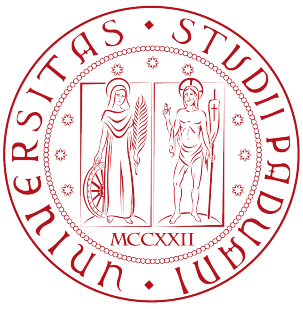

Sede Amministrativa: Università degli Studi di Padova
Dipartimento di Fisica e Astronomia G. Galilei

Corso di Dottorato di Ricerca in: Astronomia
Ciclo XXIX

Formation of dark matter halos. Statistics and dynamics of the ellipsoidal collapse model.

Coordinatore: Ch.mo Prof. Giampaolo Piotto
Supervisore: Ch.mo Prof. Giuseppe Tormen
Co-Supervisore: Ch.mo Prof. Ravi K. Sheth

Dottorando: Pietro Gambarotto

Contents

Abstract

In the standard cosmological model the formation of cosmic structures is described by the collapse of density perturbations. These perturbations have grown by the action of gravity from small Gaussian initial fluctuations. In a ΛCDM Universe cosmic structures formation is driven by the collapse of dark matter, leading to the creation of virialized systems, called dark matter halos. Baryonic matter follows the dark matter potential wells, where it cools transforming its kinetic energy into thermic energy, eventually forming visible systems, stars and galaxies (White and Rees 1978; Blumenthal et al. 1984). Structures then grow hierarchically, from smaller to larger ones. Therefore, halos containing large galaxies are formed through repeated merger of smaller halos. The collapse and subsequent formation of dark matter halos is due only to gravity, so it is simpler to explain compared to the formation of stars and galaxies themselves. However, the problem involves a high number of fluid-like particles, and does not admit an analytical solution. Therefore it is bet studied through use of N-body numerical simulations. This technique allows to evolve a large number of particles subject only to their mutual gravitational interaction. In this way we can simulate a region of Universe, analyse the motion of particles and the formation of bound structures.

The present thesis has the following structure:

- **Chapter 1**: We present the standard cosmological model for the formation of cosmic structures, and briefly describe the statistical properties of linear perturbation fields and the growth of linear perturbation in the light of *Jeans' theory*.

- **Chapter 2**: We describe the analytical approaches to the study of perturbations in the non linear regime. First of all we present the Zel'dovich approximation in case of a quasi-linear regime, and two main dynamical models: spherical and ellipsoidal collapse. Later we describe two analytic approaches to determine halo statistics starting from the initial fluctuation field: the excursion sets approach and the peaks

formalism. We also briefly describe some attempts to merge the two approaches together.

- **Chapter 3**: We describe the main approach to the study of non linear and bound structures: N-body simulations. We also describe the main features of simulations employed in this work. Finally we describe the main properties of dark matter halos, focussing on the contributions of N-body simulation to the study of halo properties.

- **Chapter 4**: Firstly we describe different halo identification methods and relaxation criteria and we explain the choice adopted in this work. Later we describe how we calculate the parameters of the ellipsoidal collapse model starting from the eigenvalues of the deformation tensor smoothed on different scales. We then present a method to describe the distribution of halo formation time separating the contribution of relaxed and non relaxed halos.

- **Chapter 5**: We explain our peaks identification algorithm and present a statistic of peaks identified in our simulations. We study the correlation functions between protohalo centres of mass, and peaks of different quantities. At a later stage we present an alternative to the peaks model for halo formation. In fact, spherical and ellipsoidal collapse model are missing a dipole term which is present in perturbation theory. We study the points where initial dipole vanishes and we correlate them to the protohalo mass centres.

- **Chapter 6**: In the first part we present a statistic of Lagrangian parameters, and compare it with results obtained by other authors. Afterwards we investigate the correlation between the Lagrangian parameters δ_L and q_L and the halo formation redshift z_{50} as a function of rescaled mass ν and identification redshift z_{id}.

- **Chapter 7**: Whereas in the previous chapter we described Lagrangian parameters, here we present profiles. Firstly we build Lagrangian profiles around the protohalo mass centres, and show how they correlate with Lagrangian shear and halo formation times. Later we study the relation between Lagrangian and (Eulerian) profile, and show how Lagrangian shear and formation times affect the final halo profiles. Finally we study the evolution of halo particle profiles with time.

Sommario

Nel quadro del modello cosmologico standard, la formazione delle strutture è descritta attraverso il collasso di perturbazioni di densità con una distribuzione iniziale generalmente assunta come gaussiana. Queste fluttuazioni erano inizialmente piccole e sono cresciute successivamente per effetto della gravità. In un universo ΛCDM la formazione delle strutture cosmiche è guidata dal collasso della materia oscura che porta alla formazione di aloni virializzati. La materia barionica cade dentro alle buche di potenziale create da questi aloni, si raffredda e conduce alla formazione di stelle e galassie, trasformando la sua energia cinetica in energia termica (White and Rees 1978; Blumenthal et al. 1984). Successivamente, le strutture crescono in modo gerarchico, dalle più piccole alle più grandi. Quindi, aloni contenenti galassie massicce si formano tramite l'accrescimento di aloni più piccoli da parte dell'alone principale. Il collasso e la successiva formazione di aloni di materia oscura è dovuto unicamente alla gravità; per questo motivo la sua descrizione è semplice in linea di principio e coinvolge un alto numero di particelle. Un modo efficace di analizzare la formazione di questi aloni passa per l'utilizzo di simulazioni a N corpi. Con questo approccio non si ricerca una soluzione analitica, bensì viene fatto evolvere un gran numero di particelle soggette alla sola interazione gravitazionale. È così possibile simulare una regione di universo e analizzare il moto delle particelle e la formazione di strutture collassate.

La struttura di questo lavoro è la seguente:

- **Capitolo 1**: Presentiamo il modello cosmologico standard per la formazione delle strutture cosmiche e descriviamo brevemente le proprietÃ statistiche dei campi di fluttuazione lineari e la crescita di perturbazioni lineari alla luce della teoria di Jeans.
- **Capitolo 2**: Descriviamo i principali approcci analitici allo studio delle perturbazioni in regime non lineare. Prima di tutto presentiamo l'approssimazione di Zel'dovich applicabile al caso di un regime quasi lineare. e descriviamo i due principali modelli dinamici per lo studio del collasso delle strutture: il modello di collasso

sferico e il modello di collasso ellissoidale. Più avanti descriviamo due approcci analitici per la determinazione la statistica degli aloni a partire dal campo di fluttuazioni iniziale: il modello degli excursion sets e il formalismo dei picchi. Descriviamo anche brevemente alcuni tentativi di fondere assieme questi due approcci.

- **Capitolo 3**: Descriviamo il metodo principale utilizzato per lo studio di strutture fortemente non lineari: le simulazioni a N corpi. Descriviamo inoltre le caratteristiche principali delle simulazioni numeriche utilizzate in questo lavoro. Infine, descriviamo le proprietÃ principali degli aloni di materia oscura, concentrandoci sul contributo dato a questo campo dalle simulazioni numeriche.

- **Capitolo 4**: Inizialmente descriviamo i diversi metodi per l'identificazione di aloni e i criteri di rilassamento usati in questo lavoro. Successivamente descriviamo come sono stati calcolati i parametri del collasso ellissoidale a partire dagli autovalori del tensore di deformazione smussati su diverse scale. Analizziamo infine un'alternativa al formalismo dei picchi per l'identificazione della formazione degli aloni. Successivamente cerchiamo un modo per descrivere la distribuzione dei tempi di formazione dei soli aloni rilassati.

- **Capitolo 5**: Descriviamo il metodo usato per l'identificazione dei picchi nei campi iniziali e presentiamo una descrizione statistica dei picchi suddetti. Successivamente analizziamo la funzione di correlazione tra i centri di massa dei protoaloni e i picchi e le valli nella distribuzione dei parametri di interesse. Infatti, il collasso sferico e quello ellissoidale mancano di untermine di dipole presente invece nella teoria delle perturbazioni. Studiamo dunque i punti dove il dipolo iniziale svanisce e li relazioniamo ai centri di massa dei protoaloni.

- **Capitolo 6**: Nella prima parte presentiamo una descrizione statistica dei parametri Lagrangiani, svolgendo un opportuno confronto coi risultati di altri autori. Successivamente indaghiamo la correlazione tra i parametri lagrangiani δ_L e q_L e i redshift di formazione degli aloni z_{50} in funzione della massa universale ν e del redshift di identificazione.

- **Capitolo 7**:Mentre nel precedente capitolo abbiamo descritto parametri smussati su una sola scala, la scala lagrangiana degli aloni, passiamo ora all'analisi dei profili. Primariamente costruiamo i profili lagrangiani attorno al centro di massa dei protoaloni e mostriamo come correlano con il parametro di shear e il redshift di formazione. Successivamente studiamo la relazione tra i profili lagrangiani e i profili euleriani e mostriamo che l'effetto dello shear lagrangiano e del tempo di formazione Ã¨ ancora presente nei profili finali. Infine studiamo l'evoluzione dei profili delle particelle dell'alone a diversi tempi cosmici.

CHAPTER 1

The formation of cosmic structures

1.1 Cosmological background

Cosmology studies the origin and the evolution of the Universe on sufficiently large scales. Modern cosmology is founded on two fundamental assumptions: the homogeneity and isotropy of the matter in the universe on large scales and Einstein's theory of general relativity.

The former assumption is the so-called cosmological principle: although the universe is very irregular on the scales of galaxies and clusters of galaxies, its properties seems to be the same in every region and every direction when observed on large enough scales (say larger than ~ 100 Mpc). The detection of the Cosmic Microwave Background (CMB) gave a very strong confirmation of the cosmological principle: the temperature inhomogeneities present in the universe nearly 380.000 years after the Big Bang were very small, with an amplitude of the order of $\Delta T/T \sim 2 \times 10^{-5}$. Fig. 1.1 shows a comparison between the all-sky measurements of the CMB done by COBE, WMAP, and the most recent Planck experiment (Planck et al. (2011), Ade et al. (2014)).

In general relativity (the second assumption) gravitation is not a force but a property of space-time whereas the geometrical structure of space-time is given by the mass distribution in the universe through the *Einstein's field equations*:

$$\boxed{R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - g_{\mu\nu}\Lambda = \frac{8\pi G}{c^4} T_{\mu\nu}} \tag{1.1}$$

where $R_{\mu\nu}$ is the Ricci tensor, which describes the local curvature of space-time, R is the curvature scalar, $g_{\mu\nu}$ is the metric, $T_{\mu\nu}$ is the energy-momentum tensor and Λ is the cosmological constant. Solving the *Einstein's equations* together with the assumption of

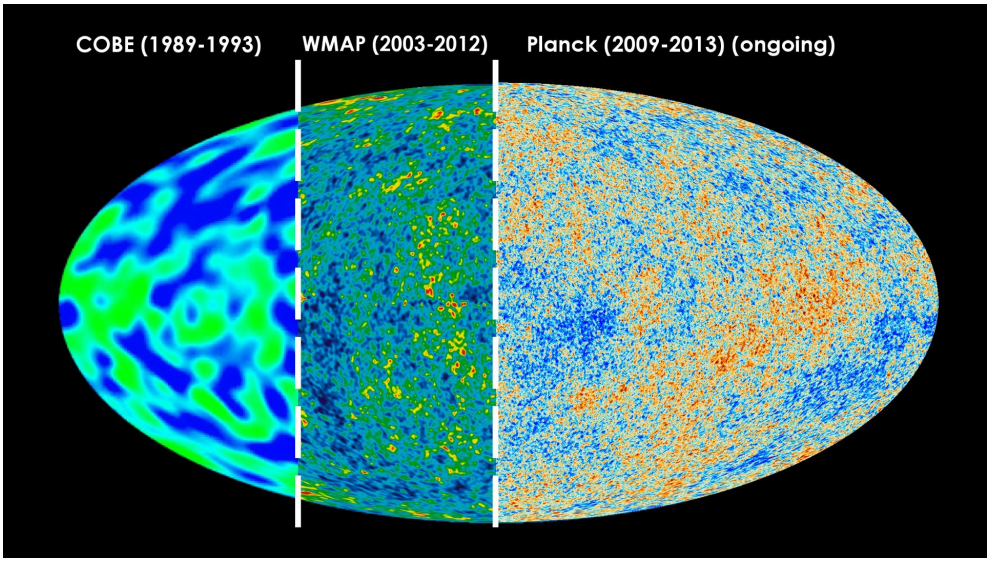

Figure 1.1: Comparison of the all-sky map of CMB anisotropies as detected by COBE, WMAP and Planck mission (here updated to 2013). Planck mission has created the sharpest all-sky map ever made of the CMB and the most accurate measurements of several key cosmological parameters.

the cosmological principle, leads to the so-called *Friedmann's equations*. The first one is derived from the 00 component of the field equations:

$$\boxed{\frac{\ddot{a}}{a} = -\frac{4\pi}{3}G\left(\rho + 3\frac{p}{c^2}\right) + \frac{\Lambda c^2}{3}} \tag{1.2}$$

and the second one is derived by putting together the first one and the trace of *Einstein's field equations*:

$$\boxed{\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{Kc^2}{a^2} + \Lambda c^2} \tag{1.3}$$

where $a(t)$ is the scale factor, which relates coordinates to distances, and K is the curvature of the Universe, which specifies the geometry of the surface. These parameters completely specify the *Robertson-Walker metric* that describes the geometric properties of a homogeneous and isotropic universe.

From the second of the *Friedmann's equations*, we can derive some important cosmological parameters, like the *critical density*:

$$\rho_c(t) = \frac{3}{8\pi G}\left(\frac{\dot{a}}{a}\right)^2 = \frac{3H^2(t)}{8\pi G} \tag{1.4}$$

and the *density parameter*, that is the ratio between the current total density of the Universe and the *critical density*:

$$\Omega(t) = \frac{\rho}{\rho_c(t)}. \tag{1.5}$$

Ω depends on the shape of the universe by way of the curvature K which can assume three different values:

- $K = 1$: Universe is closed and $\Omega > 1$,
- $K = 0$: Universe is flat and $\Omega = 1$,
- $K = -1$: Universe is open and $\Omega < 1$.

Current observational constraints suggest that $\Omega \simeq 1$ (de Bernardis et al. (2000); Planck et al. (2011); Ade et al. (2014)). Ω can be written as the sum of different components: non-relativistic matter (baryons and dark matter), Ω_m; relativistic matter (such as photons), Ω_r; vacuum energy, Ω_Λ. Standard values estimate the non-relativistic matter contribution at the present time as $\Omega_m \simeq 0.3$ and hence matter accounts for about a quarter of the total density energy of the Universe.

1.2 A model for the structure formation

One of the most important open problems in cosmology concerns how galaxies have formed. This problem can be addressed considering a theory able to describe coherently the evolution of the different constituents of the Universe through different cosmic ages.

The actual prevalent model for cosmic structures formation was in place in the middle of the 1990's (Krauss and Turner (1995),Ostriker and Steinhardt (1995)) and is known as the *Standard Cosmological Model*. Together with the two fundamental assumptions that we discussed before, this model is based on the idea that galaxies trace an underlying structure distribution, composed by a mysterious component (dark matter). Following this scenario, dark matter structures (halos) originated from initially small, Gaussian fluctuations in an otherwise homogeneous and isotropic universe. The idea that galaxies and clusters of galaxies are embedded in dark matter halos became popular in the 1970's after the works of Ostriker et al. (2012) and Einasto et al. (1974). White and Rees (1978) proposed a scenario where galaxies have formed within a hierarchically merging population of dark matter halos. Their theory was based on an analytic model proposed by Press and Schechter (1974) that we will present in Sec. 2.3. One additional issue concerns the nature of dark matter. In the White and Rees (1978) scenario dark matter was identified with an early population of low-mass stars. When Lubimov et al. (1980) measured the mass of the electron neutrino ($\simeq 30eV$), the idea came up that dark matter may be composed of non-baryonic particles; in fact, this mass was high enough to provide the critical density needed to close the universe. Other more massive particles were supposed to provide a reliable candidate for dark matter. Today dark matter particles are usually

classified into three families, depending on their typical velocities at some early time: hot, warm and cold dark matter particles. Some hybrid scenarios have also been proposed. The choice of the most probable dark matter constituent lays mainly on considerations about dissipation phenomena and the hierarchical way in which halos are supposed to have formed. Dissipation in dark matter is called free streaming: after decoupling from radiation, dark matter particles move influenced only by the mean gravitational field along geodesics in space-time and destroy every perturbation below the scale traveled from the beginning of their motion. This *free-streaming scale* can be calculated with:

$$\lambda_{FS}(t) = a(t) \cdot \int_0^t \frac{v(t')}{a(t')}\, dt' \tag{1.6}$$

where $a(t)$ is the scale factor and $v(t)$ is the velocity of dark matter particles. The mass diffused from a perturbation with a certain density is called *free-streaming mass*:

$$M_FS \propto \lambda_{FS}^3 \rho_{DM} \tag{1.7}$$

where ρ_{DM} is the density of dark matter. We notice that the free-streaming mass grows with the velocity of the dark matter particles. In a hot dark matter scenario dark matter is relativistic after decoupling and therefore the related free-streaming mass is very high (about $10^{15} M_\odot$) and a fragmentation process incurs: clusters form first and they fragment to form the smallest structures (top-down scenario). In the warm dark matter and cold dark matter scenario the free streaming mass is much smaller: small structures form first and then grow by merging and accretion, forming larger structures. Dark matter objects much smaller than galaxies (about $10^{-6} M$) can form only in a cold dark Matter scenario.

ΛCDM model considers a Universe where the main material constituent is not collisional and not baryonic cold dark matter (CDM), where the meaning of the different terms is the following:

1. **not collisional**: the cross-section for elastic collisions is about null. The motion of particles is only due to the mean gravitational field: mass elements move along geodesics in space-time.

2. **not baryonic**: dark matter is constituted by exotic particles predicted by supersymmetric theories. The evidences of not baryonic dark matter derive mainly from the comparison between nucleosynthesis and the dynamics of the Universe.

3. **cold**: the matter is not relativistic over all the ages of interest.

4. **dark**: there is no scattering due to the low scattering cross section of the particles and consequently there is not electromagnetic emission.

The favorite WIMPs (*Weakly Interactive Massive Particles*) candidate is Neutralino, which has a mass of $100 GeV$, large enough to permit a hierarchical scenario. This particle could be observed in principle from its couple annihilation with emission of $\gamma - rays$.

Visible matter is instead totally constituted by baryons. An upper and a lower limit on the baryonic density parameter is obtained from the study of the primordial nucleosynthesis, a process that went on about three minutes and produced elements during an inflationary process at the beginning of the radiative era. After recombination baryons are decoupled from the radiation and they fall into (linear) potential wells of dark matter. Gas heats up until virial temperature and then a radiation cooling occurs, leading to condensation and the formation of molecular clouds. Stars now can form and evolve, producing metals. After the explosion of the first Supernovae, gas was redistributed in the interstellar medium and the energy emitted was able to heat it up, creating new molecular clouds. The cycle is repeated until there is gas in the interstellar medium. Numerical simulations show that this process probably began when the age of the universe was less than a million years (Miralda-Escudé (2003), Gao et al. (2007)). The first proto-stellar objects are supposed to have formed with masses of $M \simeq 100 - 1000 M_{\odot}$.

