mathrm{c}}(E) = 0.5$ and $\epsilon = 1$; filled symbols are from TLR, while open symbols are from the local approximation of dynamical friction as given by solving eq. (7.24). Dots refer to $c_{\mathrm{h}} = 7$ while squares for $c_{\mathrm{h}} = 14$; the solid line corresponds to the fit given by eq. (7.23).

The first problem is to quantify the dependence of the DF time–scale on the satellite mass M_{s}. In Fig. (7.6) we plot the dynamical friction time for circular orbits T_{rig} for different M_{s} (expressed in units of M_{h}). We also explore the dependence of T_{rig} with the main halo concentration, and we notice no significant changes of the dacay time–scale for different values of c_{h}.

Using TLR data, we provide a useful fitting formula:

$$T_{\mathrm{rig}} \sim 1.3 \frac{R_{\mathrm{h}}^2 V_{\mathrm{h}}}{G\, M_{\mathrm{s}} \ln \Lambda} x_{\mathrm{c}}^2 \,. \tag{7.23}$$

This fit tries to single out the dependences of T_{rig} not only on M_{s} but also on the satellite initial orbit in a simple way and ties to the familiar expression of Chandrasekhar (1943) derived in the local approximation for the drag force (eq. [7.17]).

As done for the isothermal profile, we can apply the Chandrasekhar formalism to derive an analytical formula for T_{rig} in the case of NFW profiles. Treating the background density and dispersion velocity as local quantities the orbital equation

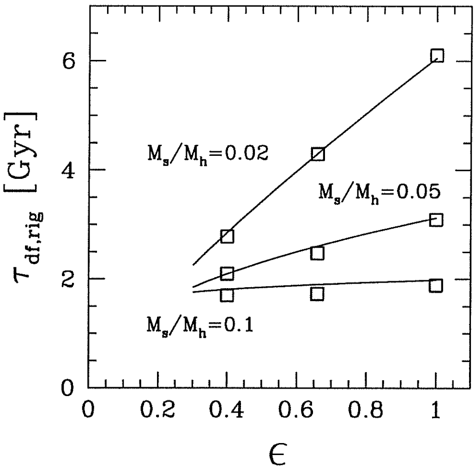

Figure 7.7: Dynamical friction time $\tau_{\rm df,rig}$ versus circularity ϵ for for $M_{\rm s}/M_{\rm h} = 0.02, 0.05$ and 0.1. The orbital energy is $x_{\rm c}(E) = 0.5$ and $c_{\rm h} = 10$. The solid lines are the model results (eq. [7.4] and eq. [7.30]) and the open squares are TLR data.

(eq. [7.3]) of a satellite spiraling down by DF along a circular orbit is:

$$\frac{1}{r}\frac{d\left[r\,V_{\rm c}(r)\right]}{dt} = -4\pi \ln\Lambda G^2 M_{\rm s}\,\frac{\rho(r,c_{\rm h})}{V_{\rm c}^2(r)} \times \left[{\rm erf}(Y) - 2\frac{Y}{\sqrt{\pi}}e^{-Y^2}\right] \tag{7.24}$$

where $Y = V_{\rm c}(r)/\sqrt{2}\sigma(r)$. This equation can be integrated grouping all quantities depending on r, on the right hand side of eq. (7.24) to give

$$T_{\rm rig} = -\frac{GM_{\rm s}\ln\Lambda}{R_{\rm h}^2 V_{\rm h}}\int_{x_{\rm c}}^{0}\Theta(x,c_{\rm h})\,dx\,. \tag{7.25}$$

The function $\Theta(x,c_{\rm h})$ has an analytical expression that can be fitted, with an average error of one part over 1000, as

$$\Theta(x,c_{\rm h}) \simeq f(c_{\rm h})\,x^{0.97}, \tag{7.26}$$

leading to a dynamical friction time–scale

$$T_{\rm rig} \sim 0.6 f(c_{\rm h})\,\frac{R_{\rm h}^2 V_{\rm h}}{GM_{\rm s}}\,\frac{x_{\rm c}^{1.97}}{\ln\Lambda}\,, \tag{7.27}$$

where $f(c_{\rm h})$ is:

$$f(c_{\rm h}) = 1.6765 + 0.0446\, c_{\rm h} \,. \tag{7.28}$$

This simple analysis explains why a fit similar to that for a singular isothermal sphere, as reported in eq. (7.23), is acceptable even in a NFW profile.

Comment: The concentration of the main halo plays an important role in eq. (7.27) while it does not sensibly affect the decay time in the TLR. This behaviour is shown in Fig. (7.6), where we plot the decay time for $c_{\rm h} = 7$ and $c_{\rm h} = 14$. We believe that, due to its local nature, the Chandrasekhar formalism is more sensitive to any variation of the density profile. On the contrary, TLR (where the drag force is determined by the reaction of the whole system) is less affected by the concentration parameter.

The decay time on eccentric orbits

As accretion of satellite haloes occurs preferentially along rather eccentric orbits we explored the dependence of $\tau_{\rm df,rig}$ on the orbital eccentricity. In agreement with the results derived for the isothermal profile (Lacey & Cole 1993) we found that:

$$\tau_{\rm df,rig} = T_{\rm rig} \cdot \epsilon^{\alpha} \,. \tag{7.29}$$

Using TLR, we found that α depends on $x_{\rm c}(E)$, and on $M_{\rm s}/M_{\rm h}$, and that whereby relatively heavy satellites decay on a time almost independent of ε, lighter satellites decay on a much shorter time when $\varepsilon \to 0$. This is shown in Fig. (7.7). A useful fit to α as a function of circularity and mass ratio is

$$\begin{aligned} &\alpha(x_{\rm c}, M_{\rm s}/M_{\rm h}) \simeq \\ &0.475 \left\{1 - \tanh\left[10.3 \left(\frac{M_{\rm s,0}}{M_{\rm h}}\right)^{0.33} - 7.5 x_{\rm c}\right]\right\} . \end{aligned} \tag{7.30}$$

7.5 The determination of the initial orbital parameters

In the previous Sections we showed that the orbital evolution is determined by the value of the initial orbital energy and angular momentum. Then, an important aspect in the study of the evolution of satellites is to provide a reliable prediction of the value of $x_{\rm c}(E)$ and ϵ in a Cosmological environment.

Cosmological simulations aimed at studying the building up of cosmic structures (Tormen 1997; Ghigna et al. 1998; Tormen, Diaferio & Seyer 1998; Kravtsov & Klypin 1998) showed that the majority of satellite's orbits have rather large eccentricities. In particular Tormen (1997) found that the distribution of the circularity is symmetric and has an average value of 0.53 ± 0.23 (1σ of the distribution). Cole et. al (2000) re-analysed Tormen simulations and they found that a particular combination of the orbital parameters

$$\Theta_{\rm orb} = \epsilon^{0.78}\left[\frac{r_{\rm c}(E)}{R_{\rm vir}}\right]^2 , \tag{7.31}$$

is distributed according to a lognormal function. In details, the distribution of $\Theta_{\rm orb}$ has mean value $\langle\log_{10}(\Theta_{\rm orb})\rangle = -0.14$ and dispersion $\langle(\log_{10}(\Theta_{\rm orb}) - \langle\log_{10}(\Theta_{\rm orb})\rangle)^2\rangle^{0.5} = 0.26$.

In the semi-analytical, EPS-based codes for galaxy formation, the orbital parameters are in general Monte Carlo piked-up from suitable distributions: for example Cole et. al (2000) use the log normal distribution, while Somerville & Primack (1999) extracts the eccentricity using the distribution proposed by Tormen (1997) and assign each satellite the same initial energy: $r_{\rm c}(E) = R_{\rm h}$.

On the contrary, 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 angular momentum of haloes presented in §5.2. Given the impact (Zel'dovich) velocity $\Delta\mathbf{v}$ and the relative distance Δr the angular momentum and the energy are computed as:

$$\begin{aligned} \mathbf{J} &= \Delta\mathbf{r}\times\Delta\mathbf{v} & (7.32)\\ E &= \frac{1}{2}\left(\Delta\mathbf{v}\right)^2 + \phi(|\Delta\mathbf{r}|)\,. & (7.33) \end{aligned}$$

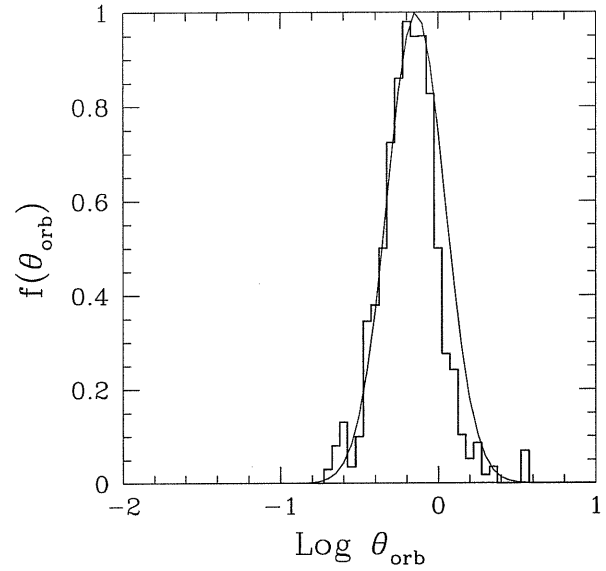

Figure 7.8: The distribution of $\Theta_{\rm orb}$ for the satellites that merge with a halo of mass $M = 2 \times 10^{14}$ $M_{\odot}$ at z=0. The solid line is the Cole et al. (2000) analyitical fit, the histogram is the PINOCCHIO result.

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

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 $M = 2 \times 10^{14}$ $M_{\odot}$. The result of our analysis are presented in Fig. (7.8), we compare the distribution of the $\Theta_{\rm orb}$ factor measured in PINOCCHIO (histogram) with the theoretical fit derived by Cole et. al (2000) (solid line). We note that the distribution measured from PINOCCHIO reproduce with good accuracy the log normal function. The average value derived by our analysis is $\langle \log_{10}(\Theta_{\rm orb}) \rangle = -0.18$ and the dispersion is $\langle (\log_{10}(\Theta_{\rm orb}) - \langle \log_{10}(\Theta_{\rm orb}) \rangle)^2 \rangle^{0.5} = 0.23$.

7.6 Conclusions

In this Chapter we discussed the orbital evolution of rigid DM satellites moving inside more massive haloes. We identified the dynamical friction as the main responsible of the "death" of such satellites as the drag force drives them towards the centre of mass of the system where they can merge with the central object.

We compared the TLR prediction with the results of high-resolution N-body simulations and we find that this semi-analytical procedure reproduces the numerical experiments with an accuracy better than 5 per cent.

Using the TLR predictions we provide a fitting formula for the decay time of a rigid satellite in a NFW profile:

$$\tau_{\rm df,rig} \sim 1.3 \frac{R_{\rm h}^2 V_{\rm h}}{G\, M_{\rm s} \ln \Lambda} x_{\rm c}^2\ \epsilon^{\alpha}\ , \tag{7.34}$$

where α is a function of the satellite mass and of the orbital energy (eq. [7.30]). This fomula allow us to determin the time–scale for the death of rigid satellites with an average error lower than 10 per cent.

Chapter 8

Dynamical Evolution of a live Satellite

> *There is a connection here...*
> *I just know it.*
>
> B. Watterson

The evolution of a rigid object is determined by the frictional drag force and its survival time corresponds to the dynamical friction time. However, a real satellite is not a rigid point mass but an extended distribution of particles moving on bound orbits inside a halo. Its life is then dramatically influenced by the tidal perturbations induced by the gravitational field of the primary halo. The global effect of the tidal perturbation is the progressive evaporation of the satellite. This process takes place during the orbital evolution and it is generally sensitive to the internal properties of the satellite and of the surrounding halo.

Our aim is to model in a realistic way the tidal effects in order to evaluate the mass that remains bound to the satellite, $M_{\rm s}(t)$, at each time along the orbit.