1.3 Linear theory

The linear growth of perturbations in the Universe can be univocally analyzed by specifying only one scalar field. When perturbations are adiabatic, this scalar field is the gravitational potential $\phi(\mathbf{x}$ or, following the *Poisson equation*, the density fluctuation field $\delta(\mathbf{x})$,

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \rho_b}{\rho_b} \tag{1.8}$$

where ρ_b is the unperturbed density of the Universe. Being $\delta(\mathbf{x})$ a stochastic field, the only possible description of the Universe starting from its initial conditions is statistical. Therefore, the study of the large scale structure of the Universe is reduced to the study of the statistical properties of $\delta(\mathbf{x})$. In this section we briefly summarize some important properties of this fluctuation field, and then we will present the main results of the perturbation theory in the linear regime, when density fluctuations were much smaller than unity. The study of the perturbations growth in a non-linear regime will be addressed in Chapter 2.

1.3.1 Statistical properties of fluctuation fields

In the standard model the density fluctuation field is a product of the inflation. This field is generally supposed to be Gaussian with a null mean; so, it can be completely defined by knowing its variance. In turn, variance is related to the power spectrum.

The variance of a density fluctuation field ($\sigma^2 \equiv \langle \delta^2(\mathbf{x}) \rangle$) is the expectation of the squared deviations of the punctual density fluctuations in the field. However, cosmological structures are not points but have a (finite) size and so we need to filter the density fluctuation field with an appropriate window function to characterize these objects. The root-mean square of density fluctuations filtered with a sphere that contains a mass M is

called *mass variance*. We can write the mass variance integrating the power spectrum of the initial fluctuation field, weighting the integral with an appropriate window function:

$$\sigma_M^2 = \frac{\langle (M - \langle M \rangle)^2 \rangle}{\langle M \rangle^2} = \frac{\langle \delta M^2 \rangle}{\langle M \rangle^2} \tag{1.9}$$

where $\langle M \rangle$ is the average mass inside a volume V with radius R. Mass variance represents the average amplitude of inhomogenities in the matter distribution on scale M. We can rewrite this quantity in Fourier space obtaining:

$$\sigma_M^2 = \frac{1}{V} \sum_{\vec{k}} \delta_k^2 W^2(kR) \tag{1.10}$$

The aim of the window function $W(kR)$ is to select only fluctuations with $\lambda \simeq k^{-1} > R$ because perturbations with $\lambda < R$ cancel each other out by interference. The choice of an appropriate filter is very important, since wrong choices lead to important consequences. Some generally used window functions are:

1. **Gaussian**

$$W_G(\mathbf{x}; R) = \frac{1}{(2\pi)^{3/2} R^3} exp\Big(- \frac{|\mathbf{x}|^2}{2R^2} \Big), \tag{1.11}$$

$$\tilde{W}_G(kR) = exp[-\frac{(kR)^2}{2}]. \tag{1.12}$$

2. **Top-hat**

$$W_{TH}(\mathbf{x}; R) = \frac{3}{4\pi R^3} \begin{cases} 1, & if\ |\mathbf{x}| \leq R \\ 0, & if\ |\mathbf{x}| > R \end{cases} \tag{1.13}$$

$$\tilde{W}_{TH}(kR) = \frac{3(sin(kR) - kRcos(kR))}{(kR)^3}. \tag{1.14}$$

Another possible choice is the *sharp k-space* filter, with the form:

$$\tilde{W}_k(kR) = \begin{cases} 1, & if\ kR \leq 1 \\ 0, & if\ kR > 1 \end{cases} \tag{1.15}$$

and a corresponding form in real space:

$$W_k(\mathbf{x}; R) = \frac{1}{2\pi^2 R^3} y^{-3} (\sin y - y \cos y). \tag{1.16}$$

This window function is often employed in the analysis of Gaussian field (for example in the excursion set model of Bond et al. (1991), see Section 2.3). It has the advantage that the change in the field strength is independent of the original field $\delta(\mathbf{x}, R)$ when the radius

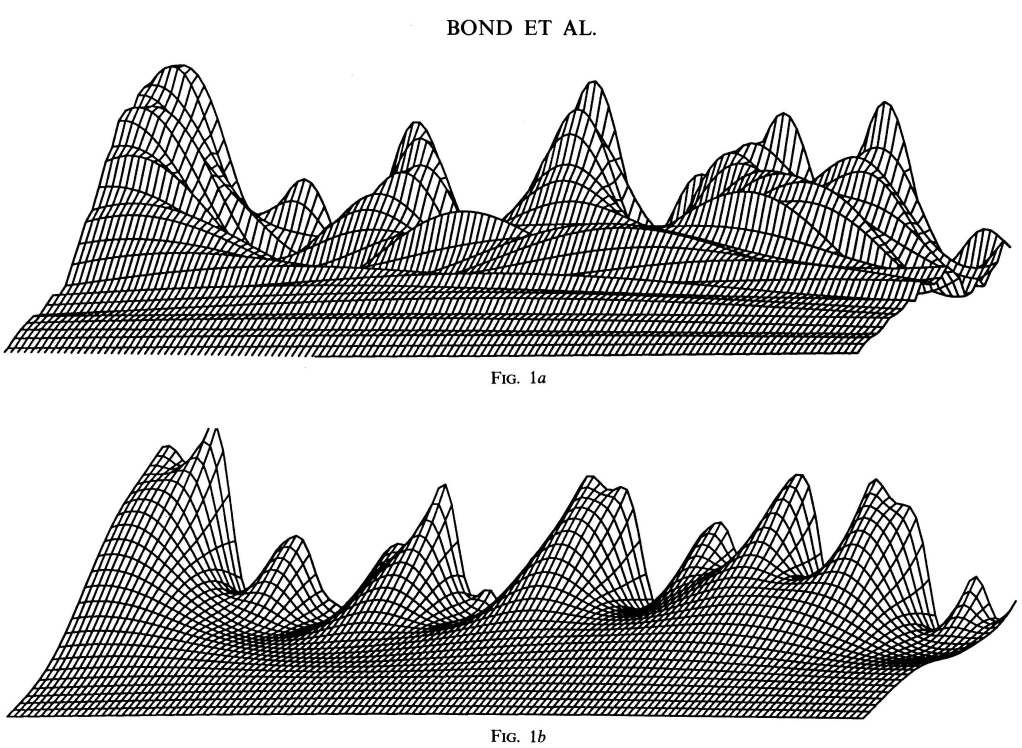

Figure 1.2: Topographical maps of a 1-dimensional stochastic Gaussian field smoothed with a *sharp k-space* (top) and a Gaussian (bottom) window function. The spectra are power law with $n = 0$. In the *x-axis* there is the position $\mathbf{x}$ and in the depth dimension there is the filtering scale in form of $log(R_f)$ i.e. small k (Bond et al. (1991)).

R is increased. However, it is not possible to assign a volume with this window, because its integration over all space diverges. This problem can be solved only by calibration against N-body computations (see also Bond et al. (1991)). A comparison between a sharp k space and a Gaussian filter can be seen in Fig. 1.2.

In this work we use a Top Hat in real space, since it is supposed to better approximate halo shapes in simulations. We will analyze a little bit deeper the consequences of a different filtering when discussing the excursion set formalism in Section 2.3.

1.3.2 Primordial Power Spectrum

Early models of structure formation generally assumed that the universe had two main components: baryons and radiation. Without a theory for the origin of perturbations, two general models were taken in account: adiabatic and isothermal initial conditions. The former considers an equal perturbation of all matter and radiation fields, whereas in the isothermal model is the ratio of components to be perturbed. Many efforts have been made

to study the linear evolution of both adiabatic and isothermal perturbations. An important result was obtained by Harrison (1970) and Zeldovich (1972); they suggested that the initial fluctuation spectrum (known as *Harrison-Zel'dovich initial power spectrum*) has the same dimensionless amplitude on every scale ($P(k) \propto k^n$ where $n \simeq 1$).

After the spread of the inflationary scenario (Guth (1981)) it was clear that quantum fluctuations of the inflaton (a scalar field) could produce a spectrum similar to the Harrison-Zel'dovich form together with a Gaussian shape and adiabatic perturbations. The low amplitude anisotropy in the CMB found by the Cobe experiment (Smoot et al. (1992)) permitted to find a good agreement with the Harrison-Zel'dovich spectrum and the hypothesis of Gaussian perturbations.

1.3.3 Perturbation growth

We now briefly summarize the growth and collapse of small Newtonian (i.e. smaller than the horizon size) perturbations in the linear regime ($\delta << 1$) through a first order perturbation theory.

Jeans (1902) studied the stability of a mass of a gravitating gas as found in a spherical nebula, showing that density perturbations grow only if they exceed a characteristic mass (or length) scale (the so-called *Jeans' mass*), making possible for gravity to win its competition with pressure. Writing the total energy equation at equilibrium ($E = E_g + E_K = 0$, where E_g and E_K are the gravitational and kinetic energy) for a spherical perturbation in a uniform fluid, we can obtain easily the *Jeans scale*:

$$\boxed{R_J = v_s \sqrt{\frac{1}{qG\rho}}} \tag{1.17}$$

where v_s is the sped of sound and q depends on the possible different definitions of this scale. The *Jeans' scale* separates two regimes:

- $R > R_J$: the fluctuation scale is larger than the *Jeans radius*: the contribution of gravity increases and fluctuations can collapse gravitationally.
- $R < R_J$: the fluctuation scale is smaller than the *Jeans radius*: diffusion wins and fluctuations are cancelled.

Jeans (1902) then studied how density fluctuations grow under the influence of their own gravity (the theory of *Jeans gravitational instability*). The starting point of this analysis are the equations for a self-gravitating fluid related to the density, velocity, pressure, entropy and gravitational field. The standard analysis, consistent with observational results, considers adiabatic and isentropic fields, resulting in an entropy constant both in time and space. After the introduction of small perturbations the evolution acts as a filter, amplifying some modes and damping others.

Static universe

This model was initially elaborated for a perfectly static, homogeneous and isotropic universe. The unperturbed solutions in this case are phisically inconsistent, because of the joint request of $\rho \neq 0$ and $\nabla^2\phi$. However, this simplified model is useful to understand what happens when introducing small perturbations. For computational ease the perturbations are decomposed into Fourier modes, transforming the equations from a differential to an algebraic form in the Fourier space. At the end we obtain stationary solutions with an exponential law

$$\delta\rho(r,t) = \delta\rho_k exp(\pm\|w\|t)exp(ikr), \tag{1.18}$$

where the growing solution describes the gravitational instability.

Expanding universe

Gamow and Teller (1939), Lifshitz (1946) and others extended the Jeans approach to an expanding background. A general dissertation of this case requires General relativity; otherwise, when considering non-relativistic perturbations and scales within the cosmological horizon, we can reduce the examination to a Newtonian analysis. Once that small perturbations are introduced in the fluid equations of an expanding Universe, we look for solutions in form of plane waves. The evolution equation of density fluctuations becomes:

$$\ddot{\delta}_k + 2\left(\frac{\dot{a}}{a}\right)\dot{\delta}_k + \delta_k\left[\frac{k^2 v_S^2}{a^2} - 4\pi G\rho_b\right] = 0 \tag{1.19}$$

The second term in the equation is the Hubble drag term, which tends to suppress perturbation growth because of the expansion of the Universe. The third term contains both gravitation and pressure contributions and its sign determines which kind of solution we have. When introducing in the equation the analytic form of $\rho_b(t)$ and $a(t)$ together with the assumption of $\omega = 1$ we obtain:

$$\ddot{\delta}_k + \frac{4}{3t}\dot{\delta}_k + \frac{2}{3t^2}\delta_k\left(\frac{k^2 v_s^2}{4\pi G\rho_b} - 1\right) = 0. \tag{1.20}$$

The general solution for the growing mode, including any possible value of ω, is

$$\boxed{\delta_+(z) = \frac{H_0}{a_0^2}(1+z)(1+\omega_0 z)^{1/2}\int_z^\infty \frac{dz(1+z)}{H_0^3(1+z)^3(1+\omega_0 z)^{3/2}}}, \tag{1.21}$$

which holds for $t > t_{eq}$ and $\lambda > \lambda_J$. This result is valid for both a baryon and a dark matter dominated Universe. By means of this equation we can descend some essential cases for different ages of Universe: before the equivalence, between equivalence and recombination, after the recombination. The age of interest for our work is the third one. We consider three different cases:

- If the scale of fluctuations is greater than the horizon scale, the fluctuations of baryons, radiation and dark matter follow the same trend:

$$\lambda > R_H : \delta_B \propto \delta_R \propto \delta_{DM} \propto a(\Omega = 1), \tag{1.22}$$

$$\delta_B \propto \delta_R \propto \delta_{DM} < a(\Omega < 1), \tag{1.23}$$

$$\delta_B \propto \delta_R \propto \delta_{DM} > a(\Omega > 1). \tag{1.24}$$

- If the scale of fluctuations is included between the Jeans scale and the horizon, the fluctuations of dark matter maintain the same trend owned over the horizon but the fluctuations of baryons and radiation swing:

$$R_J < \lambda < R_H : \delta_{DM} \propto a(\Omega = 1), \tag{1.25}$$

$$\delta_{DM} < a(\Omega < 1), \tag{1.26}$$

$$\delta_{DM} > a(\Omega > 1), \tag{1.27}$$

$$\delta_B \propto \delta_R. \tag{1.28}$$

- If the scale of fluctuations is smaller than Jeans scale, the fluctuations are canceled by diffusion.

The Meszaros effect

Dark matter has not always been the dominant component in the Universe. Before the so-called equivalence, radiation dominated. Eq. 1.21 can be rewritten like an ipergeometric equation, with a growing solution in the form:

$$\delta_{k;DM+} = 1 + \frac{3}{2}\frac{a}{a_{eq}} \tag{1.29}$$

where a_{eq} id the value of the scale factor at the equivalence. When matter perturbations come into the horizon before the equivalence, their growth slows (the *Meszaros effect*, Meszaros (1974)):

$$\frac{\delta_{eq}}{\delta_H} \leq \frac{5}{2} \tag{1.30}$$

The difference in the growth rate between radiation and matter dominated era imprints a characteristic scale corresponding to the horizon at the equivalence. Given the primordial power spectrum, the first fluctuations entering within the horizon are those on the smallest scales. For these fluctuations the Meszaros effect lasts longer and the power spectrum is strongly modified. If perturbations are large enough (i.e. larger than the scale of the

horizon at the equivalence), they do not undergo to any additional effect. The final shape of the processed Power Spectrum is:

$$\boxed{P_{fin}(k) = k^n T^2(k) \left[\frac{D(z_{fin}}{D(z_{in})}\right]^2} \tag{1.31}$$

where $T(k)$ is the transfer equation, that gives us information on how much of the primordial fluctuation remains unchanged. The final shape of the power spectrum can be seen in Fig. 1.3, where different candidates of dark matter are taken into account. The CDM density fluctuations have progressively larger amplitudes on smaller length scales. Hence structure formation is expected to proceed in a "bottom-up" manner, with smaller objects forming earlier.

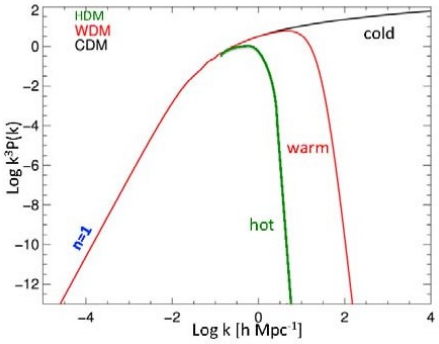

Figure 1.3: Power Spectrum in the normalized form $k^3P(k)$ for Hot (green), Warm (red) and Cold (black) dark matter Universes. We see how the *Meszaros effect* affects the power spectrum at large k so that only cold dark matter allows a hierarchical scenario.

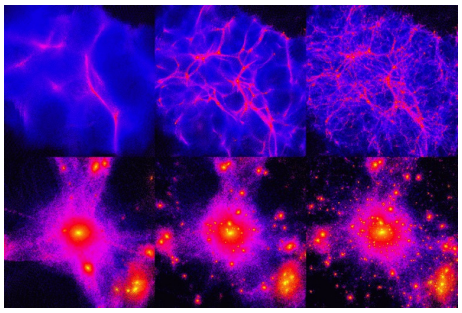

Figure 1.4: Simulations of structures at high redshift (top rows) and present time (low row) in the case of Hot (left), Warm (middle), and Cold (Right) dark matter Universes. We notice how the smallest halos form only in a Dark Matter Scenario. (courtesy ITC @ University of Zurich)

1.4 Correlation Functions

Statistical properties of the spatial distribution of a certain sample (like particles, stars, galaxies, clusters etc...) are described completely through correlation functions of every order. We present correlation functions of second order that we employ in our work.
The conditional probability that an object like a galaxy is into a volume δV_1 and another one is into a volume δV_2 separated by a distance $r_{12}^{\vec{}}$ in a great volume of the Universe is

$$\boxed{\delta^2 P = n_V^2[1 + \xi(r_{12})]\delta V_1 \delta V_2} \tag{1.32}$$

where n_V is the average number of galaxies into a unit of volume and $\xi(r)$ is the spatial two-points correlation function of galaxies. The probability depends only on the modulus of $\vec{r_{12}}$ and not on the spatial direction, because of homogeneity and isotropy of Universe. The correlation function ξ can be expressed as follow:

$$\xi(r) \equiv \langle \delta(\vec{x})\delta(\vec{x}+\vec{r})\rangle. \tag{1.33}$$

We can relate the conditional probability and the correlation function obtaining:

$$d^2P = n^2[1+\xi(r)]dV_1dV_2 \tag{1.34}$$

where $\xi(r)$ depends only on the modulus of r.
In a random distribution, $\xi = 0$ and therefore:

$$d^2P = n^2 dV_1 dV_2. \tag{1.35}$$

In Chapter 5 we will inspect the correlation function between Lagrangian peaks of certain quantities and protohalo centers using cross-correlations normalized by a random distribution.

CHAPTER 2

Non linear regime: analytic methods

Jeans theory gives us an idea of what happens in linear regime. When $\delta \to 1$ we enter in the nonlinear regime: at certain conditions there is a gravitational collapse and a dense structure formation occurs ($\delta \sim 10^2$), with an equilibrium between potential and kinetic energy (*virial equilibrium*). This leads to the formation of dark matter halos, the first nonlinear structures in the Universe. The halos carry on their mass growth by accreting material from their neighborhoods or by merging with other halos.

A model is needed to study nonlinear regime and inspect properties of dark matter halos. However, a purely analytic approach is very difficult because equations become very hard to solve. We firstly inspect the Zel'dovich approximation, initially used to describe the fluctuation field in a quasi-nonlinear regime; nowadays, it is largely employed to create the initial conditions of N-body simulations. Later, we will describe two dynamical approaches to the nonlinear regime: the spherical and the ellipsoidal collapse models. Finally, we will summarize the progress occurred in the development of two largely used analytic approaches: the Press-Schechter and the peaks formalism. These models take into account the predictions of both spherical and ellipsoidal collapse model at different stages and they are very useful to study (mainly in a statistical way) properties of the halo distribution, starting from the initial fluctuation field. We will describe the strong points of these two approaches and their weakness and we will briefly sketch the attempts provided to merge them together.