This Chapter is organized as follow. In the first two Sections we describe the tidal effects that induce the satellite mass loss. We distinguish two tidal effects:

a tidal truncation (*tidal cut*), originated by the average tidal force exerted by the main halo at the distance of the satellite (§8.1), and an *evaporation* effect induced by the rapidly varying tidal force near pericenters for satellites moving on an eccentric orbit (§8.2). In the latter case we speak of tidal shocks – short impulses are imparted to the bound particles within the satellite, heating the system and causing its dissolution. In §8.2.2 we present a model to study the combined effect of the orbital decay and of tidal stripping. Finally in §8.4 we discuss the fate of a live satellite of initial mass $M_{\rm s,0}$ and we provide a simple analytical expression for its decay time and disruption time.

8.1 The tidal truncation

A satellite is truncated at its tidal radius $R_{\rm tid}$ that is the distance between the center of mass of the satellite and the saddle point of the gravitational potential of the total system. Loosely speaking $R_{\rm tid}$ corresponds to the radius (relative to the satellite center) at which the mean density of the satellite is of the order of the mean density of the hosting halo, at the satellite position r:

$$\bar{\rho}_s(R_{\rm tid}) \approx \bar{\rho}_{\rm h}(r) \; . \tag{8.1}$$

The evaluation of the tidal radius requires a relation between $R_{\rm tid}$ and r which is customarily derived from the equivalence between the internal gravity and external tides leading to the implicit equation for $R_{\rm tid}$:

$$R_{\rm tid} = r \left\{ \frac{M_{\rm s}(R_{\rm tid})}{(2 - \partial \ln M_{\rm h} / \partial \ln r)\, M_{\rm h}(r)} \right\}^{1/3} \tag{8.2}$$

(Tormen, Diaferio & Syer 1998). The mass tidally lost, $\Delta M_{\rm tid}$, is computed subtracting spherical shells with $R > R_{\rm tid}$, using eq. (7.20) .

Warning: While strictly valid for a satellite moving on a circular orbit (where the combined potential over the system is static in the satellite's frame) $R_{\rm tid}$ gives, in the case of non circular motion, an approximate expression for the instantaneous tidal radius (Binney & Tremain 1987). This implies that, on stable orbits, $\Delta M_{\rm tid}$ is maximum at the first pericenter passage; the mass of the satellite would then remain constant.

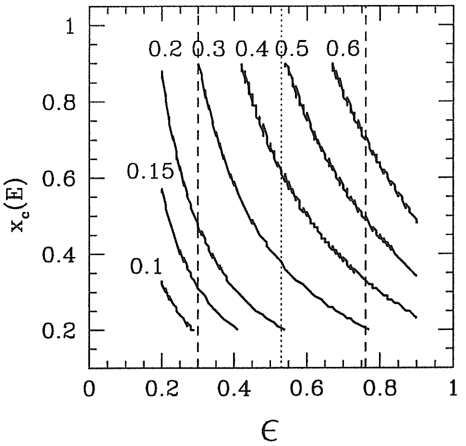

Figure 8.1: The residual mass of a satellite at the first pericenter as function of the orbital parameters, when $c_s/c_h = 2$. Each line is labeled with the corresponding $M_{s,per}/M_{s,0}$, where $M_{s,per}$ is the satellite mass at first periastron evaluated according eq. (8.2). We assume that the "tidal cut" instantaneously reduces the satellite mass. The vertical dotted line is the most probable value of the eccentricity in a cosmological environment (Tormen 1997), the dashed vertical lines are the 1σ variance.

In Fig. (8.1) we give the residual mass after instantaneous tidal cut, as a function of circularity, as computed using eq. (8.2).

Tidal stripping however does not occur instantaneously, and, following TB suggestion, we model mass loss, over a few orbital periods, adopting the expression:

$$\frac{dM}{dt} \simeq \frac{\Delta M_{\rm tid}(t)}{2\pi/\omega(t)} \tag{8.3}$$

where $\omega(t)$ is the instantaneous orbital angular velocity. This is compared with results from numerical simulations [1]. Fig. (8.5) gives the satellite mass as a function of time for a two different run. We find that mass loss by tidal cut, as described by eq. (8.3), reproduces the result of our N-body simulation only in the early phase: the satellite loses mass at a peace larger than predicted by eq. (8.3) (we refer to the dashed line of Fig. [8.5]). We believe that this is due to the action of tidal shocks and not a numerical artifact.

---

[1] The N-Body simulations were presented in §7.4.1. In this case we employ 100.000 particles for the primary halo and 50.000 particles for the satellites.

In fact, the number of particles in the simulations is chosen in order to avoid, as much as possible, numerical effects of two-body relaxation which could increase, artificially, the overall evaporation rate (Moore, Katz & Lake 1996; Gnedin & Ostriker 1999) . Numerical relaxation disperses satellite particles over a timescale related to the number of particles N

$$t_{\rm rh} = 0.138 \frac{M_{\rm s}^{0.5} R_{\rm hm}^{1.5}}{G^{0.5} m \ln(0.4N)} \ . \tag{8.4}$$

where $R_{\rm hm}$ is the half mass radius and m is the particle mass.

As shown in Tab. (8.1) the initial relaxation time is $\sim$ 100 Gyr and remains longer than 10 Gyr as mass loss continues. This is an indication that numerical two-body relaxation is unimportant. Thus, we proceed on modeling mass loss with the inclusion of tidal shocks.

8.2 Heating & evaporation

The description of the dynamical evolution of a satellite must also include heating due to compressive tidal shocks.

The theory of shock heating was developed by Ostriker et al. (1972) and Spitzer (1987) to model the evolution of globular clusters. Recent works by Gnedin & Ostriker (1997) and Gnedin, Hernquist & Ostriker (1999; GHO) also extend this theory to tidal perturbations on satellites moving on eccentric orbits inside an extended mass distribution.

8.2.1 Calculation of the tidal energy for a NFW profile

At each pericenter passage the satellite crosses very rapidly the central and more concentrated region of the primary halo. The duration of those encounters is fast compared with the dynamical time of the object. These interactions are called tidal shocks (Spitzer 1987). We will use the results derived by GHO to describe the amount of heating due to tidal shocks on a satellite moving inside an extended mass distribution.

During an orbital period $P_{\rm orb}$ the tidal force $\mathbf{f}_{\rm tid}$ per unit mass produces a global variation on the velocity of the internal fluid:

$$\Delta \mathrm{v} = \int_0^{P_{\rm orb}} \mathbf{f}_{\rm tid}\, dt \;, \tag{8.5}$$

where we have applied the impulse approximation in the hypothesis that the time scale of interaction is short compared with the dynamical time of the satellite ($t = 0$ refers to the initial satellite's position at apocenter).

In a spherically symmetric system of mass $M_{\rm h}$, the tidal force per unit mass exerted by the background on a dark matter particle of the satellite is:

$$\mathbf{f}_{\rm tid} = \frac{GM_{\rm h}}{R_{\rm h}^3}[(3\mu - \mu')(\hat{\mathbf{r}} \cdot \mathbf{R}_{\rm s})\hat{\mathbf{r}} - \mu \mathbf{R}_{\rm s}] \;, \tag{8.6}$$

where $\hat{\mathbf{r}} = \mathbf{r}/R_{\rm h}$ is the direction to the center of mass of the satellite (CMS), $\mathbf{R}_{\rm s}$ is the position of the particle respect to CMS. Note that $R_{\rm h}$ is the virial radius of the main system. Here:

$$\mu(r) = \frac{M(r)}{M_{\rm h}} \;, \tag{8.7}$$

is the adimensional mass profile, and:

$$\mu'(r) = \frac{d\mu(r)}{d\ln r} \;. \tag{8.8}$$

For a NFW profile μ and μ' are functions of the normalized radius $x = r/R_{\rm h}$ and of the concentration $c_{\rm h}$ of the primary halo:

$$\mu(x, c_{\rm h}) = \frac{\ln(1 + c_{\rm h}x) - c_{\rm h}x/(1 + c_{\rm h}x)}{\ln(1 + c_{\rm h}) - c_{\rm h}/(1 + c_{\rm h})} \;, \tag{8.9}$$

and

$$\mu'(x, c_{\rm h}) = \frac{1}{\ln(1 + c_{\rm h}) - c_{\rm h}/(1 + c_{\rm h})}\left(\frac{c_{\rm h}x}{1 + c_{\rm h}x}\right)^2 \;. \tag{8.10}$$

In the case of stable orbits the angular momentum J is conserved and we can use the identity:

$$dt = (r^2/J)\, d\theta \tag{8.11}$$

to re-write eq.(8.5) into components (GHO):

$$\Delta \mathrm{v} = \frac{GM_{\rm h}}{rJ}\{(B_1 - B_3)x, (B_2 - B_3)y, -B_3 z\} \;, \tag{8.12}$$

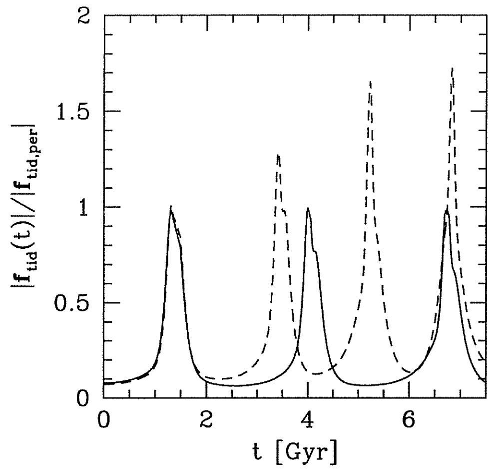

Figure 8.2: The intensity of the tidal force $|\mathbf{f}_{\rm tid}(t)|$ normalised to its value at the first periastron $|\mathbf{f}_{\rm tid,per}|$. We plot the module of the tidal force (eq. [8.6]) as a function of time for a satellite of mass $M_{\rm s,0} = 0.01 M_{\rm h}$ and $\epsilon = 0.7$ $x_{\rm c}(E) = 0.5$. The solid line refers to a stable orbit, the dashed line to an unstable one. When the drag force is active, the intensity of the tidal force, and consequently of the shock energy, grows with time.

where:

$$B_1(c_{\rm h}) = \int_{-\theta_m}^{-\theta_m} F_1(x, c_{\rm h}) \cos^2\theta \, d\theta \ , \tag{8.13}$$

$$B_2(c_{\rm h}) = \int_{-\theta_m}^{-\theta_m} F_1(x, c_{\rm h}) \sin^2\theta \, d\theta \ , \tag{8.14}$$

$$B_3(c_{\rm h}) = \int_{-\theta_m}^{-\theta_m} \frac{\mu(x, c_{\rm h})}{x} \, d\theta \ , \tag{8.15}$$

with:

$$F_1(x, c_{\rm h}) = 3\frac{[\ln(1 + c_{\rm h}x) - c_{\rm h}x/(1 + c_{\rm h}x)] - [c_{\rm h}x/(1 + c_{\rm h}x)]^2}{x[\ln(1 + c_{\rm h}) - c_{\rm h}/(1 + c_{\rm h})]} \ . \tag{8.16}$$

Here θ_m is the maximum value of the position angle.

This velocity change causes a reduction of the binding energy of the system:

$$\langle \Delta E \rangle = \left\langle \frac{1}{2} |\Delta \mathrm{v}|^2 \right\rangle \ . \tag{8.17}$$

Averaging over an ensemble of dark matter particles in a spherically symmetric satellite we have that $\langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle = R_{\rm s}^2 \ / \ 2$, and using eq.(8.12), the tidal energy gained by the satellite becomes:

$$\langle \Delta E \rangle = \left(\frac{G M_{\rm h}}{J R_{\rm h}} \right)^2 \left[\frac{(B_1 - B_3)^2 + (B_2 - B_3)^2 + B_3^2}{6} \right] R_{\rm s}^2 \,. \tag{8.18}$$

We notice that in the previous expression the contribution due to the halo and the orbital parameters (ϵ and $x_{\rm c}^2[E]$) is confined in the function:

$$\Xi\,[c_{\rm h}, x_{\rm c}(E), \epsilon] = \left(\frac{G M_{\rm h}}{R_{\rm h}^2 V_{\rm h}} \right)^2 \left[\frac{(B_1 - B_3)^2 + (B_2 - B_3)^2 + B_3^2}{6\, x_{\rm c}^2(E)\, \epsilon^2} \right] \,, \tag{8.19}$$

where $V_{\rm h}$ is the circular velocity of the main halo at virial radius.