2.1 Zel'dovich approximation

The Zel'dovich approximation (Zel'Dovich (1970)) was firstly introduced to study a fluid in a slightly nonlinear regime but it remains unsuitable in presence of strong non linearity

(like in collapsed structures). Nowadays, it is used to create initial conditions in N-body simulations and so it is worth to discuss it.

Let us consider again the nonlinear equations (in comoving coordinates) for a non-collisional fluid in the case of an expanding Universe. To study the nonlinear regime, a (numerical) solution of these equations is required. However, Zel'Dovich (1970) employ some approximations in the case of a quasi-linear regime ($\delta \sim 1$) by mixing together the non linear equations with a linear potential. This method moves an initial particle distribution along a straight trajectory, following the initial potential. The approximation does not take into account the gravitational interaction between close particles and, for this reason, is not reliable in a strong nonlinear regime.
Our purpose is to derive the final density starting from an initial density. We consider a cubic volume and we express the density (which changes with mass elements position and time) as follows:

$$\rho(\mathbf{x}, t) = \bar{\rho}\Big[det\Big(\frac{\partial x_i}{\partial q_i}\Big)\Big]^{-1} \tag{2.1}$$

where $det(\frac{\partial x_i}{\partial q_i})$ is the determinant of the Jacobian matrix J that expresses the change of coordinates. This matrix is symmetric and so the determinant is reduced to:

$$detJ = (1 + D_+\lambda_1)(1 + D_+\lambda_2)(1 + D_+\lambda_3) \tag{2.2}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3$ are the eigenvalues. Inserting Eq. 2.2 in Eq. 2.1 we obtain:

$$\boxed{\rho(\vec{x}) = \frac{\bar{\rho}}{(1 + D_+\lambda_1)(1 + D_+\lambda_2)(1 + D_+\lambda_3)}}. \tag{2.3}$$

If we consider $x \ll 1$ we can use the approximation $(1 + x)^n \simeq 1$ and we can rewrite density at the first order in this form:

$$\rho(\vec{x}, t) \sim \bar{\rho}[1 - D_+(\lambda_1 + \lambda_2 + \lambda_3)] \sim \bar{\rho}[1 + D_+\delta_+], \tag{2.4}$$

which is valid for little times.
The solution at later times depends on the sign of λ_1, λ_2 and λ_3:

1. $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$: $1 + D_+\delta_+$ is positive: volume increases and ρ decreases forming a void.

2. $\lambda_1 \geq \lambda_2 \geq 0\, \lambda_3 \leq 0$: the element of volume enlarges along two dimensions and shrinks along the third one (orbit crossing). If density in the third dimension tends to infinity we have a *pancake*.

3. $\lambda_1 \geq 0\, \lambda_1 \leq \lambda_2 \leq 0$: the element of volume enlarges along one dimension and shrinks along the other two. We obtain a filament.

4. $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq 0$: we have orbit crossing in each direction. This is the limit of the Zel'dovich approximation: trajectories do not betray of the particle interactions and penetrate each other without any gravitational stabilization.

In a Gaussian field the 8% of the points have concordant eigenvalues, which form voids or spherical structures; the 42% have an eigenvalue with a different sign forming sheets.

In the classical Zel'dovich approximation structures that form earlier (small halos) undergo orbit crossing and are ruled out. To study small structures at later times we need to filter the initial conditions (*Truncated Zel'dovich approximation*), eliminating nonlinear structures when it's needed.

2.2 Dynamical models

Most of the objects of interest in the present-day universe have very high density compared with the background; therefore, they can be properly analyzed only taking into account a model for strong nonlinear regime. In this section we summarize the efforts done to face this problem from a (simplified) analytic point of view and we describe briefly the physics involved in the gravitational collapse and the dynamics of collisionless systems.

2.2.1 Spherical collapse

The simplest attempt to describe the formation of strongly nonlinear ($\delta \ll 1$) bound structures is the so-called *spherical collapse* (SC) model. This model was firstly developed by Gunn and Gott III (1972) in the case of a flat Universe with no cosmological constant and it has been used to provide a qualitative picture of how isolated gravitationally bound structures evolve up to their final collapse. Moreover, the SC model permits to estimate when the first astrophysical objects have formed under the influence of the only gravity (see Peebles (1980)). In the SC model the formation of bound objects in the Universe is described by the evolution of a uniform overdense spherical region in an otherwise smooth background.

Let's consider a spherical region with density higher than the density of the background. This perturbation is treated as a uniform spherical Universe with $\Omega > 1$ embedded in a flat Universe with $\Omega = 1$. In General Relativity, the equations of motions of a particle on the surface of the sphere can be written using the first equation of Friedmann with two assumptions: the initial peculiar velocity is null and concentric shells remain concentric, not crossing each other. The general solution for both background and perturbation is calculated by integration of the Newtonian equation for a mass shell, obtaining:

$$\dot{R}^2 - \frac{8\pi G}{3}\rho R^2 = -kc^2. \tag{2.5}$$

To study the evolution of the density contrast:

$$\delta(t) = \frac{\rho(t)}{\rho_b(t)} - 1 \tag{2.6}$$

we need a solution for both the background and the perturbation. In the case of background the curvature is $K = 0$. Knowing that $\rho_b \propto R_b^{-3}$, we obtain:

$$\rho_b = \frac{1}{6\pi G t^2}. \tag{2.7}$$

The evolution of the perturbation density with $K = 1$ ($\Omega > 1$) can be calculated considering that the absence of shell crossing leads to mass conservation within each shell. We obtain:

$$\boxed{R(\eta) = R_\star(1 - cos\eta)} \tag{2.8}$$

$$\boxed{t(\eta) = \frac{R_\star}{c}(\eta - sin\eta)} \tag{2.9}$$

where $R_{star} = GM_0/c^2$, M_0 is the mass of the whole perturbation and η is a variable obtained by change of coordinates $dt = \frac{d\eta R(t)}{c}$. The equations for $R(\eta)$ and $t(\eta)$ give the parametric equation of a cycloid: perturbation expands until the maximum of the curve (*turn around*, TA: $\eta = \pi$; $R = 2R_\star$) when the kinetic energy is zero and all the energy is contained in the gravitational potential term. After that, the perturbation collapses and reaches the minimum corresponding to the collapse point ($\eta = 2\pi$). The evolution of the overdensity can be obtained substituting 2.8 and 2.9 into 2.6, obtaining:

$$\boxed{1 + \delta(\eta) = \frac{9}{2}\frac{(\eta - sin\eta)^2}{(1 - cos\eta)^3}}. \tag{2.10}$$

If $\eta \ll 1$ we can approximate the equation in Taylor series, obtaining the following solution for the growing mode:

$$\delta_{lin} = \frac{3}{20}\left(\frac{6ct}{R_\star}\right)^{2/3} = \frac{3}{5}\delta_{in}\left(\frac{t}{t_{in}}\right)^{2/3} \tag{2.11}$$

that corresponds to the known solution from the linear theory.

We can compare the solutions obtained with linear, nonlinear and Zel'dovich approach:

- **Turn Around**:

$$\delta_{lin} = 1.06 \tag{2.12}$$

$$\delta_{SC} = 4.55 \tag{2.13}$$

$$\delta_{ZA} = 2.70 \tag{2.14}$$

- **Collapse**:

$$\delta_{SC} \longrightarrow \infty \tag{2.15}$$

$$\delta_{ZA} = 10.90 \tag{2.16}$$

$$\delta_{lin} = 1.68647 \tag{2.17}$$

It follows that the addition of nonlinear terms makes the perturbation grow faster. In linear theory, collapse happens in an infinite time and in the Zel'dovich approximation the time of collapse is very long. In the model of Gunn and Gott III (1972) the collapse of a spherical shell leads to an infinite density. However, real perturbation obviously only reach a physical density value. The main process occurring. that prevents the complete collapse of the perturbations and permits the formation of a dark matter halo, is virialization. The orbit crossing found in the Zel'dovich approximation is not possible, because particles interact gravitationally and flatten out in dark matter halos.

The radius of a virialized halo for an Einstein-De Sitter Universe can be calculated considering the virial theorem ($2T + U = 0$) and imposing the conservation of energy ($T + U = cost$):

$$\boxed{R_{vir} = \frac{R_{TA}}{2}} \tag{2.18}$$

So, after turn-around, shells cross each other and the structure virializes. The exact value of the nonlinear overdensity at virialization is:

$$\boxed{1 + \delta(t_{vir}) = \frac{\rho(vir)}{\rho_b(vir)} = 178}. \tag{2.19}$$

Once reached this equilibrium, the virialized region no longer feels the universal expansion around it, and in the absence of subsequent accretion, its physical size will remain constant with time. However, the halo never remains isolated: its central region remains unaltered but the outer one continuously grows, accreting material from the outside. In fact, every region with overdensity δ can always be embedded in a region with overdensity δ' with $\delta' < \delta$ so that the larger region with positive curvature recollapses adding new material to the internal virialized halos.

Since virialization produces a constant density contrast relative to the background, all halos at a given redshift are predicted to have the same mean density within their virial radius, and this virial density decreases with time as the background density decreases.

2.2.2 Relaxation processes

The final equilibrium configuration follows some relaxation processes that operate during the collapse. The relaxed equilibrium state of a halo depends on the initial conditions of the process and on the physics of the collapse. We briefly show two models provided to study the relaxation process occurring in dark matter halos: *secondary infall* and violent relaxation.

Secondary infall model

The *secondary infall* model (SIM) was introduced by Gunn and Gott III (1972), Gott (1975) and Gunn (1977) to describe the collapse and the virialization of spherically symmetric halos. In this model, particles do not collapse all at the same time: those within

inner shells collapse earlier and later there is an extended infall of new shells towards the central region. Fillmore and Goldreich (1984) and Bertschinger (1985) showed that the model emerged in the Gunn and Gott III (1972) approach can be made self-similar. In this way it is possible to calculate an analytical solution for the final density profile. This solution ($\rho \propto^{-2}$) is shallower that that emerging in numerical experiments of simple gravitational collapse ($\rho \propto^{-4}$).

Violent relaxation

When a collisionless system collapses or is perturbed, changes of the gravitational potential $\phi(\mathbf{x}, t)$ generally occur. In a time-varying potential, energy is no longer an integral of motion and so particles do not conserve it:

$$\frac{d\epsilon}{dt} = \frac{\partial \phi}{\partial t} \tag{2.20}$$

The overall effect is the broadening of the energy range and the potential provides a relaxation mechanism, called *violent relaxation*. This process causes a mixing of particles also in binding energy. The time scale for violent relaxation averaged over all particles is:

$$t_{vr} = \langle \frac{1}{\epsilon^2} \Big(\frac{\phi}{t} \Big)^2 \rangle^{-1/2}. \tag{2.21}$$

Lynden-Bell (1967) showed that this is equal to the free-fall time.

Violent relaxation does not lead to mass segregation during the relaxation process, because the variation of energy with time is independent of particle mass. On the contrary this process is present in collisional systems, where the exchange in momentum leads the kinetic energy equipartition of the system. However, a potential varying in time does not guarantee this mechanism, but a mixing must occur at the same time. Although violent relaxation tends to erase the memory of the initial state of a system, numerical simulations showed that the final particle energies are correlated with their initial values.

2.2.3 Ellipsoidal collapse

A first evident limit of the spherical collapse model comes up observing that large structures in the universe are not spherical but can be more realistically approximated with ellipsoids. White and Silk (1979) introduced a more accurate model to describe the growth and collapse of homogeneous ellipsoidal perturbations in a uniform expanding background. In their model perturbations evolve through homogeneous ellipsoids with increasing eccentricity until their shortest axis goes to zero, forming a "pancake". Following the notation of White and Silk (1979) and Mo et al. (2010), we can write the potential perturbation due to the matter within the ellipsoid as:

$$\Phi_{int}(\mathbf{x}) = -\pi G \rho_e \sum_{i=1}^{3} \alpha_i x_i^2, \tag{2.22}$$

where ρ_e is the density of the ellipsoid and

$$\alpha_i = X_1 X_2 X_3 \int_0^\infty dy (X_i^2 + y)^{-1} \prod_{j=1}^{3} (X_j^2 + y)^{-1/2} \tag{2.23}$$

with $X_i(t)$ the comoving lengths of the principal axes at time t (Chandrasekhar (1970), Mo et al. (2010)). We can write the total potential by summing the contribution of internal matter with that of the smooth background density

$$\Phi_b = -\frac{2}{3}\pi G \rho_b \sum_i x_i^2 \tag{2.24}$$

obtaining the total potential within a homogeneous ellipsoidal overdensity in an unperturbed Universe:

$$\boxed{\Phi_{tot} = -\pi G \sum_i \Big[\alpha_i \rho_e + \Big(\frac{2}{3} - \alpha_i\Big)\rho_b\Big] x_i^2}. \tag{2.25}$$

To obtain the equations of motion they make an assumption: the material outside remains uniform during its evolution and the external density can be calculated by the cosmological equations. In fact, when the density within the ellipsoids is large enough, we can neglect the external potential and the evolution of perturbations is governed by its self-gravity: the ellipsoid remains ellipsoidal and homogeneous during its evolution. However, in the linear regime ($\delta \ll 1$) this approximation is valid only without considering velocity perturbations at the initial conditions, otherwise the external and internal potential are of the same order and an expression for Φ_{ext} is needed. In their analysis the only velocity field considered is the Hubble flow. So, they calculate an approximate solution of the equations of motion starting from a system of four equations (the evolution equation of the ellipsoidal axis, the evolution of the background, the conservation of the mass for both the perturbation and the background) that are integrated using a fourth-order Runge-Kutta scheme. After additional approximations, they obtain the equations describing the exact evolution of homogeneous spherical perturbations:

$$\frac{a_i(t)}{a_i(t_0)} = R_b(t) - \frac{3}{2}\alpha_i(t_0)[R_b(t) - R_e(t)] \tag{2.26}$$

In the linear regime, the analysis of White and Silk (1979) does not reduce to the Zeldóvich approximation. Bond and Myers (1996) argued that this occurs because the effect of the external tide are not included self-consistently; so, they proposed a homogeneous ellipsoidal model that reduces to the Zelídovich approximation at the lowest order. Initial conditions and external tides are chosen to recover the Zeldóvich approximation in the linear regime. This model does not assume the sphericity of the initial perturbations but most works lays on this assumption (Sheth et al. (2001); Sheth and Tormen (2002); Desjacques (2008); Lam and Sheth (2008)). They consider a patch that is spherical at the initial conditions with radius r_0. The triaxiality in the potential leads to

tidal forces which deform the sphere, leading to a final triaxial object with principal axes $X_i(t) = r_0[1-\lambda_i D(t)]$ in the linear regime, where λ_i are the eigenvalues of the deformation tensor. In the Zeldóvich approximation the principal axes of the ellipsoid are parallel to the axes of the deformation tensor.

The eigenvalues of the deformation tensor can be expressed with a combination of the parameters describing the initial shear field:

$$\lambda_1 = \frac{\delta_i}{3}(1 + 3e + p) \tag{2.27}$$

$$\lambda_2 = \frac{\delta_i}{3}(1 - 2p) \tag{2.28}$$

$$\lambda_3 = \frac{\delta_i}{3}(1 - 3e + p) \tag{2.29}$$

Therefore, the evolution of a spherical shape of radius r_0 is studied employing three parameters, combination of the eigenvalues of the deformation tensor of the Zeldóvich approximation, centered on the center of mass of proto-halos and smoothed on a scale enclosing the proto-halo mass. These parameters are derived as following:

- **Overdensity**

$$\boxed{\delta = \frac{1}{(1+\lambda_1)(1+\lambda_2)(1+\lambda_3)} - 1} \tag{2.30}$$

- **Ellipticity**

$$\boxed{e = \frac{\lambda_1 - \lambda_2}{2\delta}} \tag{2.31}$$

- **Prolateness**

$$\boxed{p = \frac{\lambda_1 + \lambda_3 - 2\lambda_2}{2\delta}} \tag{2.32}$$

where $e \geq 0$ and $|p| \leq e$; a sphere has e = p = 0, prolate configurations have $e \leq p \leq 0$, and $0 \geq p \geq e$ are oblate.

Therefore, in the EC model the three axes collapse at different times depending on the local shape of the ellipsoidal deformation tensor. In comoving coordinates protohalos shrink fastest along the axis corresponding to the direction of maximum compression, forming a pancake; later the axis corresponding to the second and the minimum compression collapse eventually resulting in a virialized object. Assuming an initial spherical shape of the mass tensor, collapse times are determined by the initial compression factors: larger compression factor means earlier collapse. Virialization happens when all three axes collapse; at this point, the first axis to collapse will also be the shortest axis of the virialized object, and the third will be longest i.e. the shortest axis at virialization will be aligned with the direction of initial maximum compression, and vice versa.

2.3 Analythic approaches

The spherical and the ellipsoidal collapse models give us the possibility to estimate linearly extrapolated overdensity (i.e. a "collapse barrier") necessary for a certain region to break from the cosmic expansion and to form a dark matter halo. When combined with the statistics of the initial density field, the collapse barrier can thereby be utilized to estimate the abundance of dark matter halos as a function of mass and redshift.

There are two main analytic approaches to determine halo statistics starting from the initial fluctuation field: the excursion set formalism (Press and Schechter (1974); Bond et al. (1991)) and the peaks theory (Bardeen et al. (1986), hereafter BBKS).

2.3.1 Excursion sets

Press-Schechter formalism

Press and Schechter (1974) combined the simple results of the SC model with the statistics of the initial fluctuation field and the subsequent evolution of the Universe in order to estimate the abundance of collapsed structures. In this way they provided a statistical method to characterize the halo population without following the nonlinear dynamics in detail. In the SC model, regions with a positive overdensity smaller than δ_c at a certain redshift z will collapse at the present day if their linearly evolved density contrast at $z = 0$ exceeds δ_c:

$$\delta(0) = \frac{\delta(z)}{D(z)} > \delta_c. \tag{2.33}$$

Press and Schechter (1974) transferred the time dependence from the density δ_s to the threshold $\delta_c(t)$ and did the following hypothesis: an infinitesimal mass element in a position $\mathbf{x}$ will be part of a halo with mass equal or larger than M at time t if the linear fluctuation $\delta_f(\mathbf{x}; R)$ centered in $\mathbf{x}$ and smoothed with a spherical filter of radius $R \propto M^{1/3}$ has a value equal or larger than the threshold for collapse at time t. So, the probability to have $\delta_s > \delta_c(t)$ is equal to the fraction of mass elements at time t that are contained in halos with mass greater than M. The mass fraction can be written as:

$$F(z, R) = \frac{1}{\sqrt{\pi}} \int_{\delta_c/D(z)}^{\infty} exp\left(\frac{-\delta^2}{2\sigma^2(R)}\right) = \frac{1}{2} erfc\Big(\frac{\nu}{2}\Big) \tag{2.34}$$

where $\nu = \delta(z)/\sigma(R)$. One of the most important results of the Press-Schechter model was the derivation of the halo mass function (also see Efstathiou et al. (1979) and Epstein (1983)). In fact, if a fraction $F(R)$ of the matter distribution collapses when smoothed on scale R, and a fraction $F' < F$ collapses when $R' > R$, the difference can be assumed to have collapsed in the mass range $[M, M + dM]$. The halo mass function is then derived by differentiating $F(z, R)$ with respect to the filtering scale R:

$$\frac{dn}{dM} dM = 2 \frac{\bar{\rho}}{M} \frac{dF}{dM}(z, R(M)) dM \tag{2.35}$$

and finally obtaining:

$$\boxed{\frac{dn}{dM}dM = \frac{\bar{\rho}}{M}\sqrt{\left(\frac{2}{\pi}\right)}exp\left(\frac{-\delta^2}{2\sigma^2(R)}\right)\frac{d\nu}{dM}dM}. \tag{2.36}$$

Excursion sets model

In the original Press and Schechter (1974) calculation, underdense regions contained within larger regions for which the threshold $\delta > \delta_c$ is satisfied were not taken in considerations (the so-called cloud-in-cloud problem; it also arises in the peaks formalism that we discuss later). Because of this, only half of the total mass is considered in the halo mass function. Press-Schechter tried to solve this problem introducing a factor of 2 without changing the shape of the mass function. Bond et al. (1991) derived the halo mass function in an alternative way, trying to give a more convincing solution to the cloud-in-cloud problem. Their approach is known as the *excursion sets model*. They filtered the initial overdensity field on different scales, taking into account the patches of low density embedded in large, high-density regions collapsing on larger scales and justifying the "factor of 2" arising in the PS formalism with a sharp k filter. They used random walks, assuming that each position $\mathbf{x}$ corresponds to a trajectory which reflects the value of the density contrast when smoothing the region around that position with a filter of radius R. When R is very large, the density contrast will be 0. As R decreases, the random walk begins to wander away from zero in a way depending on the chosen filter (see Fig. 2.1). We remark that only a top-hat in k space allows to have a Markovian random walk. The halo mass function is derived assuming the barrier of the SC model. So, at fixed z, the barrier height is independent of the mass M of the collapsed region (and so it does not depend on $\sigma(M)$).