It is then useful to write the shock energy as:

$$\langle \Delta E \rangle = \Xi\,[c_{\rm h}, x_{\rm c}(E), \epsilon] \; R_{\rm s}^2 \,. \tag{8.20}$$

When the frictional drag force is active, it is not possible to change the integration variable according to eq. (8.11). The energy change becomes:

$$\langle \Delta E \rangle = \left(\frac{G M_{\rm h}}{R_{\rm h}^3} \right)^2 \left[\frac{(A_1 - A_3)^2 + (A_2 - A_3)^2 + A_3^2}{6} \right] R_{\rm s}^2 \,. \tag{8.21}$$

Here:

$$A_1(\epsilon, x_{\rm c}[E]) = \int_0^{P_{\rm orb}} F_2(x, c_{\rm h}) \cos^2\theta \, dt, \tag{8.22}$$

$$A_2(\epsilon, x_{\rm c}[E]) = \int_0^{P_{\rm orb}} F_2(x, c_{\rm h}) \sin^2\theta \, dt, \tag{8.23}$$

$$A_3(\epsilon, x_{\rm c}[E]) = \int_0^{P_{\rm orb}} \frac{\mu(x, c_{\rm h})}{x^3} \, dt, \tag{8.24}$$

with:

$$F_2(x, c_{\rm h}) = 3\frac{[\ln(1 + c_{\rm h}x) - c_{\rm h}x/(1 + c_{\rm h}x)] - [c_{\rm h}x/(1 + c_{\rm h}x)]^2}{x^3[\ln(1 + c_{\rm h}) - c_{\rm h}/(1 + c_{\rm h})]} \,. \tag{8.25}$$

Once again we separate the contribution due to the environment:

$$\mathcal{F}[c_{\rm h}, x_{\rm c}(E), \epsilon] = \left(\frac{G M_{\rm h}}{R_{\rm h}^3} \right)^2 \left[\frac{(A_1 - A_3)^2 + (A_2 - A_3)^2 + A_3^2}{6} \right] \,. \tag{8.26}$$

For an unstable orbit:

$$\langle \Delta E \rangle = \mathcal{F}\,[c_{\rm h}, x_{\rm c}(E), \epsilon] \; R_{\rm s}^2 \,. \tag{8.27}$$

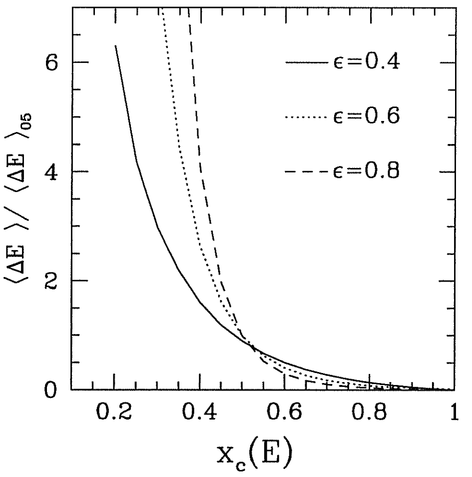

Figure 8.3: The amount of shock heating as function of the circular radius $x_c(E)$. $\langle \Delta E \rangle$ is normalized to the value of $\langle \Delta E \rangle$ when $x_c(E) = 0.5$ and $e = 0.6$. We consider three different values of the circularity: $\epsilon = 0.4$ (solid line), $\epsilon = 0.6$ (dotted line) and $\epsilon = 0.8$ (dashed line). The satellite and halo concentration are chosen such as $c_s/c_h = 2$.

The shock energy in this case must be evaluated along the perturbed orbit. As the drag force drives the satellite in the internal region of the halo the $\langle \Delta E \rangle$ increases (see Fig. [8.2]).

This model is based on the impulse approximation, it is valid to describe tidal shocks on particles whose motion is slower than the shock time. To account the conservation of adiabatic invariants of the particles which orbital period is comparable with the duration of the shock we introduce the adiabatic parameter. We define an adiabatic parameter, $x_\tau \equiv \omega\tau$ (Winberg 1995; Gnedin, Lee & Ostriker 1999; GLO), where τ is the duration of the shock, and $\omega = \sigma_s(R)/R$, here σ_s is the velocity dispersion of the satellite. The value of τ is related to the periastron crossing time: $\tau = 0.5/\omega_p$ where ω_p is the orbital angular velocity at periastron. The actual energy change is then:

$$\langle \Delta E \rangle = \mathcal{F}\left[c_h, x_c(E), \epsilon\right] \, A(x_\tau) \, R_s^2 \,, \tag{8.28}$$

where $A(x_\tau) = (1 + x_\tau^2)^\gamma$ and $\gamma = -5/2$ as suggested by GHO, and $\mathcal{F}$ is given by eq. (8.26).

We introduce a characteristic shock timescale computed, after each pericenter

Table 8.1: Characteristic timescales

Model		c_s/c_h	t_{sh}[Gyr]	P_{orb}[Gyr]	t_{rh}[Gyr]
Low concentration					
$\epsilon = 0.7$	$x_c = 0.5$	0.5	12.6	4.7	176.4
$\epsilon = 0.5$	$x_c = 0.3$	0.5	0.7	2.6	173.6
Intermediate concentration					
$\epsilon = 0.7$	$x_c = 0.5$	1	93.6	4.7	119.2
$\epsilon = 0.5$	$x_c = 0.3$	1	2.00	2.6	112.6
High concentration					
$\epsilon = 0.7$	$x_c = 0.5$	2	130	4.7	80.7
$\epsilon = 0.5$	$x_c = 0.3$	2	6.6	2.6	73.8

passage, as

$$t_{sh'} = \frac{P_{orb}}{2} \cdot \frac{E_0}{\langle \Delta E_{hm} \rangle}, \tag{8.29}$$

where $E_0 = 0.25 G M_{s,per}/R_{s,hm}$ is the binding energy of the tidally truncated satellite of mass $M_{s,per}$ evaluated according to eq. (8.3) at the time of pericentric passage. Both E_0 and $\langle \Delta E_{hm} \rangle$ are evaluated at the half mass radius $R_{s,hm}$ which is a function of the satellite concentration. A second order energy variation due to shock heating is responsible of the increase of the internal velocity dispersion and allows additional particles to leave the satellite. To account for this second order perturbation we assume that $t_{sh} = 0.43 t_{sh'}$ (GLO). Table (8.1) shows the shock time for the satellite modeled at first pericenter. The number of pericenter passages roughly necessary to unbind the satellite is t_{sh}/P_{orb}.

Lastly, we notice that $\langle \Delta E \rangle$ increases linearly with the halo concentration c_h, because in highly concentrated haloes the gradient of the gravitational force is steeper.

The amount of heating is also a function of the orbital parameters; in Fig. (8.3) we study the energy gain as a function of $x_c(E)$ for different values of the circularity. The fast growth of $\langle \Delta E \rangle$ for small values of $x_c(E)$ confirms that shocks on radial orbits are more intense: a satellite moving on a circular orbit is not subject

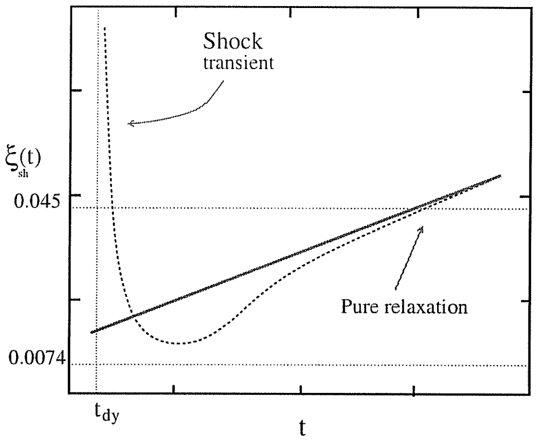

Figure 8.4: The escape probability as function of time. This cartoon qualitatively shows the function $\xi_{\rm sh}(t)$ if the satellite undergoes only one shock event (see GLO for a detailed discussion). The origin of the time axis correspond to the moment of the tidal shock. The orizontal lines show the analytical estimate of the ξ_e by Ambartsumian (1938), $\xi_{\rm e} = 0.0074$, and Hénon (1961), $\xi_{\rm e} = 0.045$. The thick solid line is the escape probability of a pure relaxation case (GLO).

to any heating.

8.2.2 Modeling the mass loss

Tidal shocks are events leading to the escape of particles. To describe globular cluster evaporation by two-body relaxation processes a so called "escape probability function" $\xi_{\rm e}$ has been introduced (Spitzer 1987; GLO). Mass loss can be predicted using the dimensionless rate of escape

$$\xi_{\rm e} \equiv -\frac{t_{\rm rh}}{M(t)}\frac{dM}{dt}\ , \tag{8.30}$$

Similarly here we define:

$$\xi_{\rm sh} \approx -\frac{t_{\rm sh}}{M(t)}\frac{dM}{dt}\ , \tag{8.31}$$

For the case of escape by two-body relaxation processes $\xi_{\rm sh}$ is a constant known to vary from $7.4 \cdot 10^{-3}$ for an isolated halo to 0.045 for a tidally truncated halo

(Spitzer 1987). On the contrary, when tidal shocks are present and dominate, the escape probability becomes a function of time; $\xi_{\rm sh}$ peaks just after each pericenter passage (GLO), rapidly decreasing until the next shock event. We find that (using both simulations and results by GLH) the escape probality can be fitted as:

$$\xi_{\rm sh}(t) \propto \left(\frac{t - t_{\rm per}}{t_{\rm tr}}\right)^{-0.5} \exp - \left(\frac{t - t_{\rm per}}{t_{\rm tr}}\right)^{0.5} , \qquad (8.32)$$

where $t_{\rm tr} \simeq 13\, t_{\rm sh}$ and $t_{\rm per}$ is the pericenter time. The shock escape probability is normalized to unity at $t = t_{\rm per} + t_{\rm dyn}$ where $t_{\rm dyn}$ is the dynamical time of the satellite.

Warning: The shock time must be evaluated at each pericenter as it varies according to the actual orbit of the satellite and its mass or equivalently its half mass radius.

The orbital timescale is often shorter that the shock timescale so that the satellite suffers a number of repeated shocks that may lead eventually to its evaporation. If it becomes unbound, further dispersal of the last particles occurs on the crossing time of the damaged system.

8.3 Testing the model

The dynamical evolution of a satellite is described using a semi-analytical code which accounts for both dynamical friction and mass loss. In this context, we use the expression of the drag force as given in eq. (7.4), since it is much faster, and in close match with TLR (see §7). At each time step we upgrade the satellite mass according to eq. (8.3) and eq. (8.31). To test the ability of our code to follow the evolution of an NFW satellite, we compare the results with those derived from a set of N-body simulations.

Tidal perturbations on stable orbits

To isolate and study the effect of a pure tidal perturbation we explore the dynamical evolution of a low concentration satellite on an unperturbed orbit. In this case, the heating by tidal shocks varies solely as a consequence of the progressive reduction of the satellite's half-mass radius. For this reason we expect a progressive reduction of the shock destructive power as time passes.

In Fig. (8.5) we show the evolution of a satellite with mass $M_{\rm s,0} = 0.01 M_{\rm h}$; the orbital parameters are chosen to reflect a typical cosmological orbit: $\epsilon = 0.65$ and $x_{\rm c}(E) = 0.5$. The bottom panel shows a low $c_{\rm s}/c_{\rm h}$ satellite, disrupted after the second pericenter passage. The top panel shows the evolution of a higher $c_{\rm s}/c_{\rm h}$ satellite surviving for more than 12 Gyr, despite having lost more than the 99 per cent of its mass. Up to the first pericenter passage only tidal cut accounts for the mass loss. The good agreement between the simulation and the code before the first pericenter suggests that the TB recipes is accurate enough to reproduce the mass loss before (or in absence of) the shock heating.

The combined effect of dynamical friction and tidal stripping

The dynamical evolution of a satellite is driven by the combined effect of dynamical friction that drives the satellite to the center of the main halo and the tidal perturbation which reduces its mass. The two processes are intimately connected as the drag force is strongly related to the mass and size of the satellite.