The differential halo mass function is proportional to the rate at which trajectories first match the threshold and can be written in the general form as:

$$\boxed{n(M)dM = \frac{\bar{\rho_0}}{M}\frac{1}{\Omega}\frac{d\Omega_v(M)}{dM}dM} \tag{2.37}$$

where ρ_0 is the average cosmological density, Ω is the density of the Universe in units of the closure density and M is a function of R_f. Without explaining the way in which trajectories are counted, we simply notice that the final form of the differential halo mass function found with the excursion sets model is identical to the one of Press-Schechter, including the factor of 2, introduced by Press and Schechter (1974) without a valid demonstration.

The excursion set formalism has been extended in order to predict not only the evolution of the mass function of dark halos, but also other properties of the hierarchical clustering process (Lacey and Cole (1993); Kauffmann et al. (1993); Kitayama and Suto (1996); Sheth and Van De Weygaert (2004)). A simple explanation of the way to investigate the halo merger tree is given by Lacey and Cole (1993). When considering a random

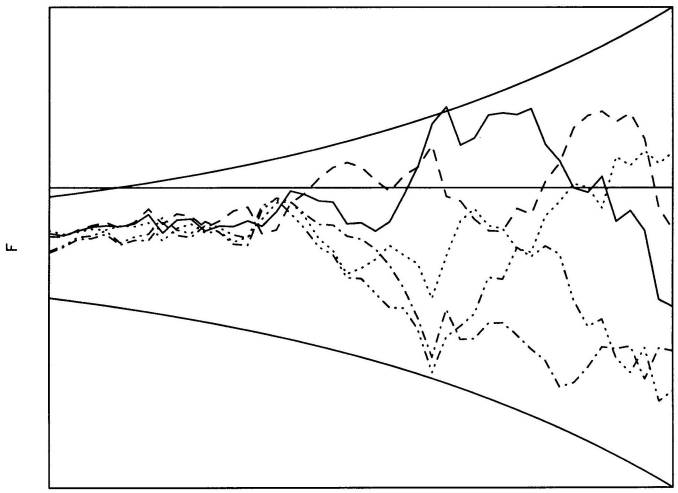

FIG. 3*a*

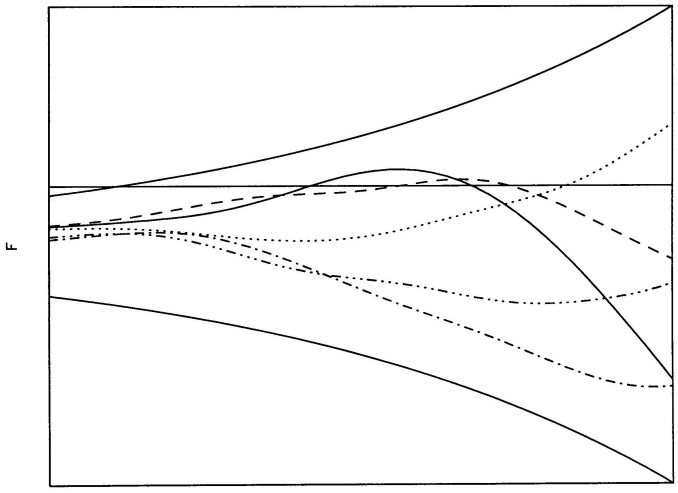

FIG. 3*b*

Figure 2.1: Trajectories of $F = \delta$ as s function of the smoothing radius R_f together with 2σ curves and the threshold for collapse. The trajectories are obtained smoothing the field with a sharp k-space (a) and a Gaussian (b). In the second case the random walk is not perfectly Markovian: the trajectories are smoothed and have long-range correlations.

walk, we can follow its merger history starting from small values of the mass and evolving the barrier with time. In the excursion set formalism is assumed that each fluid element is contained in a halo whose mass corresponds to the largest M at which the trajectory associated to the element crosses the threshold. When having a random walk going up and down, the decrease of the threshold with time cause the trajectory to match the threshold at different time and whenever it occurs the mass will jump horizontally to the value corresponding to the match with the threshold as we see in 2.2. This jump corresponds to a jump in the halo mass, whereas the small jumps found simply following δ as a function of S, are due to a mass accretion. In 2.3 we see an example of a merger tree (both figures are taken from Lacey and Cole (1993).

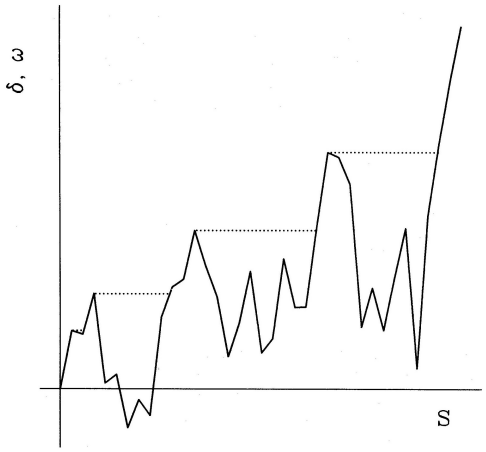

Figure 2.2: Merger history of a trajectory (solid line) of δ as a function of the smoothing radius, here described by $S(M) \equiv \sigma^2(M)$. To follow the merger history we need to start at high S and w and follow the track to the left and down. The trajectory for the halo mass during the merger history is represented by the dotted lines (a function $S(w)$). When δ increases with S, the dotted line coincides with the solid one (Lacey and Cole (1993)).

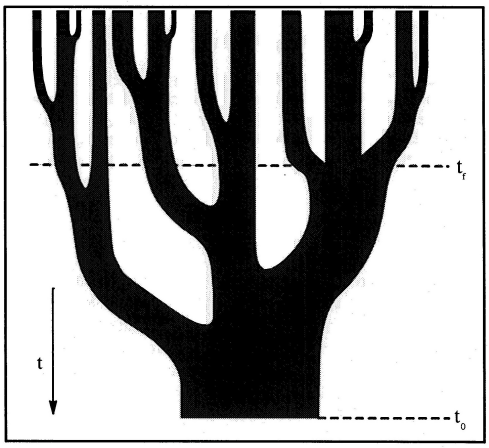

Figure 2.3: A schematic representation of a "merger tree", describing the merger history of halo, with time increasing from top to bottom. The branch sizes represent the mass of the progenitors. The distribution of the mass of parents at a certain time can be obtained considering a slice (dashed line). Here the formation time (t_f) is defined as the time at which half of the mass of the final has been assembled by a parent halo (Lacey and Cole (1993)).

The conditional distribution probability that a halo with a mass corresponding to S_2 was assembled by halos with mass corresponding to S_2 can be written in the form:

$$p(S_1, \delta_c(t1)|S_2, \delta_c(t2)) = \frac{\delta_c(t1) - \delta_c(t2)}{\sqrt{2\pi(S_1 - S_2)^{3/2}}} exp\left[- \frac{(\delta_c(t1) - \delta_c(t2))^2}{2(S_1 - S_2)} \right] \tag{2.38}$$

that can be written as a function of mass, obtaining the mass function of progenitors:

$$\boxed{p(M_1, t_1|M_2, t_2) = p(S_1, t_1|S_2, t_2)\left|\frac{dS_1}{dM_1}\right|} \tag{2.39}$$

The moving barrier

However, halo mass functions and spatial clustering of halos found in simulations were inconsistent with the prediction of Press and Schechter (1974) and in the following years accurate halo mass functions were obtained just fitting to the results of numerical simulations (Sheth and Tormen (1999); Jenkins et al. (2001); Warren et al. (2006), Tinker et al. (2008)). These discrepancies were due to some oversimplifications in the excursion set approach, like the spherical approximation. In fact, as we explained in the previous paragraph, not only perturbations in Gaussian density fields are more realistically triaxial, but also the shear field must be taken into account to study the formation of structures (Hoffman (1986); Peebles (1980); Dubinski (1992); Bertschinger and Jain (1993); Audit and Alimi (1996); Audit et al. (1997)).

Some attempts to consider ellipsoidal perturbations within the excursion set formalism have been done. Monaco (1995) employed the Zeldóvich approximation as it was a Lagrangian dynamics, computing the collapse epoch in correspondence of the first axis. They show that shear shortens the collapse time, in accordance with the lowered density threshold found in simulations (Efstathiou and Rees (1988); Carlberg and Couchman (1989); Klypin et al. (1994); Bond and Myers (1996)). In Bond and Myers (1996) both initial conditions and external tides recover the Zel'dovich approximation in the linear regime and the initial shape of an ellipsoid can be specified by its initial overdensity and by the parameters chracterizing the asymmetry of the tidal field: e and p. Lee and Shandarin (1998) extended the Press-Schechter formalism to a non spherical dynamical model, assuming that a collapse along all three direction can occur only in regions where all three eigenvalues are positive. Sheth et al. (2001) combined the Ellipsoidal Collapse model in the form of Bond and Myers (1996) with the assumption of a *moving barrier* for ellipsoidal collapse, providing a revised version of the halo mass function where the effective ellipsoidal collapse barrier is a function of mass through the peak height parameter, $\nu_c = \delta_c/\sigma(M)$ alone:

$$\frac{\delta_{EC}(e,p)}{\delta_{SC}} = 1 + \beta\left[5(e^2 \pm p^2)\frac{\delta^2_{EC}(e,p)}{\delta^2_{SC}}\right]^{\gamma} \tag{2.40}$$

where $\beta = 0.47$ and $\gamma = 0.615$. Recent calibrations have been calculated by Desjacques (2008) ($\beta = 0.412$ and $\gamma = 0.618$). Sheth et al. (2001) found a halo mass function that fits well the one found in numerical simulations (Sheth and Tormen (2002)). We can notice that the non spherical effects always lead to $\delta_{EC} > \delta_{SC}$: in correspondence of low mass halos the ellipticity is larger, and also the barrier. The problem of finding the halo mass function reduces to find the first crossing of the barrier in a higher dimensional space

defined by the parameters of the ellipsoidal collapse model. However, this method has a great computational complexity and can be simplified only considering some peculiar values of e and p. From the distribution function of the eigenvalues in a Gaussian random field (see also Doroshkevich (1970)) we find $p_{mp} = 0$ and $e_{mp} = (\delta/\sigma)/\sqrt{5}$. Inserting those values in 2.40 we obtain:

$$\boxed{\frac{\delta_{EC}(\sigma, z)}{\delta_{SC}(z)} = 1 + \beta\Big[\frac{\sigma^2}{\sigma_\star^2}\Big]^\gamma} \tag{2.41}$$

where $\sigma_\star = \delta_{sc}(z)$. This equation predicts that for massive objects ($\sigma/\sigma_\star \ll 1$) the ellipsoidal collapse barrier reduces to the spherical value, that is independent of mass. Smaller objects are more influenced by external tides and so they need a larger overdensity to hold themselves together.

When applying the moving barrier, the mass function becomes:

$$\boxed{\nu f(\nu) = 2A\Big(1 + \frac{1}{\nu^{2q}}\Big)\left[\frac{\nu^2}{2\pi} exp\Big(-\frac{\nu^2}{2}\Big)\right]} \tag{2.42}$$

where $q = 0.3$ and $A \sim 0.32222$, whereas the standard model had $q = 0$ and $A = 1/2$.

Correlated steps

Exact solutions for constant and linear barrier have been found in the case of uncorrelated steps of the random walk, together with analytic approximations for more general barriers. It is worth noticing that the steps of random walks in the ES model should not necessarily be independent; on the contrary, a correlation could be introduced, depending on the smoothing filter and the power spectrum.

Let's consider a random walk with height δ at a certain scale $s \equiv \langle\delta^2\rangle$ and a barrier B depending on that scale. When steps are uncorrelated the probability that the walk has not crossed the barrier at $S < s$ and the probability that $\delta > B(s)$ at s can be calculated separately. This is not true when considering correlated steps, because the step at S is correlated in some ways with the previous ones. Paranjape and Sheth (2012) considered completely correlated steps, for whom $\delta = \delta_0\sqrt{s/S_0}$ (where δ_0 is the value of δ on scale S_0) and calculated an expression that describes well the first crossing distribution also when the steps are not completely correlated (at least for small values of s). Musso and Sheth (2012) provided a formula usable for different kind of step correlations (i.e. for different barrier shapes and power spectrum shapes) that extends the expression for completely correlated steps. A simple case of this formula for a constant barrier is expected to fit well the halo abundances:

$$\nu f(\nu) = \frac{\nu e^{-\nu^2/2}}{\sqrt{2\pi}}\left[\frac{1 + erf(\Gamma\nu/\sqrt{2})}{2} + \frac{e^{-\Gamma^2\nu^2/2}}{\sqrt{2\pi}\Gamma\nu}\right] \tag{2.43}$$

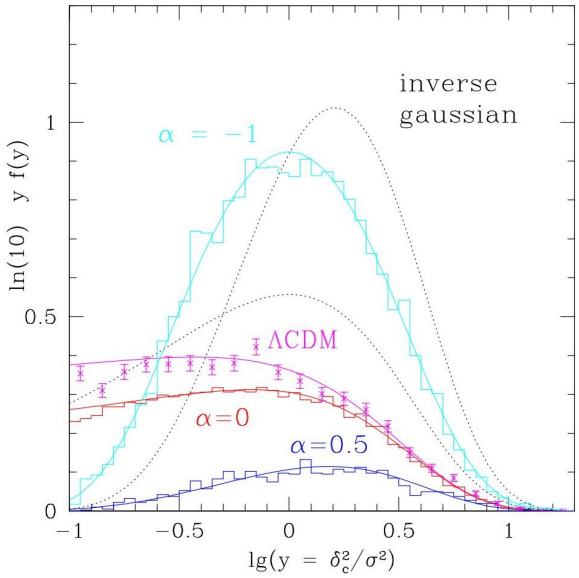

Figure 2.4: Distribution of the scale y on which random walks first cross the barrier $B(s) = \delta_c(1+\alpha s/\delta_c^2$. A comparison between Monte Carlo solutions for Gaussian smoothing of $P(k) \propto k^{-1/2}$ (for which $\Gamma^2 = 9/10$) is shown considering barriers with $\alpha = 0.5, 0$ and -1. In the last two cases also the solution for uncorrelated steps is provided (dot curves). The general solution for $f(s)$ found in Musso and Sheth (2012) (smooth curves) is in good agreement with the results of the different barrier shapes. Eq. 2.43 (smooth curves). When considering a top hat smoothing of a Λ CDM spectrum for a constant barrier ($\alpha = 0$, see points with error bars) we notice that first crossing distribution depends weakly on $P(k)$, differently of the case with uncorrelated steps. Eq. 2.43 (solid curve) shows a good agreement.

where $\nu = \delta_c/\sigma$, Γ is a function of γ that in turn was introduced by citetbardeen1986statistics to describe the spectral quantity $\sigma_1^2/\sigma_0\sigma_2$ in the case of a Gaussian smoothing filter. The expression for γ in presence of a top-hat was provided by Paranjape and Sheth (2012). The comparison between this formula and Monte Carlo simulation can be seen in Musso and Sheth (2012) for two different power spectra and filters.

2.3.2 Peaks theory

Excursion set model provides lots of significant results about statistical properties of non-linear structures. However, the formalism on which is based is not mathematically rigorous. Peaks theory would give stricter analytic foundations to the relation between non-linear structures and the initial density fields. However, it has some limitations that excursion sets have not: it is not able to model the formation of halos and the halo mass function. Attempts to put together peaks theory and excursion sets will be briefly summarized at the end.

The fundamental assumption of peaks theory lies on the idea that matter collapsing in bound structures of given mass can be related with peaks smoothed on the appropriate mass scale in the initial density field. In this scenario the condensation of matter occurs around sufficiently high local maxima of density. First applications of this peaks approach to cosmological models was performed by Doroshkevich (1970), Doroshkevich and Sandarin (1978a,b) and Peebles (1980). A mathematical development for hierarchical models as the Cold Dark Matter model has been realized by Peacock and Heavens (1985) and by Bardeen et al. (1986) (BBKS). BBKS studied extensively the properties of local maxima (peaks) in Gaussian random fields, providing a number of useful statistics, such as the number density of peaks of given height, their clustering properties and the density profile centered on peaks. These results can be related to the spatial clustering and to the number density of dark matter halos.

With the purpose of finding well defined peaks at the initial conditions, the linear density field must be smoothed with an appropriate filter. The spatial distribution of peaks located at $\mathbf{r}_{pk}$ is:

$$n_{pk}(r) = \sum_{pk} \delta^D(\mathbf{r} - \mathbf{r}_{pk}) \tag{2.44}$$

Given the smoothed density field $\delta(\mathbf{r})$, peaks are defined as usual, like points at which:

- the first derivative (gradient) of the smoothed field is zero: $\eta(r) \equiv \nabla\delta(r)$
- the second derivative tensor of the field ($\zeta_{ij} \equiv \nabla_i\nabla_j\delta(\mathbf{r})$) is negative definite at $\mathbf{r}_{pk}$.