In Fig. (8.6) we compare the semi-analytical model with the the results of N-body simulations for satellites with $c_{\rm s}/c_{\rm h} = 2$. The initial orbital parameters are $\epsilon = 0.7$, and $x_{\rm c}(E) = 0.5$. We study two different cases: a light satellite of mass $M_{\rm s,0} = 0.02\, M_{\rm h}$ and a massive one with $M_{\rm s} = 0.1\, M_{\rm h}$. The mass loss rate and the orbital evolution are well reproduced in both cases.

The massive satellite loses mass during the orbital evolution, yet a core of bound particles survives having 5 per cent of its initial mass but sinks to the center merging with the main halo in 3 orbital periods. On the contrary, the light satellite loses 99 per cent of its mass but a bound core remains which moves on an inner orbit stable against dynamical friction, following mass loss.

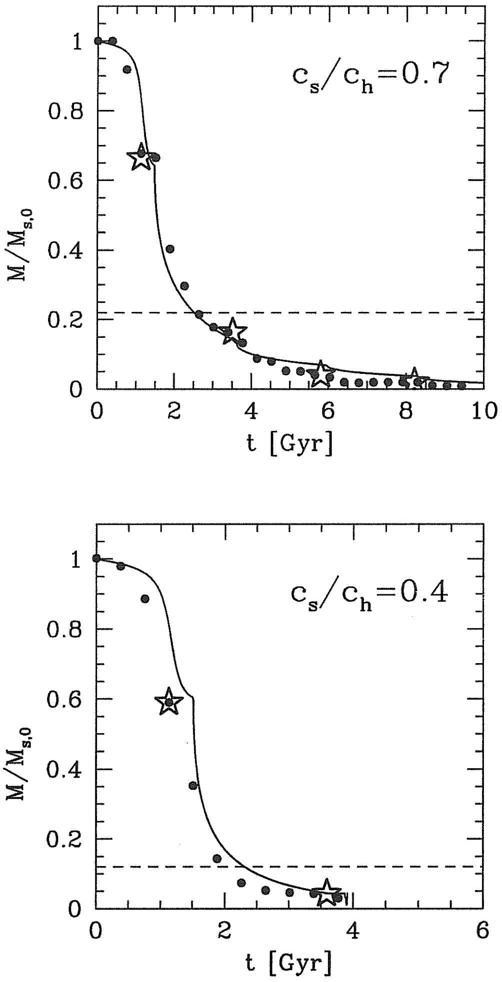

Figure 8.5: The mass loss rate of a satellite moving on a stable orbit. The orbital parameters are chosen as $\epsilon = 0.65$ and $x_c(E) = 0.5$. The halo concentration is $c_s/c_h = 0.7$ (top panel) and $c_s/c_h = 0.4$ (bottom panel). The symbols are the N-body data and the continuous line the semi-analytical model. We show the bound mass of the satellite normalized to the initial one as a function of time. Stars identify each pericenter passage. The dashed line is the asymptotic mass loss if we apply only the tidal cut using eq. (8.2).

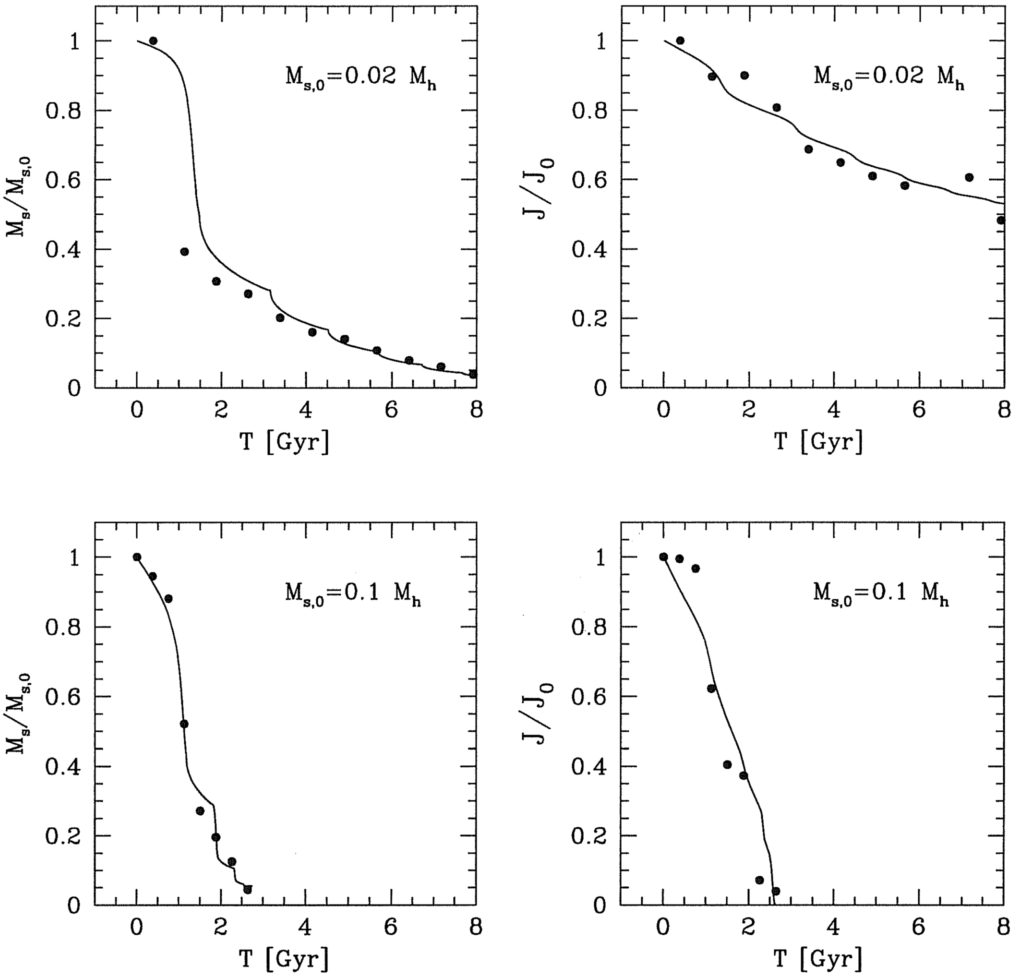

Figure 8.6: The orbit of a live satellite with $c_s/c_h = 2$, the orbital parameters are $\epsilon = 0.7$ and $x_c(E) = 0.5$. The points are the N-body data and the line the semi-analytical model. We plot the time evolution of the mass normalized to the initial one (left panels) and the orbital angular momentum scaled to the initial one (right panels). The initial mass of the satellite increase from bottom to top, the adopted values are: $M_{s,0} = 0.1\,M_h$ and $M_{s,0} = 0.02\,M_h$.

8.4 The satellite fate

Now, we use our semi-analytical model to investigate the fate of a live satellite. An individual satellite is labeled by four parameters, $x_{\rm c}[E]$ and ϵ identify the orbit, while initial mass $M_{\rm s,0}$ and concentration $c_{\rm s}$ identify the internal properties. Each combination of the four parameters leads to a different final state for the satellite: rapid merging toward the center of the main halo (M), disruption (D), or survival (S) (when a residual mass $M_{\rm s}$ remains bound and maintains its identity, orbiting in the main halo for a time longer than the Hubble time).

How much does mass loss affect the orbital decay? In Fig. (8.7) we give as a function of the initial satellite mass, $M_{\rm s,0}/M_{\rm h}$, the ratio of the dynamical friction time of a rigid satellite $\tau_{\rm df,rig}$ to the same time for a homologous live satellite $\tau_{\rm df,live}$.

Warning: We use highly concentrated satellites ($c_{\rm s}/c_{\rm h} = 2$) that are not rapidly disrupted by tidal interactions.

The figure shows that massive satellites ($M_{\rm s,0}/M_{\rm h} > 0.1$) sink to the center of the main halo and the value of $\tau_{\rm df,rig}$ for a satellite unaffected by the mass loss, gives the correct timescale over which merging occurs. On the contrary, light satellites with $M_{\rm s,0} < 0.01\, M_{\rm h}$ survive. Their $\tau_{\rm df,live}$ is longer than the Hubble time and we assume that the ratio $\tau_{\rm df,rig}/\tau_{\rm df,live}$ vanishes.

In the mass range $M_{\rm s,0} = 0.01 - 0.1 M_{\rm h}$ the satellites sink toward the center of the main halo but lose mass so efficiently that the dynamical friction time becomes two or tree times longer than $\tau_{\rm df,rig}$, while still remaining shorter than the Hubble time. Thus dynamical friction still decides their final fate (merging for the present choice of the concentration ratio). In §8.4.1 we will give a simple expression for the disruption time $t_{\rm dis}$ that can be used for a comparison with the other timescales.

In the following, we estimate the analytical expression for $\tau_{\rm df,live}$ in the three regimes.

For massive satellites, $M_{\rm s,0} \geq 0.1\, M_{\rm h}$, the dynamical friction time is not affected by the mass loss so

$$\tau_{\rm df,live} \sim \tau_{\rm df,rig} \simeq 0.5\, \frac{R_{\rm h}^2 V_{\rm h}}{G M_{\rm s}}\, \mathcal{A}_{\rm rig}\left[M_{\rm s,0}/M_{\rm h}, x_{\rm c}(E), c_{\rm h}\right] \,, \qquad (8.33)$$

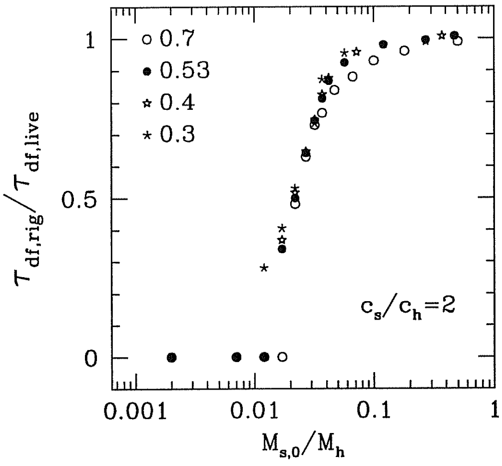

Figure 8.7: The life time of a satellite with $c_{\rm s}/c_{\rm h} = 2.$. We plot the ratio of the dynamical friction time of a rigid and a live satellite of equal initial mass, as function of $M_{\rm s,0}/M_{\rm h}$. $x_{\rm c}(E) = 0.5$ and we vary the circularity labeled with different symbols.

where

$$\mathcal{A}_{\rm rig}\left[\frac{M_{\rm s,0}}{M_{\rm h}}, x_{\rm c}(E), c_{\rm h}\right] = f(c_{\rm h})\,\frac{x_{\rm c}^{1.97}(E)}{\ln\left(1 + M_{\rm h}/M_{\rm s,0}\right)}\,, \tag{8.34}$$

where $f(c_{\rm h})$ is given by eq. (7.28). In this case, $\tau_{\rm df,live}$ is not a function of the orbital eccentricity.

For $0.01\ M_{\rm h} < M_{\rm s,0} < 0.1\ M_{\rm h}$, we provide a useful fit:

$$\tau_{\rm df,live} \sim \frac{R_{\rm h}^2\,V_{\rm h}}{GM_{\rm s,0}}\mathcal{A}_{\rm live}\left[\frac{M_{\rm s,0}}{M_{\rm h}}, \frac{c_{\rm s}}{c_{\rm h}}, x_{\rm c}(E), \epsilon\right]\,, \tag{8.35}$$

where

$$\begin{aligned}\mathcal{A}_{\rm live}\left(\frac{M_{\rm s,0}}{M_{\rm h}}, \frac{c_{\rm s}}{c_{\rm h}}, x_{\rm c}, \epsilon = 1\right) &= \left[\frac{0.25}{(c_{\rm s}/c_{\rm h})^6} - 0.07\,\frac{c_{\rm s}}{c_{\rm h}} + 1.123\right] \\ &\times \left[B(x_{\rm c})\left(\frac{M_{\rm s,0}}{M_{\rm h}}\right)^{0.12} + C(x_{\rm c})\left(\frac{M_{\rm s,0}}{M_{\rm h}}\right)^{2}\right]\,. \end{aligned} \tag{8.36}$$

Here

$$B(x_{\rm c}) = -0.050415 + 0.33554\,x_{\rm c} + 0.32807\,x_{\rm c}^2\,, \tag{8.37}$$

$$C(x_{\rm c}) = 2.1516 - 14.176\,x_{\rm c} + 27.383\,x_{\rm c}^2\,. \tag{8.38}$$

This fitting formula reproduces the semi-analytic estimate of the decay time of a live satellite on circular orbits with an error of $\gtrsim$ 10 per cent, for 0.2$<$ $x_{\rm c}(E)$ $<$1.