Moreover we consider only maxima whose heights are in a certain range, for example: $\delta/\sigma \geq \nu$ where $\nu \equiv \delta_c/\sigma$ and σ is related to the power spectrum by $\sigma^2(R) \equiv \langle\delta^2(\mathbf{x}; R)\rangle$. The density of peaks can be written in differential form as:

$$N_{pk}(\nu)d\nu = \left\langle \delta^D(\delta/\sigma - \nu)|\Lambda_1, \Lambda_2, \Lambda_3|\mathcal{H}(\Lambda_3)\delta^(D)(\eta)\rangle d\nu \right\rangle \tag{2.45}$$

where Λ_i (with i=1,2,3) are the three eigenvalues of the field tensor (the capital letter Λ needs to distinguish them from the ones of the Zel'dovich approximation) and $\mathcal{H}$ is the Heaviside step function (equal to zero for negative values of the lowest eigenvalues Λ_3 and 1 elsewhere). We compute the average considering the distribution function in the form

of a multivariate Gaussian, obtaining:

$$\boxed{\mathcal{N}_{pk}(\nu)d\nu = \frac{1}{(2\pi)^2 R_*^3} e^{-\nu^2/2G(\gamma,\gamma\nu)} d\nu} \tag{2.46}$$

where the spectral quantity γ and the comoving length R_* are in relation to moments of power spectrum $P(k) = \langle|\delta(k)|^2\rangle$:

$$R_* \equiv \sqrt{3}\frac{\sigma_1(R)}{\sigma_2(R)}, \qquad \gamma \equiv \frac{\langle k^2\rangle}{\langle k^4\rangle^{1/2}} = \frac{\sigma_1^2}{\sigma_2\sigma_0} \tag{2.47}$$

and $G(\gamma, \gamma\nu)$ is a complicate function for which they found an accurate fit.

Fig. 2.5 and 2.6 show the differential and the cumulative number density

$$\boxed{n_{pk} \equiv \int_\nu^\infty \mathcal{N}_{pk}(\nu)d\nu} \tag{2.48}$$

that permits to count peaks above a certain threshold.

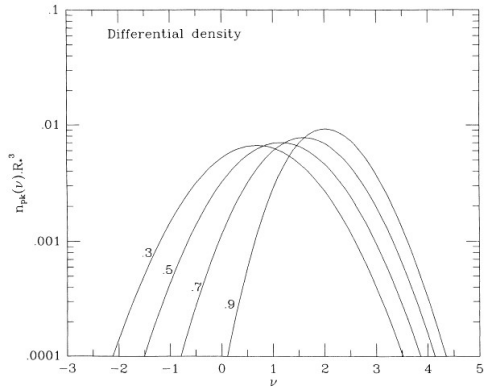

Figure 2.5: Differential number density $N_{pk}(\nu)$ of peaks between ν and $\nu + d\nu$ for various values of γ (Bardeen et al., 1986).

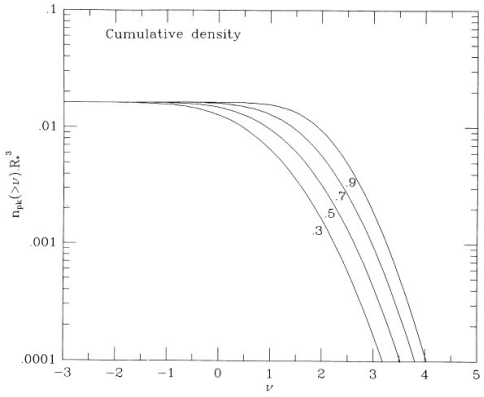

Figure 2.6: Cumulative number density $N_{pk}(\nu)$ of peaks with height above a certain threshold ν for various values of γ (Bardeen et al., 1986).

The differential and the cumulative peak densities can be written in the case of high peaks:

$$\boxed{\mathcal{N}_{pk}(\nu)d\nu = \frac{(\sigma_2^2/3\sigma_0^2)^{3/2}}{(2\pi)^2}(\nu^3 - 3\nu)e^{-\nu^2/2}d\nu} \tag{2.49}$$

$$\boxed{n_{pk}(\nu) = \frac{(\sigma_2^2/3\sigma_0^2)^{3/2}}{(2\pi)^2}(\nu^2 - 1)e^{-\nu^2/2}} \tag{2.50}$$

These equations agrees with previous results of other authors (see for example Eqs.19 and 20 in Doroshkevich (1970)).

However, a plausible physical way to select peaks which form halos is most likely less sharp respect to the form found in 2.50, where ν_t is the threshold height. Therefore, Bardeen et al. (1986) introduce a selection function $t(\nu/\nu_t, q)$ that takes into account also peaks below the threshold (although with a lower probability) and gives the probability that a peak of height ν forms one of the objects. The differential number density for $\nu_t = 3.5$ and different values of q's is shown in 3.1. Although the probability to have a peak below the threshold is always low, their number increase as q decreases and low peaks dominate for small values of q.

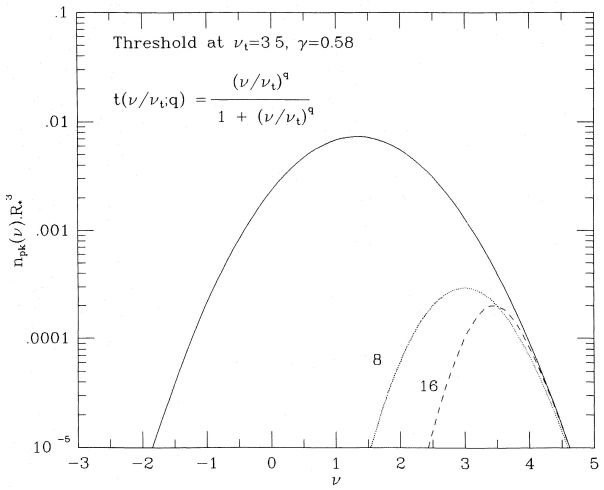

Figure 2.7: Product of the selection function and the differential peak number density $n_{pk}(\nu)$ for $\nu_t = 3.5$ and $q = 8$ and 16. Bardeen et al. (1986)

The difficult to associate a number density of halos to the number density of peaks is due to the cloud-in-cloud (or peak-in-peak) problem, previously found in the Press-Schechter formalism. In fact a mass element could be associated to peaks of different heights. So to relate peaks and halos, a statistics of peaks as R is varied is needed. However, this problem is pretty hard to treat mathematically. Manrique and Salvador-Sole (1995) calculate the halo mass function taking into account this 'peak-in-peak' problem. They consider a differential density of peaks for a fixed density contrast $\mathcal{N}_{pk}(R, \delta)dR$ in the form proposed by Bond (1989):

$$\mathcal{N}_{pk}(R, \delta)dR = N_{pk}(\nu, R)\frac{\partial \nu}{\partial R}dR \tag{2.51}$$

(where $\mathcal{N}_{pk}(\nu, R)$ is the same in Eq. 2.49 but R is treated as a parameter) and then

transform it to the mass function of objects at the time t:

$$\boxed{N(M,t)dM = N(R,\delta_c)\frac{dR}{dM}dM} \tag{2.52}$$

where $N(R,\delta_c)$ is expressed in the form of a Volterra type integral equation of the second kind that takes into account the cloud-in-cloud problem and is solvable numerically by iterations, $M(R)$ is the mass of collapsing clouds associated with peaks at scales between R and $R+dR$, and $\delta_c(t)$ is the typical overdensity for collapse.

Attempts to test how the peaks formalism works from an "object-to-object" point of view found initially a low confirmation. Ludlow and Porciani (2011) used two high-resolution cosmological N-body simulations to find a pretty good correspondence between halos at late times and peaks in the initial conditions. They identify halos at $z=0$ with a friends-of-friends algorithm (Davis et al. (1985), see next chapter to a better explanation of different algorithms for halos identification) and identify peaks as values of the smoothed field δ_s denser that all 26 neighbouring points. The evolution of peaks is followed by tagging the closest particle. The comparison between the number density of peaks in their simulations and the predictions of (Bardeen et al., 1986) is shown in Fig. 2.8.

They associate to each halo a single "main" halo peak, defined as the one that is still within the halo at $z=0$ and for which the filter mass is closest to the true FOF halo mass. They find that the majority of halos are associated with peaks of the same characteristic mass. The cumulative fraction of halos above a given mass that also contain peak particles is showed in Fig. 2.9. Figure shows that all halos with more than 100 particles contain at least one peak particle and the majority of them contain a peak particle on the halo mass scale with a fraction depending on the halo mass: higher for larger halos. They also found some (typically low mass) halos forming close to peaks smoothed on very different scales ('peakless' halos). These halos are on average tidally compressed along two axes of the initial tidal tensor and evolve from more oblate initial configurations (see right panel of Fig. 2.10).

Hahn and Paranjape (2013) found similar results using a power spectrum with a small-scale cut-off as in warm dark matter cosmologies. In their simulation all halos form close to initial density peaks and they also found a strong correlation between protohalo density and ellipticity.

2.3.3 Excursion set model of peaks

The excursion set approach considers all points in space, looking for a statistical description of mass structure in bound objects. The peak model can be combined with excursion set theory by imposing that peaks on a given smoothing scale are counted only if they satisfy a first crossing condition.

Bond and Myers (1996) used a numerical prescription (the 'peak-patch' approach) studying an excursion set formalism for peaks in order to solve the cloud-in-cloud problem. Sheth et al. (2001) claimed that an excursion set formalism that takes into account only

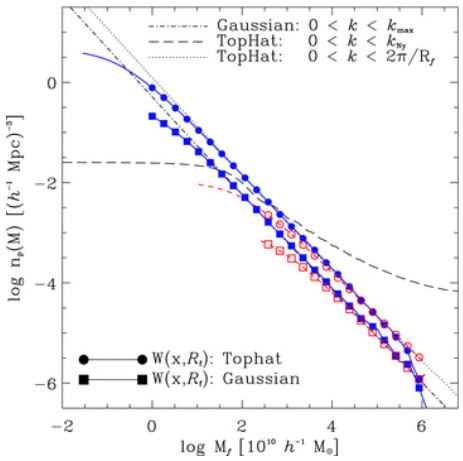

Figure 2.8: Comoving number density of peaks found in two simulations (red and blue curve) as a function of the filter mass. The range of smoothing scales go from $M_f = M_{part}$ (unsmoothed density field) to $M_f = 1.16 \times 10^{16} h^{-1} M_\odot$. The theoretical peak number density expected from the peaks theory is shown for three different filters (Ludlow and Porciani, 2011).

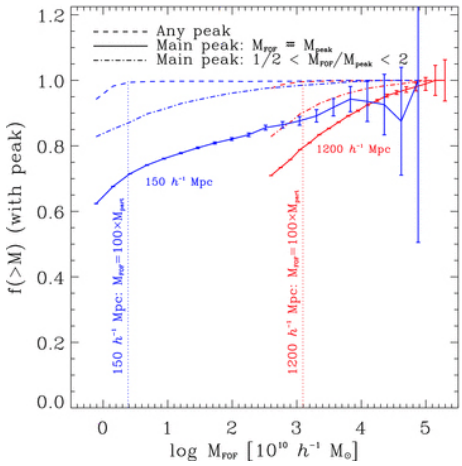

Figure 2.9: Cumulative fraction of halos founded in two simulations (red and blue curve) that contains peak particles of every smoothing scale (dashed) and in the range $1/2 < M_{FOF}/M_{peak} < 2$ (dot-dashed) Solid curves identifies the Poisson errors to the cumulative fraction of halos associated with peaks of the same characteristic mass (Ludlow and Porciani, 2011).

special positions (like peaks) is necessary. Moreover, recent numerical simulations (Ludlow and Porciani (2011)) showed that a large fraction of halos forms in correspondence with peaks. So, different attempts have been provided to unify the two approaches. Sheth et al. (2001) wrote a form for the number density of halos in a certain mass range, depending only on ν. However, this formula underpredicts massive clusters. As a consequence, one more parameter (q) has been taken into account in the fit $f(\sqrt{q}\nu)$, finding a universality of q. Paranjape and Sheth (2012) investigated the origin of $q \neq 1$, given the fact that $q < 1$ emerges only in a statistical description of the number density of halos but almost never in an object-by-object analysis. Therefore, they search for a way of integrating the peaks approach within the excursion set formalism creating an excursion set model for peaks.

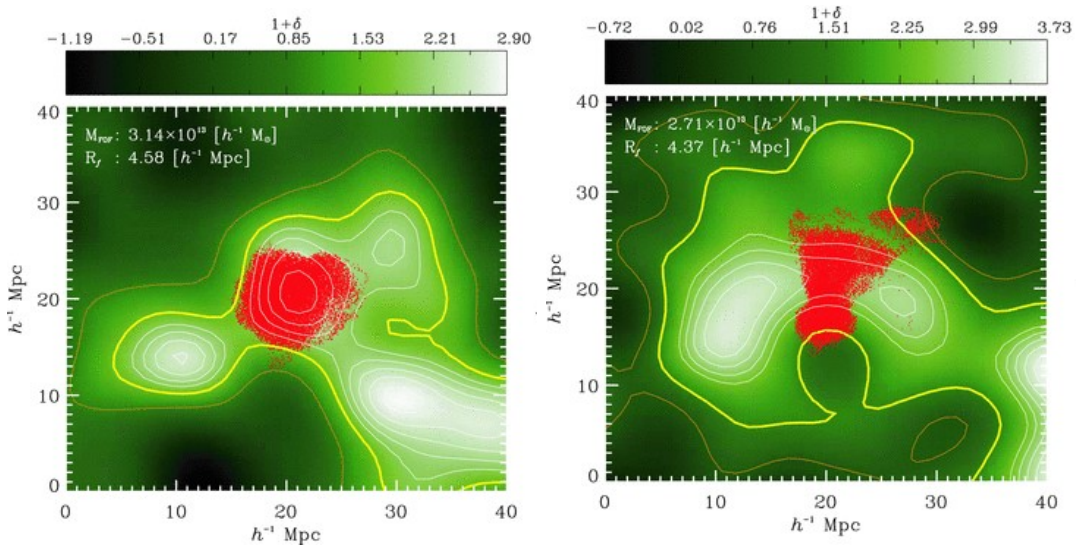

Figure 2.10: Overdensity field in proximity of a halo with $M_{pk} \simeq M_{FOF}$ (left) and $M_{pk} \leq M_{FOF}/4$ (right; peakless halos). Red dots are the proto-halo particles for a halo identified at $z = 0$. The orange curve corresponds to a density contrast $\delta_s = 1$ and the yellow one corresponds to the linear collapse threshold ($\delta_s = 1.686$) (Ludlow and Porciani, 2011).

CHAPTER 3

A numerical approach: N-body simulations

3.1 General overview

Formation and properties of dark matter halos can be addressed with both N-body simulations and analytic models. In Chapter 2 we briefly inspected the main analytic approaches; among them, the Press-Schechter formalism is the most developed to date. However, the study of the inner properties of halos is very complicated to approach analytically, because of the strong non linearity of the involved equations. N-body simulations are instead more suitable to study the clustering in the non-linear regime and provide numerical solutions for the equation of motion of many particles that interact gravitationally.

N-body simulations of the gravitational collapse of a collisionless system of particles pre-dates the CDM model. The first cosmological simulations of structure formation date back to Press and Schechter (1974) and in the 1980's simulations began to be widely used. In the 1990's there was a huge growth both in the size of cosmological simulations and in the sophistication of the physics and in the last years simulations with a huge number of particles and very different mass resolutions have been run.

The employment of N-body simulation needs to setup the initial conditions, to compute the force field for a certain configuration of particles and to move each particle following this force field. In order to run a cosmological N-body simulation with N_p particles populating a region with volume V, we need to take into account the following physical requirements (Bagla and Padmanabhan, 2004):

- **Periodic boundary conditions**: the real universe does not have a boundary. A method is hence needed to fill with matter the region outside the simulation box. Periodic boundary conditions are of fundamental importance because otherwise most of matter within the simulation volume would collapse towards the box center.

- **Average density**: the average density of the box should be equal to the average density of the universe.

- **Mass resolution**: the mass resolution should be high enough, in order to resolve the object of interests; otherwise, the employment of N-body simulations is totally unuseful. In this work we use N-body simulations with different mass resolutions in order to have a large halo statistic over a wide range of masses and to study inner properties of halos.

- **Collisionless**: Each particle in an N-body simulations has a mass equal to a huge amount of real particles (whatever they are). Therefore the simulation particles must interact in a purely collisionless manner.

One of the issues arising with N-body simulations is the huge computational cost required, expecially in the calculation of forces between particles. Even simulations evolving few hundreds of particles requires too much time when a direct computation of force is employed. Some schemes have been provided to avoid this problem and reduce the computational time:

- **Particle Mesh (PM)**: particles are converted in a grid/mesh of density values choosing between different interpolation schemes. The Poisson equation for gravitational potential is then solved in the Fourier space:

$$\phi(\vec{k},t) = 4\pi G a^2 \frac{\rho(\vec{k},t)}{k^2} \tag{3.1}$$

  where ρ and ϕ are the discrete Fourier transforms of respectively the mass density and the potential. The gravity field is then obtained by transforming the potential back to the spatial domain. Finally, the force is interpolated from the grid back to particles that can now be moved. There are different ways to interpolate a system of particles on a grid. The simplest one is the **ngp** (Nearest Grid Particle) method, where each particle is assigned to the closest grid point with no contribution of mass to any other one. This method produces large truncation errors (Efstathiou et al 1985, Hockney and Eastwood 1988) and so another method becomes to be largely used, the cloud-in-cell (**cic**) scheme, where particles are interpolated to the eight grid points defining the cubical mesh cell containing the particle. So, one particle can contribute mass to several cells. The force fluctuations introduced by the sharp edges can be reduced by using a higher-order interpolation scheme, the Triangular Shaped Cloud (**tsc**), which uses the nearest 27 grid points.

  The PM method requires O(N)+O($N_p \log N_p$) operations to evaluate the force acting on all particles and hence it has a good speed. However, the inverse square law is poorly approximated by the forces for pair separations less than several grid points.

- **Particle-Particle Particle Mesh (P^3M)**: This method improves the resolution of PM method and was first applied in cosmology by Efstathiou and Eastwood (1981). The force obtained in a PM algorithm is supplemented by a direct summation of contribution from particles (Hockney and Eastwood 1988 and Efstathiou et al 1985) with separation equal or lower than a cell.

- **Tree**: In high resolution cosmological simulations the P^3M scheme is not a good chance when strong clustering occurs and the cost of the direct summation becomes dominant. Therefore the direct summation is usually replaced by a tree code (Xu 1995) scheme in which the particle distribution is organized in a hierarchical tree structure. In tree codes the force between particles of distant regions is calculated using the centers of mass of these regions rather than calculating forces between every particle, strongly reducing the number of operations to $O(NlogN)$ (Barnes and Hut 1986, Bouchet and Hernquist 1988). The employment of a Fast Multipole Method permits to reduce the computational effort to $O(N)$, calculating forces between two nodes rather than between individual particles and nodes. This method require a force softening for gravitational interactions in order to avoid hard scatterings between nearby particles (typically either with a Plummer or a cubic spline kernel). The force resolution is then given by the softening length ϵ_{soft}.

3.2 Cosmological simulations

Despite the importance of the excursion set theory and the peaks formalism to understand some properties occurring in mass distribution, spatial clustering and assembly history of dark matter halos, it is clear that the future of structure formation will be hold by numerical simulations. Cosmological simulations are usually split in full-box and zoom-in simulations. Zoom-in simulations focus on one halo of interest in order to study its internal structure and its substructures with high resolution. Full-box (or cosmic scale) simulations usually cover a significant part of universe, employing only one particle mass and particle resolution and studying in this way the large scale structure and statistical properties of dark matter halos. Some important cosmic scale simulations are:

- **Millennium-XXL**. It was run in 2010 by the Virgo Consortium at the Julich Supercomputer Centre in Germany with the aim of studying the impact of galaxy formation on cosmology by making evolve 303 billion particles.