For eccentric orbits we find that:

$$\mathcal{A}_{\rm live}\left(\frac{M_{\rm s,0}}{M_{\rm h}}, \frac{c_{\rm s}}{c_{\rm h}}, x_{\rm c}, \epsilon\right) = \mathcal{A}_{\rm live}\left(\frac{M_{\rm s,0}}{M_{\rm h}}, \frac{c_{\rm s}}{c_{\rm h}}, x_{\rm c}, \epsilon = 1\right)$$
$$\times \left[0.4 + \mathcal{Q}\left(\frac{M_{\rm s,0}}{M_{\rm h}}, x_{\rm c}\right) \times (\epsilon - 0.2)\right] , \qquad (8.39)$$

where

$$\mathcal{Q}\left(\frac{M_{\rm s,0}}{M_{\rm h}}, x_{\rm c}\right) = 0.9 + 10^8 \left(12.84 + 3.04\, x_{\rm c} - 23.4\, x_{\rm c}^2\right)$$
$$\times \left(\frac{M_{\rm s,0}}{M_{\rm h}} - \frac{0.0077}{1 - 1.08 x_{\rm c}} - 0.0362\right)^6 \qquad (8.40)$$

This formula reproduces the semi-analytical data with an average error lower than 15 per cent, when $0.9 < x_{\rm c} < 0.3$ and $0.8 < \epsilon < 0.3$.

Comment: As a consequence of the shock heating that enhances the mass loss, the decay time on eccentric orbits can be longer than the corresponding circular one.

Light satellites, $M_{\rm s,0} \leq 0.01\ M_{\rm h}$, evolve on slightly perturbed orbits; the dynamical friction timescale in this case is two times that of the rigid satellite (CMG found a factor e longer for $M_{\rm s,0} = 0.02 M_{\rm h}$).

An important role is played by the concentration ratio as shown by the life diagrams in Fig. (8.8). These predict the final fate of a satellite with $M_{\rm s,0} = 0.01\, M_{\rm h}$, as function of the $c_{\rm s}/c_{\rm h}$ and of the orbital parameters. The fractional area in this parameter space leading to disruption, survival or decay is an estimate of the relative importance of these processes in determining the satellite's fate. Disruption due to the tidal perturbation is the fate of those satellites that initially move on close orbits despite $c_{\rm s}/c_{\rm h}$. Satellites moving along typical (plunging) cosmological orbits survive over a Hubble time only if they had a concentration higher than that of the main halo at the time of their infall.

In Fig. (8.9) we have drawn the probability distribution relative to the three final states: direct merging (by dynamical friction), which dominates at large

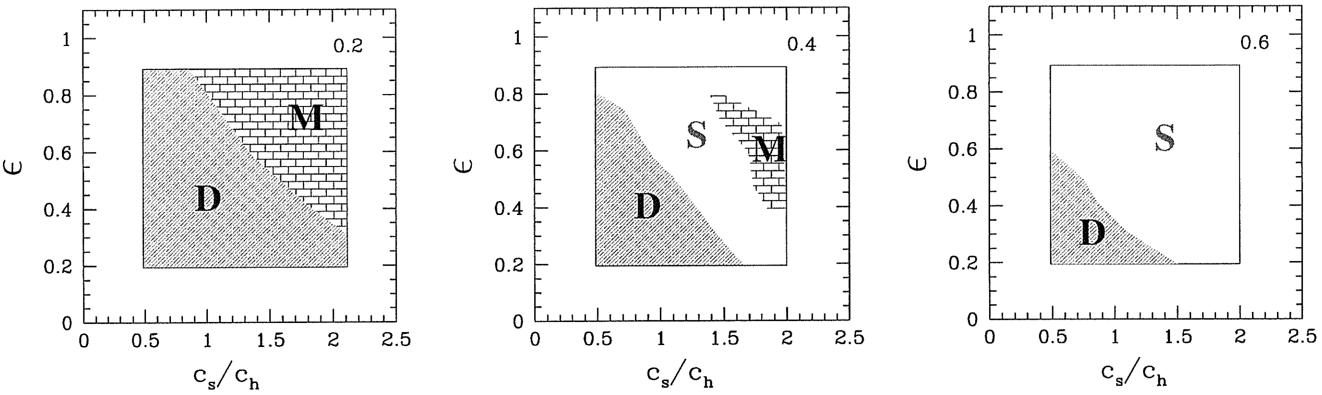

Figure 8.8: The life diagram of a satellite with $M_{s,0} = 0.01\,M_h$. Each plot is labeled with the value of $x_c(E)$. We identify the region of the parameters space where the satellite sinks to the center of the main halo (M), evaporates in the background (D) or survives (S).

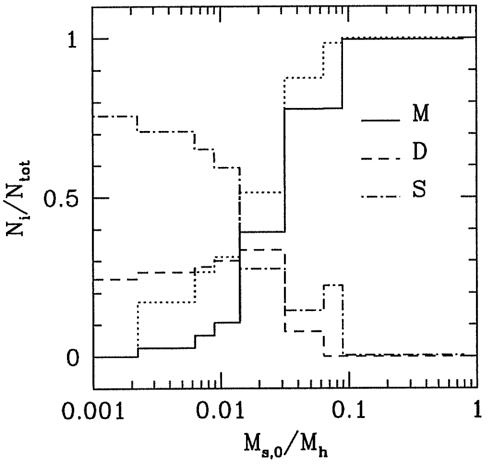

Figure 8.9: Probability distribution for the three final endpoints: merging (M), disruption (D) or survival (S) as function of the initial satellite mass. The dotted line refers to the case of a rigid satellite.

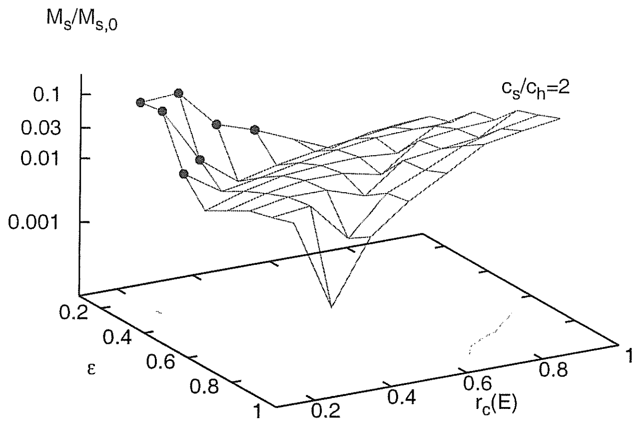

Figure 8.10: The mass of a satellite with initial $M_{s,0} = 0.01\,M_{\rm h}$ as a function of the orbital parameters after 15 Gyr. Dots identify the regions where the satellite is disrupted. The contour line on the xy plane identify the loci where $M_s/M_{s,0} = 0.1$.

masses, survival and/or disruption which is the most likely end for satellite with $M_{s,0} < 0.01\,M_{\rm h}$.

Comment: To produce the probability distribution relative to the three final states we have generated evolutionary paths (ending after a time equal to the Hubble time) for satellites starting from a uniform distribution of orbital parameters and concentrations.

Our study suggests that those satellites that survive have lost memory of their initial state: dynamical friction perturbs the orbit and tidal stripping reduces the satellite mass. In Fig. (8.10) we compute the mass of the satellites that remains after a Hubble time. The figure refers to a high concentration case, but we also extend our analysis to low concentrated satellites as shown in Fig. (8.11), where we compute the mass distribution for all the initial orbital parameters. On average, much less then 10% of the initial mass remains bound. Of course, in general circular orbits do not cause serious damages to the satellite as shock heating is less intense (an exception is represented by satellites on very tightly bound orbits). In

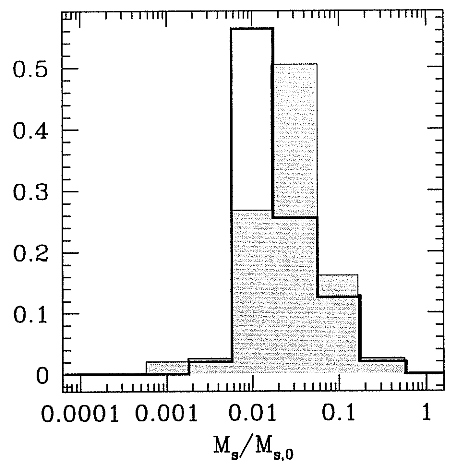

Figure 8.11: The distribution of the final mass of a satellite $M_s = 0.01 M_h$ after 15 Gyrs. Histograms show the mass distribution for a uniform distribution of orbital parameters, for two values of the concentration ratio: $c_s/c_h = 2$ (filled area) and $c_s/c_h = 1$.

Fig. (8.10), dots show the final mass just prior evaporation. As expected, radial orbits can more easily dissolve a satellite.

The strength of the orbital decay can be estimated measuring the reduction of the apocenter. In Fig. (8.12) we plot the distribution of apocenters for a satellite with $M_{s,0} = 0.01 M_h$ after 15 Gyr of orbital evolution. The strength of the drag force reduces the apocenter by a factor of two for cosmological orbits and it is not significantly affected by the concentration.

8.4.1 An rough estimate of the disruption time–scale

Because the lifetime of light satellites is mostly set by tidal disruption we estimate the disruption time. If the main halo density profile is isothermal (ISO) then GHO showed that the shock energy change is:

$$\langle \Delta E \rangle_{\rm ISO} = \left(\frac{V_{\rm h}}{R_{\rm h}}\right)^2 \frac{2\sin^2\theta_{\rm m} + 4\theta_{\rm m}^2}{6(\epsilon\, x_{\rm c})^2} A(x_\tau) \cdot r^2 \,, \qquad (8.41)$$

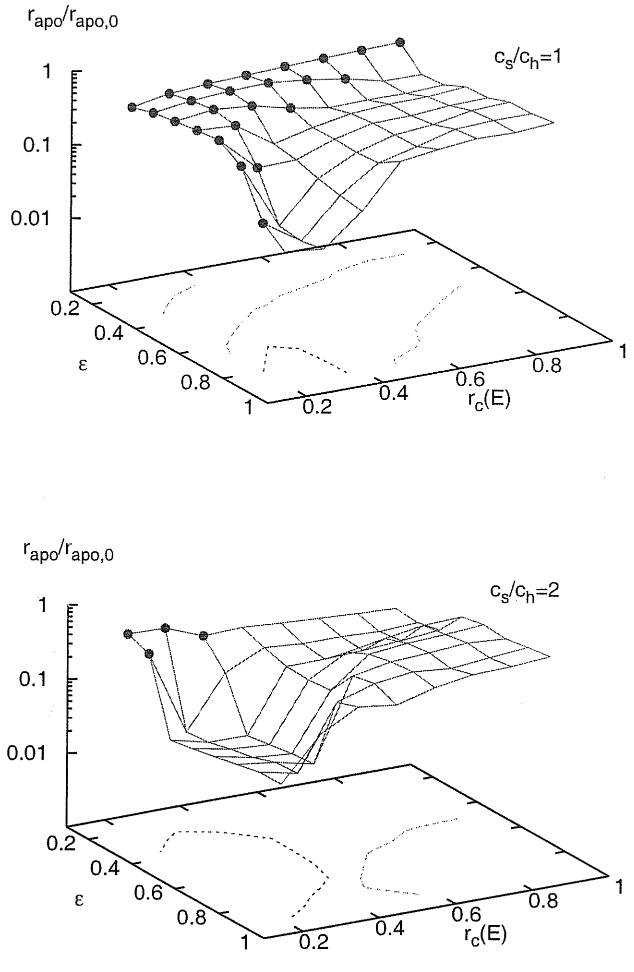

Figure 8.12: The apocenters in units of its initial value for a satellite $M_{s,0} = 0.01M_{\rm h}$ evaluated after 15 Gyr. The two plots refer to $c_{\rm s}/c_{\rm h} = 2$ (bottom) and $c_{\rm s}/c_{\rm h} = 1$ (top). Dots identify satellites that evaporate before 15 Gyrs. The contour lines on the xy plane select the regions where the relative reduction of $r_{\rm apo}$ is 0.1 (dotted lines) and 0.5 (dashed lines).

where $\theta_{\rm m}$ is the maximum value of the position angle which varies from $\pi/2$ to π. Using the orbit equation (eq. [7.2]) we can evaluate $\theta_{\rm m}$:

$$\theta_{\rm m} = 2\,\epsilon\, x_{\rm c} \int_{r_{\rm per}}^{r_{\rm apo}} \frac{dx}{x^2\sqrt{\ln(r_c/x)^2 - (r_c/x)^2\epsilon^2 - 1}} \tag{8.42}$$

and the orbital period

$$P_{\rm orb} = 2\frac{R_{\rm h}}{V_{\rm h}} \int_{r_{\rm per}}^{r_{\rm apo}} \frac{dx}{\sqrt{\ln(r_c/x)^2 - (r_c/x)^2\epsilon^2 - 1}}\ . \tag{8.43}$$

The shock in the ISO profile equals the shock of an NFW case when $c_{\rm h} = 30$. We have than:

$$\langle \Delta E\rangle_{\rm NFW} \sim \langle \Delta E\rangle_{\rm ISO} \times (0.029\, c_{\rm h} + 0.13)\ . \tag{8.44}$$

At each pericenter passage the satellite is shock heated and its radius r is reduced of a factor Δr. As a first order approximation $\Delta r \sim r/N$ where N is the number of pericenter passages necessary to destroy the satellite. Then, we have an implicit equation for N

$$N + \frac{1}{N}\sum_{i=1}^{N-1} i^2 = \frac{E_0}{\langle \Delta E\rangle_{\rm NFW}}\ , \tag{8.45}$$

where $E_0 = 0.5 G M_{\rm s}/R_{\rm s,0}$. The disruption time can be written as:

$$t_{\rm dis} \sim P_{\rm orb}\cdot N\ . \tag{8.46}$$

This formula provides a simple estimate of the disruption time valid on cosmological relevant orbits with a precision of the 20 per cent.

8.4.2 Cosmological examples

Now, we apply our analysis to some cosmologically relevant examples. For all cases, the initial orbital parameters are chosen as $\epsilon = 0.6$ and $x_{\rm c}(E) = 0.5$.

We discuss in detail the evolution of different satellites which orbit in cluster-like and galaxy-like haloes. The cluster halo is a Coma-like cluster with mass $M_{\rm h} = 5\cdot 10^{15} M_\odot$ and concentration $c_{\rm h} = 3.44$.

Group in Coma

We consider a group-like satellite of mass $M_{s,0} = 3 \cdot 10^{13} M_\odot$ and $c_s = 7.5$ which enters the Coma-like halo at $z = 0.5$. In a ΛCDM Cosmology it evolves for ~4.8 Gyr inside the halo until $z = 0$.

As suggested by the high value of c_s/c_h the satellite is not disrupted. Since $M_{s,0} = 0.006 M_h$ the orbit is stable and, with this choice of the initial orbital parameters, the satellite evolves for ~ 1.5 P_{orb}. The final apocenter is $r_{apo} \simeq 0.85 r_{apo,0}$ and its final mass is $M_s = 7.2 \cdot 10^{12} M_\odot$.

Milky Way in Coma

A Milky Way-like satellite has mass $M_{s,0} = 10^{12} M_\odot$ and $c_s = 10.44$. If it enters the Coma-like halo at $z = 0.5$ it evolves for ~ 1.5 P_{orb}. The orbit remains almost unperturbed ($r_{apo} \simeq 0.99 r_{apo,0}$) as strength of the drag force is extremely weak since $M_{s,0} = 0.0002 M_h$. Due to the extremely high relative concentration, $c_s/c_h \simeq 3$ the satellite does not evaporate and its final mass ad $z = 0$ is $M_s = 2.5 \cdot 10^{11} M_\odot$

Large Magellanic Cloud in Milky Way

A Large Magellanic Cloud halo has $M_{s,0} = 10^{11} M_\odot$ and $c_s = 11.9$. As expected due to its relative high mass, the satellite merges with the Milky Way in ~ 4 Gyr. Before merging, the satellite loses 97% of its mass that is dispersed in the Milky Way Halo.

Dwarf in Milky Way

We consider a Dwarf-like satellite of mass $M_{s,0} = 5 \cdot 10^9 M_\odot$ and concentration $c_s = 13.6$. If it enters the Milky Way halo at $z = 0.5$ it evolves on an almost unperturbed orbit for ~ 2 P_{orb} and its final mass is $M_s = 2 \cdot 10^8 M_\odot$.

Dwarf in Milky Way at high redshift A Dwarf-like satellite enters a Milky Way halo at $z = 2$, when the Milky Way has mass $M_h = 10^{11} M_\odot$ and concentration $c_h = 6.3$. The satellite has $M_{s,0} = 5 \cdot 10^9 M_\odot$ and $c_s = 13.6$. The dwarf evolves for ~11 Gyr. Due to its low relative concentration it loses 99% of its initial mass during the first orbital period; its orbit then becomes stable ($r_{apo} \simeq 0.36 r_{apo,0}$).

Note that we are not accounting for the evolution of the main halo which accretes mass during the 11 Gyr.

8.5 Summary and discussion

The fate of substructures is a complicated process to model. Our simple analysis shed light on the fate of satellites, whether they merge, evaporate or survive under the simultaneous action of dynamical friction and tidal mass loss.

In particular we provide a simple analytical recipe for the decay time of a satellite on the basis of its mass, at the time of infall into the main halo (here we refer to typical cosmological orbits):

- High mass satellites ($M_{\rm s,0} > 0.1 M_{\rm h}$) sink rapidly toward the center of the main halo without significant mass dispersal: the dynamical friction timescale for a rigid satellite (eq. [8.33]) gives the correct timescale of merging.

- For satellites of mass $0.01 M_{\rm h} < M_{\rm s,0} < 0.1 M_{\rm h}$ dynamical friction is still strong and drives the satellite toward the center where tidal mass loss becomes severe. Low concentration satellites are disrupted, while high concentration satellites, severely pruned by the tidal field, survive with masses $0.01 M_{\rm s,0}$, and settle into inner orbits with a typical reduction of the apocenters of a factor ~ 0.1 relative to the initial value. The dynamical friction timescale for these satellites is longer than for the rigid counterpart, and is given by eq. (8.35).

- Light satellites with mass $M_{\rm s,0} < 0.01 M_{\rm h}$ are almost unaffected by dynamical friction which is operating on a rather long timescale. Mass loss by the tidal field, which is not severe on these cosmological orbits, further stabilizes the orbit.

- Low concentration satellites below $0.1 M_{\rm h}$ can be disrupted by tides before their orbital decay is complete. Comparison of the dynamical friction timescale and the disruption timescale as provided by eq. (8.46), allows to describe the actual lifetime of satellite haloes.

Our results have several implications:

1- Because of the combined action of stripping and dynamical friction, a primary halo at $z = 0$ will host preferentially satellites with mass $M_{\rm s}/M_{\rm h} \ll 0.01$ as the heavier ones would have been accreted or/and dispersed in the background, leaving a "depression" in the mass function of substructure above 0.01 $M_{\rm s}/M_{\rm h}$,(of course we are neglecting effects due to the evolution of the main halo itself). This feature should be more evident in Milky Way-size halos than in cluster halos as in the former bound satellites had more time to evolve.

2- Since the destructive power of the tidal field (and in particular of tidal shocks) depends sensitively on the degree of circularity of the satellite's orbit, a large galaxy halo like that of the Milky Way ($> 10^{12} M_{\odot}$) should host satellites moving preferentially on circular orbits as a consequence of the selective action of the tidal field. Also, because dynamical friction seems unable to render the satellites' orbit circular, the low eccentricities should have been already present as initial conditions. This "selection effect" will be extremely weak for smaller satellites (below $0.01 - 0.001 M_{\odot}$) because their orbit barely decays and thus will have in general long survival times (only low concentration satellites could disappear quickly but they are not common in CDM models; see e.g. Eke, Navarro & Steinmetz 2001; Bullock et al.2001). This mass regime corresponds to that of the dwarf spheroidal satellites of the Milky Way. On the other end, the Magellanic Clouds, the dwarf elliptical satellites of M31 and perhaps the dwarf spheroidal Fornax are all massive enough to fit in the intermediate regime where destruction is still possible; thus these galaxies could have survived because their host haloes had nearly circular orbits. In the case of the Magellanic Clouds a nearly circular orbit is indeed measured (Kroupa & Bastian 1999). There is, however, at least one caveat to this interpretation, namely that both the dwarf ellipticals of M31 and the Magellanic Clouds could be dense enough to survive shocks on even very eccentric orbits (Mayer et al. 2001b). Only when all the orbits of the satellites will be accurately determined we will know whether eccentricity or internal structure was more important in determining their survival.

3- A key result of our analysis, and one that is in agreement with the high resolution cosmological simulations from which the initial orbits were drawn, is that the inner, most bound part of small satellites as concentrated as expected in CDM models (Eke, Navarro & Steinmetz 2001; Bullock et al. 2001) survive for timescales comparable to or longer than the age of the Universe. This residual has a size corresponding to a few percents of the initial virial radius; this is comparable to the scale of the baryonic component in galaxies, so we can argue that galaxies will mostly survive within the main halo. This result is also confirmed by high-resolution SPH simulations of the formation of Milky Way-like galaxy (Governato et al. 2002). Indeed, dissipation could make the inner part of the haloes even more robust against tides (Navarro & Steinmetz 2001). On the other end, additional tidal shocks occurring during encounters between substructures, i.e. galaxy harassment (Moore et al. 1996, 1998), might have a counteracting effect and could actually increase mass loss. However, detailed simulations of this mechanism have shown that only very fragile, LSB-type galaxies would be severely damaged by harassment (Moore et al. 1999); halo profiles of these galaxies likely correspond to the low concentration satellites studied in this Chapter (Van den Bosch & Swaters 2001) which we have shown are easily disrupted even by the tides of the primary halo alone. Thus, adding harassment would only accelerate the disruption of a few satellites while not affecting the survival of the majority of them which, in CDM models, have high concentrations. Hence the picture emerging from the life diagrams of the satellites shown in this Chapter is robust. Satellites close to disruption at the present time, like Sagittarius in the Milky Way subgroup, must have been much bigger in the past in order for dynamical friction to drag them to an inner orbit where dissolution can easily take place; alternatively they could have entered the halo at fairly high redshift, which would place them naturally on an inner, tightly bound orbit (Mayer et al. 2001b). In clusters, dwarf galaxies cannibalized by giant CD galaxies might also trace an early population.

4- Satellites infalling at redshift one or lower in the main haloes will complete several orbits and eventually undergo morphological changes by tidal stirring (Mayer et al. 2001a,b) and harassment (Moore et al. 1996, 1998). These will produce diffuse streams of stars while they are orbiting (Helmi & White 1999;

Johnston, Sigurdsson & Hernquist 1999), contributing to the build up of an extended stellar halo population. Such population should be present out to more than 200 kpc in the Milky Way halo, as the plunging orbits of satellites seen in cosmological simulations go this far out. On the contrary, a less extended stellar halo should be expected if dynamical friction were more efficient in dragging satellites to the center. The amount of stellar halo substructure out to large distances could thus reveal the original mass function of observed dwarf spheroidal galaxies in the Local Group. Components decoupled in their kinematics as well as in the metallicity and age of their stars should be present, but tracking such properties might be a daunting task observationally if enough phase mixing occurs (Ibata et al. 2001a,b); however, while in the inner halo fast orbital precession and heating by other clumps might blur the streams, the phase space distribution of the outer halo material should still carry the memory of the initial orbits of the satellites (Mayer et al. 2002).