- **DEUS FUR**. It employs more than 10 billion particles within a box of $21h^{-1}Gpc$ (the entire observable universe). It uses a particle mass of $1.2\times10^{12}M_{\odot}$ that prevents the resolution of individual galaxies. This simulation was run with a modified version of the RAMSES code with the main purpose of studying the imprint of dark energy on the cosmic structure.

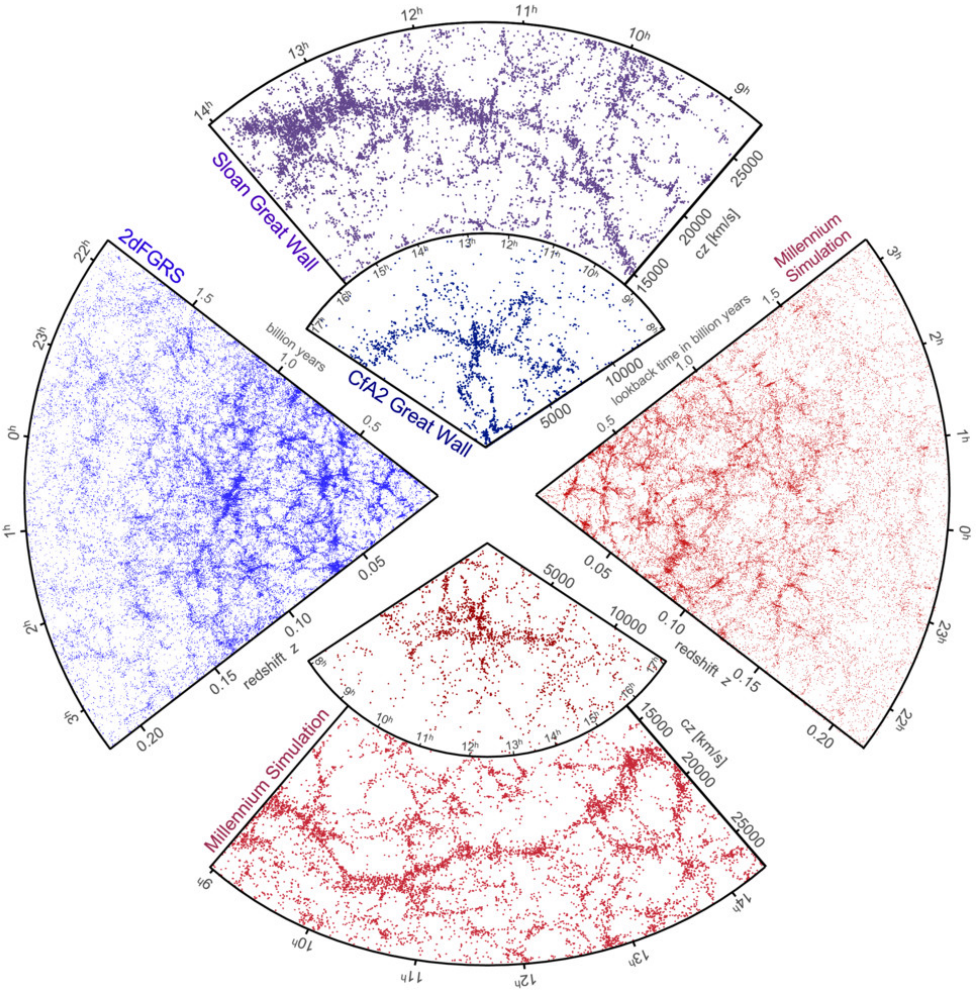

Figure 3.1: Comparison of the galaxy distribution in redshift surveys and in mock catalogues realized by applying semi-analytic galaxy formation to the assembly trees of the Millennium simulation (Frenk and White, 2012).

3.2.1 Le SBARBINE simulations

This work is based on data obtained employing four simulations, part of Le SBARBINE simulations cosmological set.

Le SBARBINE simulations set is composed by six Dark-Matter-only cosmological simulations run in 2013-2014 by the Padova cosmology group using the publicly available code GADGET-2. The initial conditions and the assumed background are consistent with recent Planck results (Ade et al., 2014): $\Omega_m = 0.30711$, $\Omega_L = 0.69289$, $\Omega_b = 0.04825$, $h0 = 0.6777$. They follow the evolution of 1024^3 particles and provide a good spatial and mass resolution by varying the box and the mass resolution accordingly to each other in order to avoid the risk of resolution effects (see also Despali et al. (2014)). The main parameters of the six simulations are listed in 3.1.
In this work we use only four simulations (Ada, Bice, Cloe, Dora) in order to have a good mass resolution on the mass range in which we are interested.

	N	*M (Ms/h)*	*Box (Mpc/h)*	*z*	*random*	*soft (Kpc)*
Ada	1024	1.94×10^7	62.5	124	919374	1.5
Bice	1024	1.55×10^8	125.	99	834512	3.
Cloe	1024	1.24×10^9	250.	99	726351	6.
Dora	1024	9.92×10^9	500	99	574656	12.
Emma	1024	7.94×10^{10}	1000.	99	129485	24.
Flora	1024	6.35×10^{11}	2000	99	564738	48.

Table 3.1: Main parameters of the cosmological simulation set Le SBARBINE.

3.2.2 Initial conditions

Starting from an initial uniform state ('glass' distribution) of particles, initial conditions are imposed by way of perturbations. The 'glass' distribution has been realized by evolving particles from a Poisson distribution using an inverted sign of Newton constant. The position of particles has been perturbed according to the growing mode solution of linear theory in order to obtain a Gaussian random field. The initial Power Spectrum was generated using the CAMB Code (Lewis et al., 2000) (see results for different simulations in Fig. 3.2).

Initial conditions are produced by perturbing a glass distribution with N-GenIC, which requires as input parameters:

- the initial Power Spectrum
- a glass file
- the properties of simulations (like the mass resolution, the box size, ...)

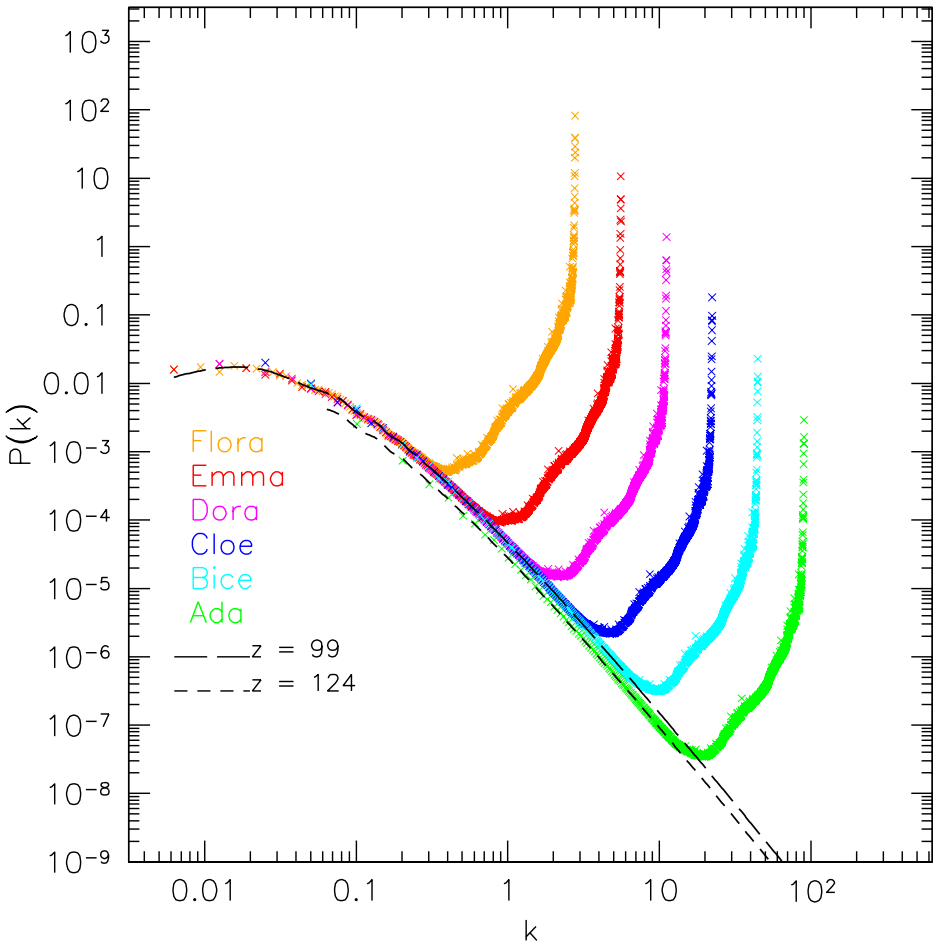

Figure 3.2: Initial power spectrum of the six simulations. The two black shaded lines show the input linear power spectrum at z=99 and z=124.

- a random number used to generate the initial particles positions from the glass file

3.3 Halo properties

In this section we briefly summarize the main properties of dark matter halos, taking into account their internal structure, their formation histories and their abundance and distribution. We focus mainly on the properties inspected with the employment of N-body simulations.

3.3.1 Halo formation

As we discuss in the previous chapter, the first structures to collapse are small. Larger structures form by merging of smaller ones and by accreting matter from the intercluster medium. Most mergers are minor, not altering the halo structure. Major mergers (involving halos of similar mass) can bring material to the inner core, but without modifying critically the internal mass distribution (Wang and White 2009). An inside-out formation scenario is generally assumed: a bound core collapses earlier and then other material accretes.

Halo formation time

Lacey and Cole (1993) found an expression for the halo formation times using halo counting argument. They define the formation time as the redshift at which half of the final halo mass has been assembled in a single progenitor. Their equation has the form:

$$\boxed{\delta_c(t_f) = \delta_c(t_0)\tilde{w}\sqrt{\sigma^2(M_0/2) - \sigma^2(M_0)}} \tag{3.2}$$

where the scaled variables

$$\boxed{\tilde{w} \equiv (w - w_0)/\sqrt{S_h - S_0}} \tag{3.3}$$

and:

$$\tilde{S} \equiv (S - S_0)/(S_h - S_0) \tag{3.4}$$

permit to express the distribution of formation times independently of the spectral index n in the range of physical interest. Knowing $\delta_C(t_f)$ the expression is easy to solve. The work of Lacey and Cole (1993) lays on the assumption of a spherical collapse for halos. However, this choice leads to underestimate the redshift of halo formation (Lin, Jing and Lin, 2003; Giocoli et al. 2007)). Giocoli et al. (2007) searched for a better description of halo formation using a model based on the ellipsoidal collapse model. A comparison with GIF2 numerical simulations showed that the EC prediction describes quite well the median formation redshift. The fact that their predicted expressions are broader than the one found in simulation is ascribable to the assumption of uncorrelated steps in the excursion set walks (Sheth and Tormen, 2002), whereas some correlations between halo

formation and environment have been found (Sheth and Tormen (2004); Gao et al. (2005); Harker et al. (2006); Wechsler et al. (2006)).

Giocoli et al. (2012a) found a fitting formula useful for different definition of formation times in terms of f:

$$\boxed{P(> w_f) = \frac{\alpha_f}{e^{w_f^2} + \alpha_f - 1}} \tag{3.5}$$

where:

$$\alpha_f = 0.815 e^{-2f^3/f^{0.707}} \tag{3.6}$$

.

3.3.2 Halo density profile and concentration

The first simulations studying the formation of elliptical galaxies found virialized structures with de Vaucouleurs or Einasto type density profiles, similar to observations of elliptical galaxies. Self-similar infall models predict scale free, nearly isothermal profiles $\rho \propto r^{-2}$, whereas simulated profiles are steeper in the outer parts and shallower in the inner regions. CDM halos have some similarities with spherical infall models:

- particles that collapse in early peaks are located closer to the potential minima of the entire turn-around region and end up closer to the center of the final halo
- the typical particle apocenter distances are close to their turnaround radii.

Because of this, modified infall models can reproduce some of the features of halo density profiles found in cosmological simulations.

Over a large mass range the spherically averaged CDM halo density profile can be approximated with the same universal form (Navarro and White, 1996):

$$\boxed{\rho(r) = \frac{\rho_{crit}\delta_c}{(r/r_s)^\gamma (1 + r/r_s)^{3-\gamma}}} \tag{3.7}$$

where $\gamma = 1$, the scale radius r_{-2} is related to the peak circular velocity scale by $r_{Vmax} = 2.163 r_s$, ρ_{crit} is the critical density of the universe and δ_c is a dimensionless parameter related to concentration by

$$\boxed{\delta_c = \frac{200}{3} \frac{c^3}{[ln(1+c) + c/(1+c)]}}. \tag{3.8}$$

The halo concentration has been defined in different ways. One of the most common methods involves the scale radius of the NFW profile and the halo virial radius:

$$\boxed{c_{vir} = \frac{r_{vir}}{r_s}} \tag{3.9}$$

Therefore, the NFW profile can be fully specified by the concentration and the halo mass. These two parameters have shown a correlation in simulations: the average concentration of a halo decreases weakly as a function of mass. Therefore, the NFW density profile can be described by a single free parameter, the concentration, which can be related to virial mass (Duffy et al. (2008)):

$$c = 6.67(M_{200}/2 \times 10^{12} h^{-1} M_{\odot})^{-0.092} \tag{3.10}$$

for relaxed halos. Halo concentrations are related to the halo formation time, so that halos that form first tend to have higher c_{vir} and ρ_s at $z = 0$. Small halos form earlier than bigger ones and so their concentration tends to be higher.

The universal density profile in the form of Navarro-Frenk-White is believed to arise from the inside-out process of halo formation. A possible explanation is that the relaxation mechanism that produces equilibrium is independent of the initial conditions and merger history (NFW 1997). However Syer and White (1998) proposed that the NFW profile could arise by the halo mass history, suggesting a possible universal halo mass history profile (Sheth and Tormen (2004);Giocoli et al. (2012a); Salvador-Solé et al. (2012)). Ludlow et al. (2013) compared the mass accretion history, expressed in terms of the critical density of the Universe, with the NFW density profile, expressed in units of enclosed mass and mean density within r, in a mass-density plane, showing that the halo mass histories, if opportunely scaled, follow the NFW profile. Correa et al. (2015) try to link the halo mass history with the halo concentration.

High resolution numerical simulations resolving scales around $0.01 r_{200}$ found systematic deviations from the NFW profile, with much denser inner parts (Navarro et al. (2004); Hayashi and White (2008); Navarro et al. (2010); Navarro et al. (2010); Diemer and Kravtsov (2014)), leading sometimes to the use of Einasto profiles where the logarithmic slope is a power law of radius (Einasto, 1965):

$$\boxed{dln\rho/dlnr \propto (r/r_s)^{\alpha}} \tag{3.11}$$

The Einasto profile fits better not only in the center suggesting that the addition parameter α really characterizes dark matter halos (Merrit et al. 2006; Klypin et al. 2014). A very interesting result of Ludlow et al. (2003) is that halos white profiles well approximated by Einasto form, present also accretion histories that deviate from the NFW shape in a similar way.

3.3.3 Halo bias

A fundamental problem in cosmology is the relation between the spatial distribution of galaxies and galaxy clusters and the underlying matter. In the standard cosmological model galaxies are assumed to form within dark matter halos. Therefore the problem of galaxy biasing can be tackled by first understanding how dark matter halos are distributed respect to the dark matter distribution (Mo and White 1996; Catelan et al. 1997; Sheth

and Lemson 1999). The initial Gaussian distribution of density fluctuations is completely described by its power spectrum. Therefore, it is possible to use the statistics of the initial conditions in order to understand the clustering of halos once they are associated to specific regions in the initial density field. This approach has been used by Kaiser (1984) in order to explain the clustering of the Abell clusters in the light of the peaks statistics in an initial Gaussian field. Bardeen et al. (1986) developed this formalism, showing that the "galaxy biasing" arises when galaxies are associated to high peaks of the initial density field. Some issues emerging in Bardeen et al. (1986) are discussed by Bond and Myers (1996) and Mo and White (1996).

The halo fluctuations can be related in the linear bias model to the local matter density fluctuations

$$\delta_h = b\delta \tag{3.12}$$

where $\delta \equiv \delta\rho/\vec{\rho}$ is the overdensity and b is the bias parameter. However the halo clustering does not depend only by the initial peaks clustering (δ_L) but also by the motion of peaks along large scale flows of matter. At later time and linear order the previous equation becomes:

$$\delta_h(a) = \delta_L + \delta_m(a). \tag{3.13}$$

On large scales the second term is simply the matter overdensity and we can relate the halo and mass field at the time in which halos are identified (Eulerian bias) as:

$$b(a) = \delta_h(a)/\delta_m(a) = 1 + \delta_L/\delta_m(a) = 1 + b_L(a). \tag{3.14}$$

While the matter overdensity depends on time, the Lagrangian clustering of peaks does not, only reflecting the clustering of peaks in the initial conditions. It follows that the former one will grow over time and the latter will become less important. So, a halo, assigned in the PS formalism to a Lagrangian patch with coordinates q, will not collapse in the same position but in the Eulerian point:

$$\mathbf{x}(\mathbf{q}, z) = \mathbf{q} + \mathbf{S}(\mathbf{q}, z) \tag{3.15}$$

where $\mathbf{S}(\mathbf{q},z)$ is the displacement vector at the epoch of the halo collapse $z = z_f(\mathbf{q},M)$. The Lagrangian patches from which the halos originated can can be written as a Taylor series in the dark matter density fluctuation field. If the formation of halos is assumed to be influenced by quantities other than the local density, Lagrangian bias will be non local and these variables should be taken into account in the Taylor series.

Assembly bias

The term *Assembly Bias* is used to describe the fact that the clustering of halos depends not only on their mass but also on the details of their assembly history. Recent N-body simulations have found a dependence at fixed mass between halo formation times and the environment: halos in dense regions form earlier than halos in less dense environments (Sheth and Tormen (2004); Avila-Reese et al. (2005); Harker et al. (2006); Zentner

et al. (2005)). Using the Millennium Simulation of Springel (2005), Gao et al. (2005) demonstrate that the standard ΛCDM paradigm predicts a dependence of the spatial distribution of halos at fixed mass and their formation times. This dependence is negligible for high-mass halos ($M > M_*$) but increases for lower mass objects. This result seems to be incompatible with the Press and Schechter (1974) approach for the growth of structure that predicts that halo clustering depends only on halo mass. Moreover it contradicts the assumption that the galaxy content of a halo of given mass is statistically independent of its larger scale environment (Kauffmann et al. (1997); Peacock and Smith (2000); Benson et al. (2000) and others) that is an important assumption of the Halo Occupation Distribution (HOD) model. Some modifications to the Press-Schechter model have been explored in order to take into account the assembly bias. A dependence on the environment can be introduced using the ellipsoidal dynamics, where the time needed for a region to virialize increases with the initial shear (Sheth et al., 2001) and so the tidal field generated by the large scale environment produces an environmental effect (Wang et al., 2007). The way in which the ellipsoidal collapse model can produce an environmental effect similar to the one found in N-body simulations is extensively inspected with different approaches in Desjacques (2007) and Sandvik et al (2007). However in both cases some issues remain unsolved. Assembly bias has been inspected using the peak model by Dalal et al. (2008). At large scales they compare predictions from the statistics of Gaussian random fields and numerical simulations finding a very good match. At low masses they suggest that assembly bias arises mainly due to the cessation of mass accretion onto a subpopulation of small halos.