Chapter 9

Conclusions

So far, so good... so what?

D. Mustaine

In this thesis I have tackled the problem of structure formation and evolution in a hierarchical cosmological scenario from a semi–analytical point of view.

Since dark matter traces the visible matter and forms the potential wells where galaxies arise and evolve, we have focused on the problem of the formation and evolution of dark matter haloes, which only involve pure gravitational processes. In doing this, we have considered two alternative ways of proceeding.

1- As a first attempt, I have developed a semi-analytical code able to generate merging histories of haloes based on the extended Press & Schechter formalism. Such Monte Carlo procedure is a standard tool to construct synthetic catalogs of haloes (Somerville & Primack 2000; Cole et al. 2000), and provides a full description of the formation path of a given halo. However, this semi–analytical approach shows discrepancies at a level of a factor of 2 or more when its outputs (e.g. mass function and conditional mass function) are compared with quantities measured from N-body simulations. This can lead to errors both in the number of progenitor haloes and in the halo conditional mass function. Moreover, this procedure does not produce the information regarding the position, the spin, and the internal structure of the haloes. For this reason, we were not completely satisfied and tried a different approach.

2- To overcome the limitations of the extended Press & Schechter procedure, we have developed the PINOCCHIO code (Monaco et al. 2002; Monaco, Theuns & Taffoni 2002). We have followed a two step approach which mimics the hierarchical build-up of haloes through accretion and merging. The first step identifies orbit-crossing as the instant when a mass element undergoes collapse. Orbit-crossing is computed numerically by applying the local ellipsoidal collapse approximation of the full Lagrangian perturbative expansion. In the second step, the collapsed particles are grouped into disjoint haloes, using an algorithm similar to the one used to identify haloes in N-body simulations. Loosely speaking, a particle accretes onto a halo if it is sufficiently close to it at the moment of collapse. We have used the Zel'dovich (1972) approximation to compute the positions of haloes and particles. This second step automatically determines the full merger history of haloes and requires negligible computer time. Compared to simulations, the first step determines when a simulation particle enters a high-density region, whereas the second one identifies the haloes.

   The PINOCCHIO code generates catalogues of dark matter haloes with known mass, position, velocity, merging history and angular momentum. I have demonstrated that predictions from our code are very accurate when compared with the results of large N-body simulations covering a range of cosmological models, box sizes and numerical resolutions (Taffoni,Monaco & Theuns 2002).

Both PINOCCHIO and the Merger Tree code based on the extended Press & Schechter formalism are only aimed at identifying the haloes, while do not attempt to compute their internal properties or their substructure. A post–run analysis is required to assign a density profile to each halo and to follow the dynamical evolution of its satellites.

The evolution of substructures of dark matter haloes has increasingly become an extremely hot topic since, increasing computer power and code performances have recently allowed to carry out extremely high-resolution simulations that can resolve the evolution of substructures (Ghigna et al. 1998, 2000; Mayer et al. in preparation). These represent the new ground where CDM models have been tested and their predictions compared to observations. These simulations however remain computationally expensive and usually only one system at the time can be simulated down to very small scales. On the other hand, resolving the mass function of substructures in depth is an important issue, given the problem of overabundance of satellites (Moore et al. 1999; Klypin et al. 1999). Such mass function can be viewed as the convolution of the mass function of satellites before their entrance in to the main halo, with an evolutionary filter function that depends on the dynamical mechanisms that govern the evolution of substructures. Under the assumption that haloes are well described by a Navarro, Frank & White (1996) density profile, I have analysed these dynamical mechanisms and developed a semi–analytical code which is able to predict if a satellite of given mass, orbital eccentricity and infall redshift, will merge, evaporate or survive under the simultaneous action of dynamical friction and tidal mass loss (Taffoni et al. 2002).

The code combines a model to compute the dynamical friction based on the Chandrasekhar formalism with a novel recipe which accounts for the satellite mass loss due to tidal stripping. The ability of this code to follow the orbital decay and the satellite mass evolution has been successfully tested by comparing its results with those derived from a set of N-body simulations.

I have also provided useful analytical prescriptions for the decay and disruption rates of satellites within haloes which considerably improve the predictive power of PINOCCHIO and the semi–analytical codes based on the extended Press & Schechter formalism.

As a final remark about the future prospects of this research I would like to stress a number of crucial points.

- Due to its ability to describe the evolution of clustering of haloes as a function of mass, PINOCCHIO is perfect to be combined with semi-analytical models for galaxy formation (White & Frenk 1991, Kauffmann, White & Guiderdoni 1993, Cole et al 1994, Somerville & Primack 1999). PINOCCHIO can be used to reliably generate mock galaxy catalogs, with the correct evolution of galaxy clustering built-in, while requiring orders of magnitude less computational time than numerical simulations.

- PINOCCHIO easy and accurate production of large halo catalogues can be extremely useful when interpreting data and estimating errors from galaxy or galaxy cluster surveys, as it is for instance the case of galaxy bias (Diaferio et al. 1999; Benson et al. 2000), power-spectra (e.g. Efstathiou & Moody 2001), determination of the shear from weak lensing measurements (van Waerbeke et al. 2000, Wittman et al. 2000, Bacon, Refregier & Ellis 2000, Kaiser, Wilson & Luppino 2000) or the study of the intrinsic galaxy alignments (Crittenden et al. 2001, Brown et al. 2000).

- By using our analytical prescriptions for the decay and disruption rates, we can approach the substructure problem in a statistical way which is orders of magnitude faster than with N-body simulations; as an example we are currently working on the possibility of exploring a large number of dynamical histories of satellite haloes by randomly varying their orbital and structural parameters in the range which is typical of cold dark matter cosmogonies (Taffoni et al. in prep). In this thesis I have presented a first attempt which considers uniform distributions for the above parameters. Clearly, the time dependent potential of the growing primary halo, whether it is a galaxy or a cluster, is an additional ingredient that only simulations can incorporate and which could in principle affect the orbital dynamics of the satellites. The latter limitation however can be partially overcome by using the merger tree extracted from PINOCCHIO.

In conclusion, while a lot of work as been done, much more still has to come.

Bibliography

[1] Antonuccio-Delogu V., Becciani U., Pagliaro A., van Kampen E., Colafrancesco S., Germanà A., Gambera M., 2002, MNRAS, 332, 7

[2] Audit E., Teyssier R., Alimi J.M.,1997, A&A, 325, 439

[3] Balbi, A., 2001, in Cosmology and Particle Physics, ed. Durrer, Garcia-Bellido, Shaposhnikov, AIP Conference Proceedings, Vol. 555

[4] Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15

[5] Barnes J., Efstathiou G. P., 1987, ApJ, 319, 575

[6] Benson A.J., Cole S., Frenk C.S., Baugh C.M., Lacey C.G,. 2000, MNRAS, 311, 793

[7] Bertschinger E., Jain B., 1994, ApJ, 431, 486

[8] Binney, J., Tremaine, S. 1987, Galactic Dynamics, Princeton University Press, Princeton

[9] Bode P., Bahcall N.A., Ford E.B., Ostriker J.P., 2001, ApJ, 551, 15B

[10] Bond J.R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

[11] Bond J.R., Myers S.T., 1996a, ApJS, 103, 1

[12] Bond J.R., Myers S.T., 1996b, ApJS, 103, 41

[13] Bontekoe T.R., van Albada T.S., 1987, MNRAS, 224, 349

[14] Borgani S., Coles P., Moscardini L., 1994, MNRAS, 271, 223

[15] Borgani S., Rosati P., Tozzi, P.,2001, ApJ, 561, 13B

[16] Bouchet F., 1996, in Dark Matter in the Universe, ed. S. Bonometto et al. IOS, Amsterdam

[17] Brainerd T.G., Goldberg D.M., Villumsen J.V., 1998, ApJ, 502, 505

[18] Buchert T., 1992, A&A, MNRAS, 254, 729

[19] Buchert T., 1996, in Dark Matter in the Universe, ed. S. Bonometto et al. IOS, Amsterdam

[20] Buchert T., Ehlers J., 1993, MNRAS 264, 375

[21] Buchert T., Melott, A., L., Weiss, A., G., 1994, A&A, 288, 349B

[22] Bullock J.S., Dekel A., Kolatt T.S., Kravtsov A.V., Klypin A.A., Porciani A., Primack J.R., 2001, ApJ, 555, 240

[23] Bullock J.S., Kolatt T. S., Sigad Y., et al., 2001, MNRAS, 321, 559

[24] Bullock J.S., Kolatt T.S., Sigad Y., et al., 2001, MNRAS, 321, 559

[25] Bullock J.S., Kravtsov A.V., Weinberg, D.H., 2000, ApJ, 539, 517

[26] Carlberg R.G., 1994, ApJ, 433, 468

[27] Carr B.J., Lacey C.G., 1987, ApJ, 316, 23

[28] Catelan P., 1995, MNRAS, 276, 115

[29] Catelan P., Lucchin F., Matarrese S., Porciani C., 1998, MNRAS, 297, 692

[30] Catelan P., Theuns T., 1996a, MNRAS, 282, 436

[31] Catelan P., Theuns T., 1996b, MNRAS, 282, 455

[32] Cavaliere A., Colafrancesco S., Menci N., 1992, ApJ, 392, 41

[33] Cavaliere A., Menci N., Tozzi P., 1996, ApJ, 464, 44

[34] Chandrasekhar S. 1943, ApJ, 97, 255

[35] Colafrancesc, S., Lucchin F., Matarrese S., 1989, ApJ, 494, 96

[36] Cole S., 1991, ApJ, 367, 45

[37] Cole S., Kaiser N., 1988, MNRAS, 233, 637

[38] Cole S., Lacey C.G., 1996, MNRAS, 281, 716

[39] Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

[40] Coles P., Melott A.L., Shandarin S.F., 1993, MNRAS, 260, 765

[41] Colberg J.M., White S.D.M., Yoshida N., et al, 2000, MNRAS, 319, 209

[42] Colpi M., 1998, ApJ,502,167C

[43] Colpi M., Pallavicini A., 1998, ApJ, 502, 150C

[44] Colpi M., Mayer L., Governato F., 1999, ApJ, 525, 720

[45] Cora S.A., Muzzio J.C., Vergne M.M., 1997, MNRAS, 289, 253

[46] Couchman H., Thomas P., Pearce F., 1995, Apj, 452, 797

[47] Dikaiakos M. & Stadel J., 1996, Conf. Proc. of the International Conference on Supercomputing (New York: Assoc. for Computing Machinery)

[48] de Bernardis P., Ade P.AR., Artusa R. et al. 1999, NewAR, 43, 289

[49] de Bernardis P., Ade P.A.R., Bock J.J., et al. 2000, Natur 404, 955D

[50] Domínguez-Tenreiro R., Gómez-Flechoso M.A., 1998, MNRAS, 294, 465

[51] Dubinski J., Carlberg R.G., 1991, ApJ, 378, 496

[52] Duncan M.J., Farouki R.T., Shapiro S.L., 1983, ApJ, 271, 22

[53] Efstathiou G.P., Frenk C.S., White S.D.M., Davis M., 1988 MNRAS, 235, 715

[54] Eisenstein D.J., Hut P., 1998, ApJ, 498, 137

[55] Eke V.R., Navarro J.F., Steinmetz M., 2001, ApJ, 554, 114

[56] Freedman W.L., Madore B.F., Gibson B.K., et al. 2001, ApJ, 553, 47

[57] Frenk C.S., White S.D.M., Davis M., Efstathiou G., 1988, ApJ, 327, 507

[58] Fukushige T., & Makino J., 2001, ApJ, 557, 533

[59] Gardner J.P., 2001, ApJ, 557, 616

[60] Gelb J. M., Bertschinger E., 1994, ApJ, 436, 467

[61] Ghigna S. Moore B., Governato F., Lake G., Quinn T., Stadel J. 1998, MNRAS, 300, 146

[62] Ghigna S.,Moore B., Governato F.,Lake G., Quinn T., Stadel J., 2000, ApJ, 544, 616