CHAPTER 4

Statistics of halos and parameters

In this chapter we present the basis of the analysis developed in following chapters. In section 4.1 we describe how the bound objects employed in our work (i.e. dark matter halos) have been identified and selected. We go through the main algorithms provided for the halo identification and we account for our choice. Moreover, we present halo relaxation criteria and we describe how halo mass centers and moving centers have been defined. In section 4.2 we explain how the deformation tensor elements have been calculated in the simulation initial conditions. From the eigenvalues of the deformation tensor we calculate the parameters of the ellipsoidal collapse and the traceless shear q, emerging from a perturbative analysis of the ellipsoidal collapse model. We show the distribution of these parameters calculated over all field and smoothed on different scales, and we compare our results with theoretical predictions.

4.1 Halos

4.1.1 Halo identification criteria

Some different methods can be found in literature to identify halos in numerical simulations. One of the most employed is the *Friends-of-Friends* (FoF) method (Davis et al. (1985)), a percolation algorithm that links particles closer than a certain value given by bd (where $b < 1$ is the linking length and d is the mean inter-particle distance). The halos are therefore identified on the base of physical proximity: they appear to be realistic, closely following the isodensity contours of the identified structures. The surface of bound structures is $\rho/\bar{\rho} \sim 3/2\pi b^3$ and so it is defined by a free parameter, the linking length b: all particles within a linking length from each other are linked into a single group. A common value for the free parameter is $b = 0.2$. If we assume a density profile decreasing

with r^{-2} for spherical halo, we obtain $\rho/\bar{\rho} \sim 60$ and a mean overdensity of about 180, very close to the density threshold of the spherical sollapse model ($\delta \simeq 178$). However, a problem of the FoF algorithm is that it may link together dynamically different halos into a single virialized structure. Moreover, it does not employ any method to check if the identified halos are physically bounded. This last problem is present also in the *Spherical Overdensity* (SO) method. However, contrary to the FoF algorithm, the SO method forces halos to have a well defined shape (a sphere) and avoids the risk of connecting different halos. Indeed, it works by growing spherical shells around the halo center (or a density peak) in order to find the radius at which the mass overdensity reaches the threshold

$$\delta_{vir} = \frac{\rho(< R) - \rho_b}{\rho_b} \tag{4.1}$$

where ρ_b is the background density of the Universe and $\rho(< R)$ is the mean density inside the radius R. Despite a more realistic algorithm has been provided by Despali et al. (2013) in order to identify triaxial halos (the *ellipsoidal halo finder* (EO)) and find more realistic shapes, the SO method reproduces sufficiently well the halo shape. Moreover, the centers of mass of halos identified with SO and EO methods are very similar. We then decide to employ the SO identification method because it is more suitable to compare halos with quantities smoothed with spherical filters.

4.1.2 Halo mass selection

The simulation with the highest mass resolution (Ada) has a particle mass of $1.9 \times 10^6 M_\odot$. We select halos with at least 10^3 particles; so, the smallest halo that we can resolve has a mass of about $1.9 \times 10^7 M_\odot$, nearly equal to $M_*/256$, where $M_* = 4.9 \times 10^{12} h^{-1} M_\odot$ is the typical mass at which a virialized halo forms at the actual cosmic time ($z = 0$). We then consider fifteen mass bins $[M/\sqrt{2}, \sqrt{2}M]$, where the mass M covers a range from $M_*/128$ to $128 M_*$. In Tab. 4.1 we show the number of particles with which halos of different masses are resolved in the the simulations. We consider halos identified at $z_{id} = 0$ together with three more identification redshifts: $z_{id} = 1$, $z_{id} = 2$ and $z_{id} = 4$. When the identification redshift grows, the typical mass forming at that z_{id} shifts towards smallest masses. As a consequence, the halo statistics in the highest mass bins is highly reduced. Nevertheless, we consider the same mass bins for computational ease and we compare results at different redshifts using scaled units $\nu = \delta_{SC}(z_{id})/\sigma(M)$, where

$$\delta_{SC;z_{id}} = \delta_{SC}(z = 0) \frac{D(z_{id})}{D(z = 0)} \tag{4.2}$$

is the threshold of the spherical collapse model rescaled with the growth factor $D(z_{id})$ at the different redshifts and $\sigma^2(M)$ is the variance of the initial density fluctuation field when smoothed on the scale $R = (3M/4\pi\bar{\rho})^{1/3}$. By definition, the typical mass forming at each redshift corresponds to $\nu = 1$. So, at $z_{id} = 0$ the selected mass bins selected are distributed around this value, whereas at larger redshifts the mass range expressed in units of ν is

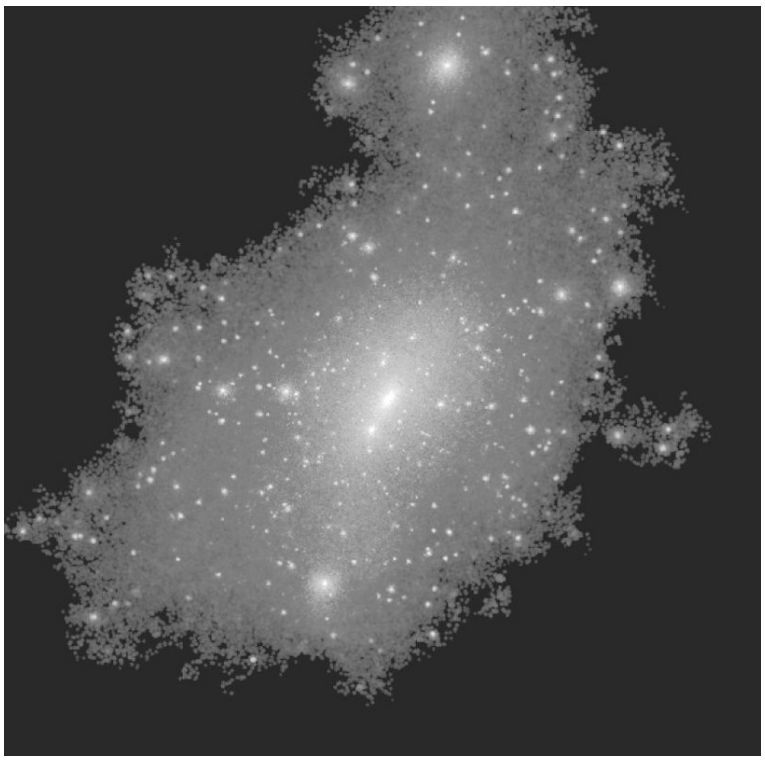

Figure 4.1: Friends-of-friends group containing a cluster. Particles have been weighted proportionally to the local dark matter density in order to highlight the substructures (Springel et al., 2001a).

always larger. So, we can investigate a large range of ν but a comparison of halo properties at different redshifts can be addressed only in a small range. Moreover, the statistics of halo at the highest masses happens to be largely reduced as the identification redshift grows. Tab. 4.1 shows the range of masses spanned by each simulation together with the corresponding value of ν at the different identification redshifts. Given the different mass resolution, each simulation covers a different mass range. In what follows, we consider only mass bins with at least 30 halos in order to have a sufficiently large statistics. The number of halos identified at different masses and identification redshifts are listed in A.1. We remark that in this work the notation M_* refers exclusively to the typical mass forming at $z = 0$. When comparing different redshifts we will discuss preferably in terms of ν to avoid any misunderstanding.

4.1.3 Relaxation criteria

Halos identified at different redshifts not only cover a different range in masses but also have a different formation history. For this reason some halos could present oddities that

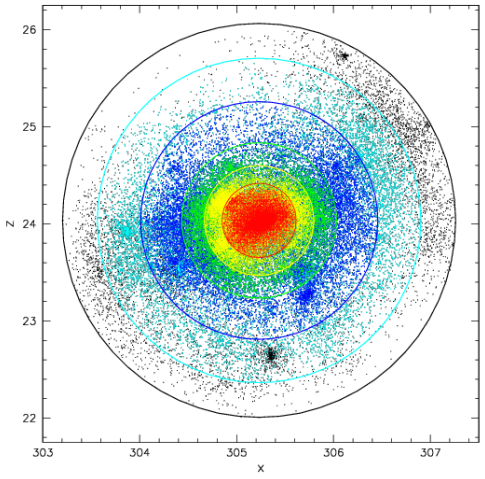

Figure 4.2: Halo identified with a Spherical Overdensity algorithm in the Dora simulation, part of Le SBARBINE cosmological set (Despali).

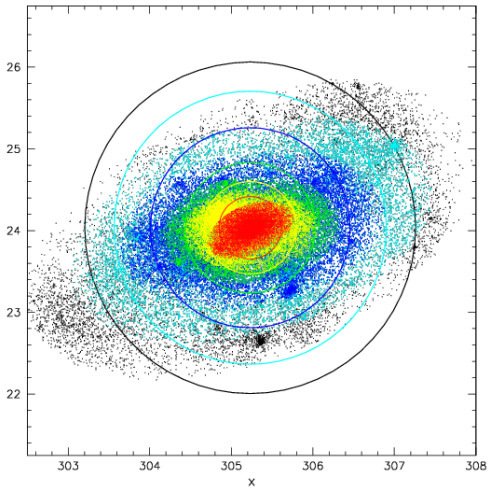

Figure 4.3: Halo identified with a Ellipsoidal Overdensity algorithm in the Dora simulation, part of Le SBARBINE cosmological set (Despali).

	Number of particles									
	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$
Ada	7903	15807	31613	15807	31613	63226	126452	252904	505808	
Bice				1976	3952	7903	15807	31613	63226	126452
Cloe							1976	3952	7903	15807
Dora										1976

	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Bice	252904	505808			
Cloe	31613	63226	126452	252904	505808
Dora	3952	7903	15807	31613	63226

Table 4.1: Number of particles with which the four simulations employed in this work resolve the fifteen characteristic masses. We show only mass bins spanned by the different simulations.

make them unsuitable for our analysis and make our results more difficult to interpret. We consider some relaxation criteria and in the next chapers we often compare results obtained with both all and selected halos. In this way we can understand how much non relaxed halos affect final results. The selection criteria we take into account are the following:

1. The mass of the main progenitor must not surpass the mass of the final halo for more than 10% at any time.

2. The total energy must be negative.

3. The center of mass of mass offset from the potential minimum must be less than 5% or 10% the virial radius.

4. The halo mass does not decrease more than 10% between two consecutive snapshots.

5. The mass of a halo at the fourth from last snapshot must be at least $10^{-1.5}$% of the final mass i.e. the halo must not accrete too much mass in the latest four snapshots.

The first criterion ensures that the halo does not lose too much mass before its identification. The second criterion is related to the binding energy of the halo; in fact, SO identification criterion does not assure that a halo is physically bounded. The third criterion is related to the distance between the center of mass and the most bound region and it assures that the halo shape has not been strongly distorted during recent merger events. Neto et al. (2007) use a similar criterion (more specifically, they impose an offset lower than 7% with respect to the mass center of FoF halos) and they find that this criterion removes the vast majority of unrelaxed haloes. The last two criteria are restrictions on the merger history rate. Clearly, some halos can be affected by more than one restriction. We investigate these criteria mixed in different combinations. In Fig. 4.4 we show the fraction of relaxed halos respect to the total number as a function of ν for halos with $z_{id} = 0$. We notice that the first two criteria affect mainly low mass halos, whereas the importance of the third criterion grows with mass. When considering the first three criteria together we notice that the fraction of halos is almost mass independent; a similar trend (but a slightly lower fraction) has been found at larger z_{id}'s. When taking into account all five criteria together we reduce drastically the number of halos to be considered relaxed. We generally use the first three criteria together in order to reduce the number of non relaxed halos equally with the mass and to maintain a good statistics. Unless otherwise specified, the wording *relaxed halos* will refer to halos selected with this combination. However, in this work we often employ different combinations of relaxation criteria in order to inspect how contraints on halo properties and formation history affect our measures. The number of relaxed halos at different masses and identification redshifts is listed in Tab. A.2.

In the last section we show how the different relaxation criteria affects the distribution of formation times.

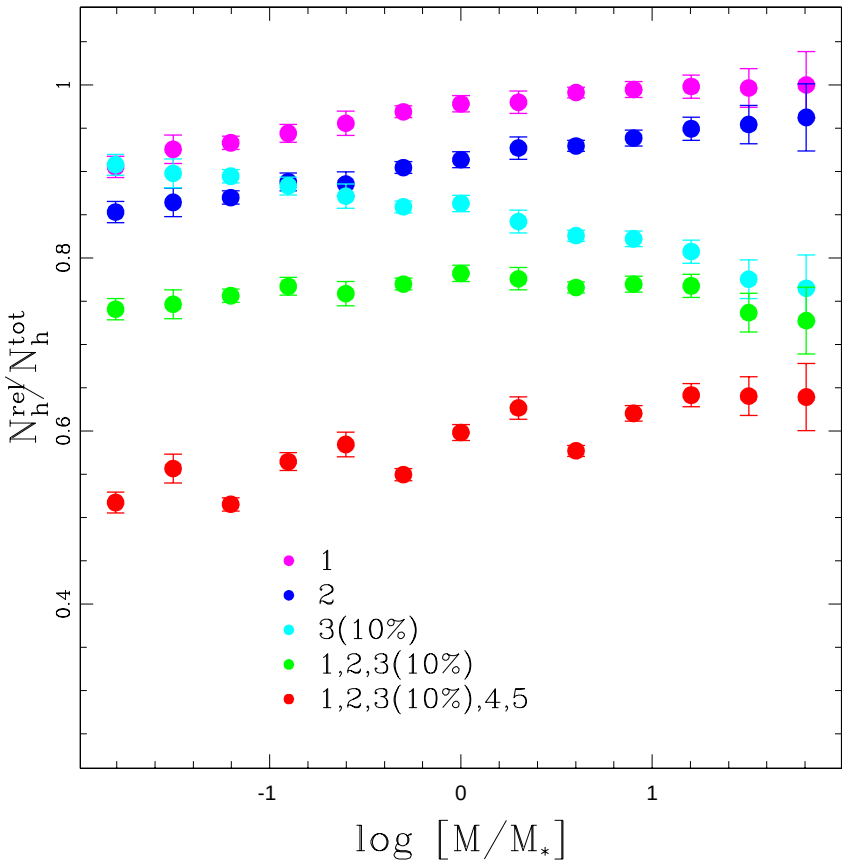

Figure 4.4: Fraction of relaxed halos identified at $z_{id} = 0$. We consider separately the first three relaxation criteria and two combinations of different criteria. The error bar is given by $\sqrt{N_h^{rel}}/N_h^{tot}$. We notice that the third criterion selects halos equally with the mass; a similar trend (but a slightly lower fraction) has been found at larger z_{id}'s.

4.1.4 Definition of the halo center

Once halos have been identified, their particles can be traced back to different redshifts in order to study the time evolution of halo properties. Despite the spherical halo shape imposed by the SO method at the identification redshift, the bunch of halo particles at the different redshifts has generally a strongly irregular shape, depending on the halo formation history at different times. So, it can be useful to define two different halo centers:

- **Center of Mass (CM)**. Since the particle mass is costant in a given simulation, center of mass is simply obtained by averaging the distances of the different halo particles.
- **Moving Center (MC)**. This center marks the most strongly bounded region of the halos and is obtained in this way: we draw a sphere around the halo center of mass enclosing all the protohalo particles and we shrink it of a certain factor; at this point we calculate the center of mass of the halo particles within the shrunk sphere, we center upon that and we go on until the decreasing radius reaches a certain threshold. The mass center that we calculate in the last iteration is what we call the halo moving center.

The moving center is useful when inspecting the densest halo regions, whereas the center of a mass characterizes better the center of the whole protohalo and so it is more respectful of its shape.

In this work we generally use the center of mass because it is more consistent with our analysis. However, we employ sometimes both centers to check some halo properties.

4.2 Parameters

One of the aim of this work is to study the Lagrangian properties of dark matter protohalos. To do so we need to compare the halo seed locations with the distribution of certain quantities. These quantities are calculated from the eigenvalues of the deformation tensor, obtained by smoothing the initial density field using an appropriate window function.

4.2.1 Calculation of the deformation tensor

The deformation tensor at each point is defined as the second derivatives of the gravitational potential Φ and it could be seen as a tidal force. We use the code *smooth_Gauss.f90* in order to calculate the elements of the deformation tensor after interpolating the particle distribution on a grid. This code is split in two parts:

1. The general aim of the first part of the code is to read a discrete particle distribution and to output a 3-dim grid for displacements. Each side of the grid has 1024 cells,

so we have 1024^3 cells in the entire box: as much as the number of particles. At first the code reads the input parameters and the input data (i.e. positions and velocities of particles in the initial conditions). Since positions are in $h^{-1}Mpc$, velocities are converted to positions using $h = 70$. Now we calculate the unperturbed positions and we wrap to take into account the boundary conditions. The displacements are now expressed in units of the box, and rescaled to $z = 0$ using the scale factor $(a = 1/(z + 1))$. Finally the code interpolates the distribution on a grid. Different interpolation methods have been discussed in 3.1. In this work we compare the distribution of the deformation tensor elements using the different interpolations and we finally choose a tsc interpolation, as we discuss in 4.2.1. We then save the displacements in order to use them in what follows.

2. The second part of the code smooths the grid on certain scales using a window function and calculates the elements of the deformation tensors.
We smooth displacements with a Top Hat (TH) filter, working in the Fourier space for computational ease. In this space the filter takes the form shown in 1.14.

   In the Zel'dovich approximation the deformation tensor corresponds to the first derivatives of the initial displacements,

$$\zeta_{ij} = -\frac{\partial^2 \Phi}{\partial x_i \partial x_j} = -\frac{\partial \psi}{\partial x}. \tag{4.3}$$

   Hence, for each grid point, we calculate the nine elements of the deformation tensor: D_{xx}, D_{xy}, D_{xz}, D_{yx}, D_{yy}, D_{yz}, D_{zx}, D_{zy}, D_{zz}.

Protohalos and radii of interest

One of the aim of this work is the study of the distribution of protohalo Lagrangian properties. We define protohalos as follows. Let's consider a halo identified at z_{id} within a bin of characteristic mass M. The protohalo is obtained by tracing the halo particles back to $z = 99$, when the density field can be considered to be linear. We choose smoothing scales corresponding to the characteristic bin masses M shown in Tab. 4.1. Since $\sigma^2(M) << 1$ at $z = 99$, the smoothing scales can be calculated as

$$R_L = \sqrt[3]{\frac{3M}{4\pi\rho_0}} \tag{4.4}$$

where ρ_0 is the initial density. Tab. 6.1 shows R_L in units of $h^{-1}Mpc$ together with the number of grid points spanned by these scales in the different simulations.