[63] Gnedin O.Y.,Hernquist L., Ostriker J.P., 1999, ApJ, 514, 109

[64] Gnedin O.Y., Lee H.M., Ostriker J.P.,999, ApJ, 522, 935

[65] Gnedin O.Y., Ostriker J.P., 1997, ApJ, 474, 223

[66] Gnedin O.Y. & Ostriker J.P., 1999, ApJ, 513, 626;

[67] Governato F., Babul A., Quinn T., Tozzi P., Baugh C.M., Katz N., Lake G. 1999, MNRAS, 307, 949

[68] Governato F., Mayer L., Wadsley J., Gardner J.P., Willman B., Hayashi E., Quinn T., Stadel J., Lake G., astro-ph/0207044

[69] Hanami H., 2001, MNRAS, 327, 721H

[70] Hanany S., Ade P., Balbi A., et al. 2001, AAS, 199, 3403H

[71] Hayashi E., Navarro F.J., Taylor J.E., Stadel J., Quinn T., astro-ph/0203004

[72] Heavens A., Peacock J., 1988, MNRAS, 232, 339

[73] Helmi A., White S.D.M., 1999, MNRAS, 307, 495

[74] Hoyle F., 1949, in Burgers J.M., van der Hulst H.C., eds., Problems of Cosmical Aerodynamics, Central Air Document Office, Dayton. p. 195

[75] Huang S., & Carlberg, R., G., 1997, ApJ, 480, 503

[76] Huss A., Jain B., Steinmetz M., 1999, ApJ, 517, 64

[77] Ibata R., Irwin M., Lewis G., Ferguson A.M.N., Tanvir N., 2001, Nature, 412, 49

[78] Ibata R., Lewis G.F., Irwin M., Totten E., Quinn T., 2001, ApJ, 551, 294

[79] 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

[80] Jing Y.P., 1998, Apj, 503,9

[81] Jing Y.P., 1999, Apj, 515,45

[82] Jing Y.P., Suto Y., 2000, ApJ, 529L, 69

[83] Johnston K.V., Sigurdsson S., Hernquist L., 1999, MNRAS 302, 771

[84] Kaiser N., 1984, ApJ, 284L, 9

[85] Kaplinghat M., Turner M.S., 2001, PhRevLett, 86, 385

[86] Katz N., Quinn T., Gelb J.M., 1993, MNRAS, 265, 689

[87] Kauffmann G., White S.D.M., Guiderdoni B., 1993, MNRAS, 264, 201

[88] Klypin A., Gottlober S., Kravtsov A.V., Khokholov A.M., 1999, Apj, 516, 530

[89] Klypin A., Kravtsov A.V., Valenzuela O., Prada F., 1999, ApJ, 522, 82

[90] Kolatt T.S., Bullock J.S., Sigad Y., Kravtsov A.V., Klypin A.A, Primack J.R., Dekel A., 2000, MNRAS submitted (astro-ph/0010222)

[91] Kravtsov A.V., Klypin A.A., 1998, in Large Scale Structure: Tracks and Traces, eds Mueller, Gottloeber,Muecket, Wambsganss

[92] Kroupa P., Bastian U., 1997, New Astronomy, 2, 77

[93] Lacey C., Cole S. 1993, MNRAS, 262, 627

[94] Lacey C., Cole S., 1994, MNRAS, 271, 676

[95] Lanzoni B., Mamon G.A., Guiderdoni B., 2000, MNRAS, 312, 781

[96] Lemson G., Kauffmann G., 1998, MNRAS, 302, 111

[97] Lee J., Shandarin S.F., 1998, ApJ, 500, 14

[98] Lewis G.F., Babul A., Katz N., Quinn T., Hernquist L., Weinberg D.H., 2000, ApJ, 536, 623

[99] Linde A. D., 1990, Particle Physics and Cosmology, Harwood Academic, Chur, Switzerland

[100] Lineweaver C.H., Tenorio L., Smoot G.F., Keegstra P., Banday A.J., Lubin P., 1996, ApJ, 470, 38L

[101] Maller A.H., Dekel A., Somerville R.S., 2002, MNRAS, 329, 423

[102] Manrique A., Salvador-Solé E., 1995, ApJ, 453, 6

[103] Maoz E., 1993, MNRAS, 263, 75

[104] Mather J.C., Fixsen D.J., Shafer R. A., Mosier C., Wilkinson D.T., 1999, ApJ, 512, 511

[105] Mayer L., Governato F., Colpi M., Moore B., Quinn T., Wadsley J., Stadel J., Lake G., 2001, ApJ, 547, 123

[106] Mayer L., Governato F., Colpi M., Moore B., Quinn T., Wadsley J., Stadel J., Lake G., 2001, ApJ, 559, 754

[107] Melott A.L., Buchert T., Weib A.G., 1995, A&A, 294, 345M

[108] Melott A.L., Pellman T.F., Shandarin S.F., 1994, MNRAS, 269, 626M

[109] Mo H. J.,Jing, White S.D.M., 1997, MNRAS, 284, 189

[110] Mo H.J., White S.D.M., 1996, MNRAS, 282, 347

[111] Mohr J.J.; Mathiesen, B., Evrard, A. E., 1999, ApJ, 517, 627M

[112] Monaco P., 1995, ApJ, 447, 23

[113] Monaco P., 1997a, MNRAS, 287, 753

[114] Monaco P., 1997b, MNRAS, 290, 439

[115] Monaco P., 1998, Fund. Cosm. Phys., 19, 153

[116] Monaco P., Theuns T., Taffoni G., Governato F., Quinn T., Stadel J., 2002, ApJ, 564, 8

[117] Monaco P., Theuns T., Taffoni G., 2001, MNRAS,

[118] Moore B., Katz N., Lake G. 1996, ApJ, 457, 455

[119] Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J ., Tozzi P., 1999, ApJ, 524, 19

[120] Moutarde F., Alimi J.M. ,Bouchet F.R., Pellat R., Ramani A., 1991, ApJ 382, 377

[121] Myers S.T., Baker J.E., Readhead A.C.S., Leitch E.M., Herbig T., 1997, ApJ, 485, 1

[122] Naab T., Burkert A., Hernquist L., 1999, ApJ, 523, 133

[123] Navarro J.F., Frenk C.S., White S.D.M., 1996, ApJ, 462, 563

[124] Navarro J.F., Frenk C.S., White S.D.M., 1997, ApJ, 490, 493

[125] Navarro J.F., Steinmetz M., 2000, ApJ, 528, 607

[126] Nagashima M., Gouda N., 1998, MNRAS, 301, 849

[127] Okamoto T., Habe A., 1999, 1999, ApJ, 516, 591

[128] Ostriker J.P., Spitzer L.J., Chevalier R.A., 1972, ApJ, 176, 51

[129] Peacock J.A., 1999, Cosmological Physics, Cambridge University Press, Cambridge

[130] Peacock J.A., Heavens A.F., 1985, MNRAS, 217, 805

[131] Peacock J.A., Heavens A.F., 1990, MNRAS, 243, 133

[132] Peebles P.J.E., 1980, The Large Scale Structure of the Universe (Princeton: Princeton Univ. Press)

[133] Peebles P.J.E., 1993, Principles of Physical Cosmology, Princeton University Press, Princeton

[134] Penzias A.A., Wilson R.W., 1965, ApJ, 142, 419

[135] Perlmutter S., Aldering G., Goldhaber G., et al., 1999, ApJ, 517, 565

[136] Percival W.J.; Baugh C.M., Bland-Hawthorn J., et al., 2001, MNRAS, 327, 1297P

[137] Pettini M., Bowen D.V., 2001, ApJ, 560, 41

[138] Porciani C., Catelan P., Lacey C., 1999, Apj, 513, 99

[139] Porciani C., Matarrese S., Lucchin F., Catelan P., MNRAS, 298, 109

[140] Porciani C., Dekel A., Hoffmann Y., 2002a, MNRAS, 332, 325

[141] Porciani C., Dekel A., Hoffmann Y., 2002b, MNRAS, 332, 339

[142] Press W.H., Schechter P., 1974, ApJ, 187, 425

[143] Quinn P.J., Goodman J., 1986, ApJ, 309, 472

[144] Rodrigues D.D.C., Thomas P.A., 1996, MNRAS, 282, 631

[145] Scoccimarro R., Sheth R.K., 2002, MNRAS, 329, 629

[146] Sahani V., Coles P., 1995, Phys. Rep., 262, 1

[147] Séguin P., Dupraz C., 1994, A&A, 290, 709

[148] Shandarin S.F., Zel'dovich Ya.B.,1989, Rev. Mod. Phys., 61, 185

[149] Sheth R.K., Lemson G., 1999a, MNRAS, 304, 767

[150] Sheth R.K., Lemson G., 1999b, MNRAS, 305, 946

[151] Sheth R.K., Mo H.J., Tormen G., 2001, MNRAS, 323, 1

[152] Sheth R.K., Tormen G., 1999, MNRAS, 308, 119

[153] Sheth R.K., Tormen G., 2002, MNRAS, 329, 61S

[154] Somerville R.S., Kolatt T.S. 1999, MNRAS, 305, 1

[155] Somerville R.S., Lemson G., Kolatt T.S., Dekel A., 2000, MNRAS, 316, 479

[156] Somerville R.S., Primack J. R. 1999, MNRAS, 310, 1087

[157] Spitzer L. Jr. 1987, Dynamical Evolution of Globular Clusters (Princeton: Princeton Univ. Press)

[158] Taffoni G., Mayer L., Colpi M., Governato F., 2001, in Chemical Enrichment of Intracluster and Intergalactic Medium, ed. Fusco-Femiano and Matteucci, ASP Conference Proceedings Vol. 253.

[159] Taffoni G., Mayer L., Colpi M., Governato F., 2002, MNRAS submitted

[160] Taffoni G., Monaco P., Theuns T., 2002, MNRAS, 333, 623

[161] Taylor J., Babul A., 2001, ApJ, 559, 716T

[162] Theuns T., Leonard A., Efstathiou G., Pearce F.R., Thomas P.A., 1998, MNRAS, 301, 478

[163] Tóth G., Ostriker J.P., 1992, Apj, 389, 5

[164] Tormen G., 1997 , MNRAS, 290, 411

[165] Tormen G., Bouchet F.R., White S.D.M., 1997, MNRAS, 286, 865

[166] Tormen G., Diaferio, A., Seyer, 1998 , MNRAS, 299, 728

[167] Tremaine S., Weinberg M.D., 1984, MNRAS, 209, 729

[168] van den Bosch F.C., Lewis G., Lake G., Stadel J., 1999, Apj 515, 50

[169] van den Bosch F.C., Swaters R.A., 2001, MNRAS, 325, 1017

[170] Ueda H., Shimasaku K., Suginohara T., Suto Y., 1994, PASJ, 46, 319

[171] Vitvitsaka M., Klypin A.A., Kravtsov A.V., Bullock J.S., Wechsler R.H., Primack J.R., astro-ph/0105349

[172] Velázquez H., & White S.D.M., 1999, MNRAS, 304, 254

[173] Weinberg S., 1972, Gravitation and Cosmology. Wiley, New York

[174] Weinberg M.D., 1986, ApJ, 300, 93

[175] Weinberg M.D., 1994a, ApJ, 108, 1398

[176] Weinberg M.D., 1994b, ApJ, 108, 1403

[177] Weinberg M.D., 1994c, ApJ, 108, 1414

[178] Weinberg M.D., 1995 ApJ, 455, 31

[179] White S.D.M., 1976, MNRAS, 174, 467

[180] White S.D.M., 1978, MNRAS, 184, 185

[181] White S.D.M., 1984, ApJ, 286, 38

[182] White S.D.M., 1996, in Schaeffer R. et al., eds, Cosmology & Large-scale structure, Proc. 60th Les Houches School, Elsevier, p. 349

[183] White S.D.M., Frenk C.S., 1991, ApJ, 379, 52

[184] White S.D.M., Rees M.J. 1978, MNRAS, 183, 341

[185] White S.D.M., Silk J., 1979, ApJ, 232, 1

[186] Zel'dovich YA. B., 1970, Astrofizika, 6, 319 (translated in Astrophysics, 6, 164 [1973])