Symmetrization of the deformation tensor

The corresponding symmetrical elements of the deformation tensor are: D_{xy} vs D_{yx}; D_{xz} vs D_{zx}; D_{yz} vs D_{zy}. The matrix should be symmetrical and so the six elements are

Smoothing scales										
	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$
$M(10^{12}h^{-1}M_\odot)$	0.39	0.78	0.156	0.312	0.625	1.25	2.45	4.9	9.8	19.6
$R[h^{-1}Mpc]$	0.48	0.6	0.76	0.95	1.2	1.51	1.90	2.40	3.02	3.80
$R_A[gp]$	7.78	9.81	12.36	15.57	19.62	24.72	31.14	39.24	49.43	
$R_B[gp]$				7.78	9.81	12.36	15.57	19.62	24.72	31.14
$R_C[gp]$							7.78	9.81	12.36	15.57
$R_D[gp]$										7.78

	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
$M(10^{12}h^{-1}M_\odot)$	39.2	78.4	156.8	313.6	627.2
R [Mpc]	4.79	6.04	7.6	9.6	12.08
$R_A[gp]$					
$R_B[gp]$	49.43				
$R_C[gp]$	19.62	24.72	31.14	39.24	49.43
$R_D[gp]$	9.81	12.36	15.57	19.62	24.72

Table 4.2: Lagrangian scales in units of Mpc and corresponding number of grid points for the different simulations. Capitol letters specify the different simulations: A=Ada; B=Bice; C=Cloe; D=Dora.

supposed to be very similar. Fig. 4.5 shows a comparison between the corresponding elements of the deformation tensor obtained with GIF2 simulations. More specifically, we consider a box slice and we plot symmetrical elements calculated employing all three interpolation methods. We compare results using three smoothing radii: 1,5 and 9 grid points (given a box side of $110h^{-1}Mpc$ for GIF2 simulations, these scales correspond to about 0.28, 1.38 and 2.48 $h^{-1}Mpc$). The dispersion of the values decreases going towards larger smoothing scales and is minimum when using the tsc interpolation. We then decide to employ the tsc interpolation and moreover we average symmetrical values (see Tab. 4.3) in order to obtain a perfectly symmetrical deformation matrix.

Average of the symmetrical elements	
symmetrical elements	averaged element
D_{xy}, D_{yx}	$\bar{D}_{XY} = \frac{D_{xy}+D_{yx}}{2}$
D_{xz}, D_{zx}	$\bar{D}_{XZ} = \frac{D_{xz}+D_{zx}}{2}$
D_{yz}, D_{zy}	$\bar{D}_{YZ} = \frac{D_{yz}+D_{zy}}{2}$

Table 4.3: Symmetrization of the deformation tensor elements.

Distribution of the elements of the deformation tensor

As we explained earlier, the deformation tensor is very important for this work. Indeed we use its eigenvalues to calculate the parameters of the ellipsoidal collapse model (δ, e

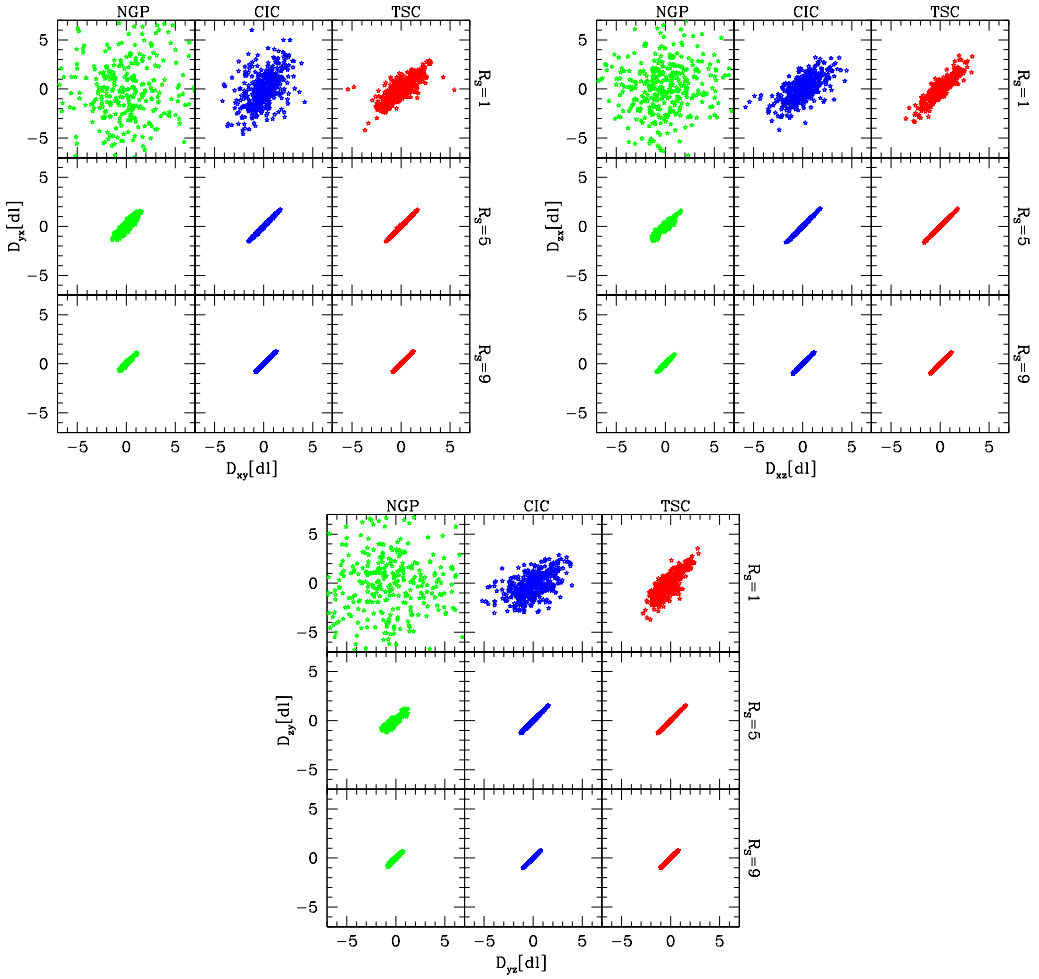

Figure 4.5: Comparison between D_{xy} and D_{yx}, D_{xz} and D_{zx} and D_{zy} and D_{yz} obtained with tsc interpolation in GIF2 simulations. The dispersion of the values is minimum when considering larger smoothing radii and when using a tsc interpolation.

and p) and the traceless shear parameter q. The elements of the deformation tensor are

$$D_{11} = (-y_1 - 3y_2/\sqrt{15} - y_3/\sqrt{5})/3 \tag{4.5}$$

$$D_{22} = (-y_1 + 2y_3/\sqrt{5})/3 \tag{4.6}$$

$$D_{33} = (-y_1 + 3y_2/\sqrt{15} - y_3/\sqrt{5})/3 \tag{4.7}$$

$$D_{12} = D_{21} = y_4/\sqrt{15} \tag{4.8}$$

$$D_{23} = D_{32} = y_5/\sqrt{15} \tag{4.9}$$

$$D_{13} = D_{31} = y_6/\sqrt{15} \tag{4.10}$$

where the y_i are independent Gaussian variates with zero mean and variance $\sigma(M)$ (Bardeen et al. (1986), Sheth and Tormen (2002)). We notice that the mean is less than 10^{-6} for all the elements and the variance

$$\sigma^2_{D_{i,j}}(r_S) = \langle D^2_{i,j}(r_S)\rangle - \langle D_{i,j}(r_S)\rangle^2 \tag{4.11}$$

has been listed in Tab.B.1 for all simulations and smoothing scales. Fig. 4.6 shows how the (logarithm of) variance of the deformation tensor elements depends on the smoothing scale.

4.2.2 Eigenvalues and parameters of interest

In the ellipsoidal collapse model the evolution of a patch in the initial conditions can be described by the eigenvalues of the 3×3 deformation tensor. In this work we neglect the mass tensor and we assume that the final shape of the object is determined only by the initial deformation tensor. This means that tidal torques, induced by the misalignment between the deformation and the mass tensor, are subdominant (see Porciani et al. (2002)).

In the first part of this work we are interested in the study of parameters determined in the whole field. So, we calculate the eigenvalues of the deformation tensor elements on every grid point. We store the eigenvalues in descending order so that $\lambda_1 > \lambda_2 > \lambda_3$ and we study their distribution. To have an idea of how much their distribution moves away from a Gaussian we calculate skewness and kurtosis. The skewness (γ_1)is ultimately the third standardized moment and is defined as:

$$\gamma_1 = E[(\frac{X-\mu}{\sigma})^3] \tag{4.12}$$

where σ is the standard deviation. Asymmetry can be estimated in some different ways. We simply expresses skewness in terms of the non central moment $E[x^3]$:

$$\gamma_1 = \frac{E[X^3] - 3\mu\sigma^2 - \mu^3}{\sigma^3}. \tag{4.13}$$

We can have a negative skewness when the left side of the curve is longer than the right side; otherwise we have a positive skew. A symmetric distribution has a skewness equal to 0. We estimate the kurtosis using the fourth standardized moment of the distribution:

$$\beta_2 = \frac{E[(X-\mu)^4]}{([(X-\mu)^2])^2} \tag{4.14}$$

where σ is the standard deviation. The value of mean and variance for the eigenvalues can be seen in TabB.2. The mean has been predicted by the linear theory Doroshkevich (1970):

$$\bar{\lambda}_1 = \frac{3\sigma(M)}{\sqrt{10\pi}} \tag{4.15}$$

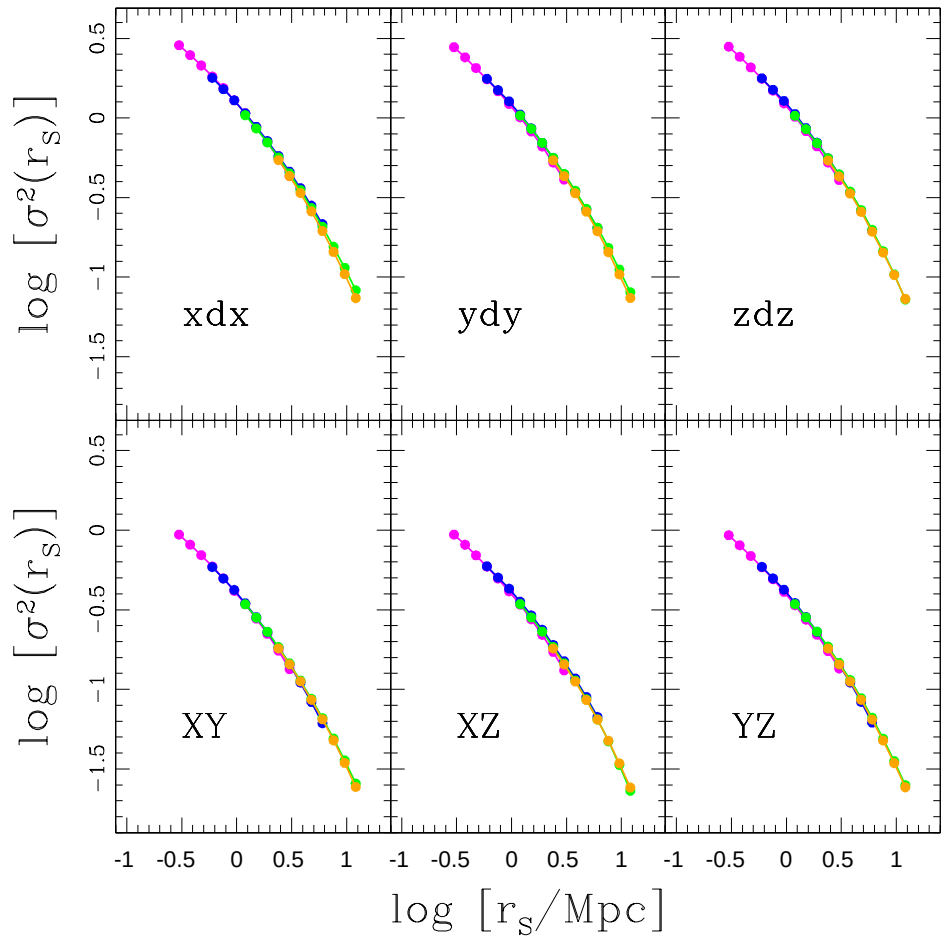

Figure 4.6: Variance of the three diagonal elements of the deformation tensor and the averaged symmetrical elements, obtained by smoothing with a top-hat filter. Different colors refer to different simulations: Ada (magenta), Bice (blue), Cloe (green) and Dora (orange). As expected from Eq. 4.5-4.10 the variance of the diagonal elements is larger than the variance of the symmetric ones.

$$\bar{\lambda}_2 = 0 \tag{4.16}$$

$$\bar{\lambda}_3 = -\frac{3\sigma(M)}{\sqrt{10\pi}} \tag{4.17}$$

and it is consistent with our results (see Fig.4.7 where we show the distribution of the eigenvalues for three different smoothing scales).

We can also compare (Tab. B.3) the number of triplets with concordant and discordant sign with the values found in literature. We found a perfect correspondence. As we explained earlier, these percentages mean that in a Gaussian fluctuation field about 42% of the initial volume forms sheets or filaments and the 8% forms void or spherical structures.

The parameters of the ellipsoidal collapse model are: δ, e and p. Although this work doesn't concern directly this theory of collapse, we use these quantities to inspect the initial conditions. We remind the derivation of these parameters from the eigenvalues of the deformation tensor.

- Density parameter:

$$\delta = \lambda_1(t_i)\lambda_2(t_i)\lambda_3(t_i) \tag{4.18}$$

- Ellipticity:

$$e = \frac{\lambda_1(t_i) - \lambda_3(t_i)}{2\delta(t_i)} \tag{4.19}$$

- Prolateness:

$$p = \frac{\lambda_1(t_i) + \lambda_3(t_i) - 2\lambda_2(t_i)}{2\delta(t_i)} \tag{4.20}$$

When $\delta > 0$, we have $e \geq 0$ and $|p| \leq 0$. The probability to have the three eigenvalues in descending order can be derived from the linear theory in the form:

$$p(\lambda_1, \lambda_2, \lambda_3) = \frac{15^3}{8\pi\sqrt{5}\sigma^6} exp\left(-\frac{3I_1^2}{\sigma^2} + \frac{15I_2}{2\sigma^2} \right) \times (\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_3) \tag{4.21}$$

where $\sigma \equiv \sigma(R_f)$,$I_1 \equiv \lambda_1 + \lambda_2 + \lambda_3$, $I_2 \equiv \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3$ and $\delta \equiv I_1$. The integration of $p(\lambda_1, \lambda_2, \delta - \lambda_1 - \lambda_2)$ over $(\delta - \lambda_1)/2 \leq \lambda_2 \leq \lambda_1$, and then over $\delta/3 \leq \lambda_1 \leq \infty$ leads to a Gaussian distribution for δ with variance σ^2 and mean 0, consistent with our results.

The equation of e and p has δ at the denominator. It follows that if δ tends to be zero, these quantities skyrocket. Although e and p are largely employed to describe the shape of the gravitational potential, some alternatives have the advantage of not depending on a ratio of eigenvalues. In this work we employ the square root of the traceless shear q, arising from a perturbative analysis of the ellipsoidal collapse model. It can be expressed using the parameters of the ellipsoidal collapse model

$$q^2 = I_1^2 - 3I_2 = \delta^2(3e^2 + p^2) \tag{4.22}$$

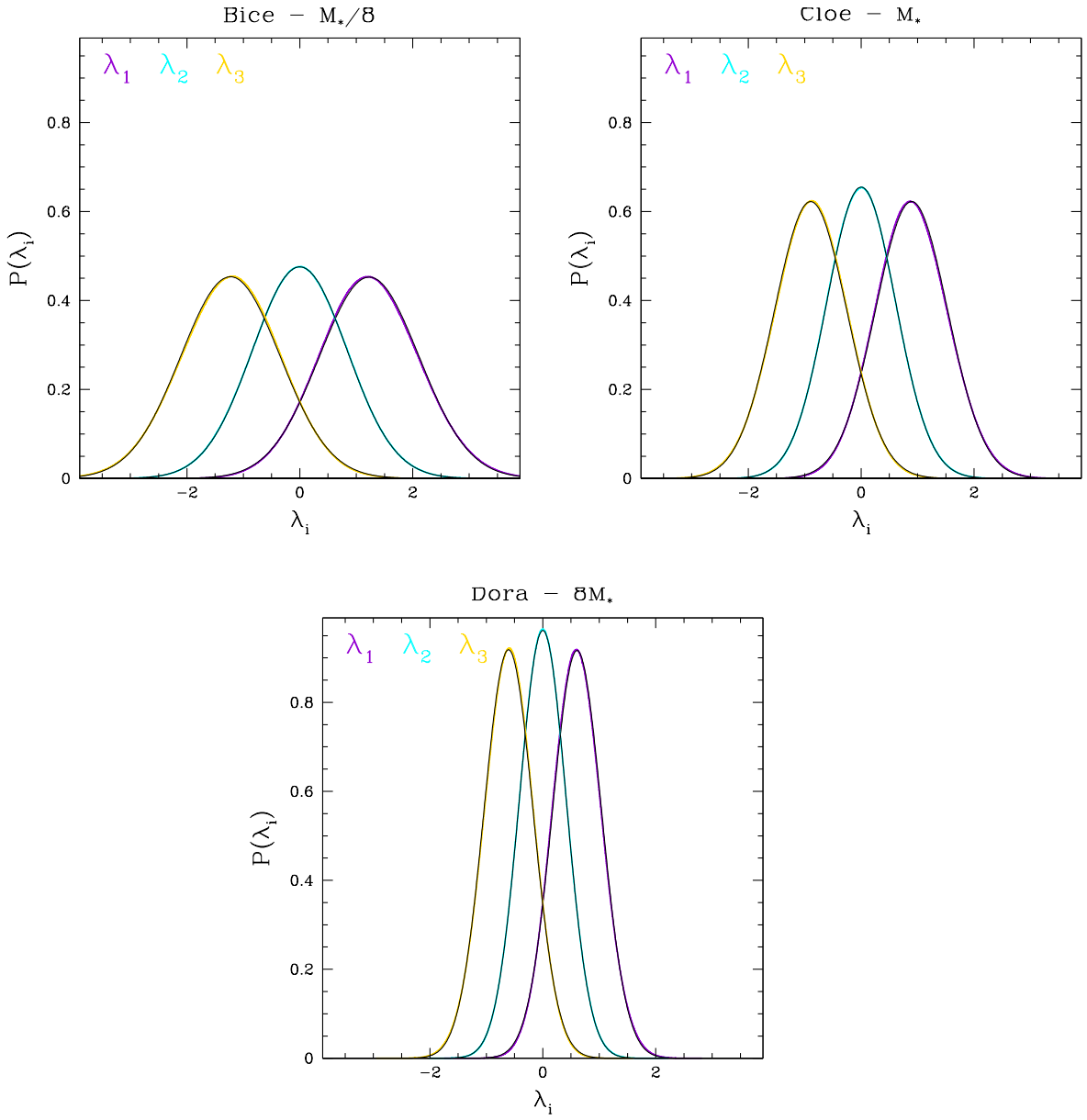

Figure 4.7: Distribution of $\lambda_1, \lambda_2, \lambda_3$ smoothed with three different scales using three different simulations: Bice, $R_S(M_*/8)$; Cloe, $R_S(M_*)$; Dora, $R_S(8M_*)$. The black curves are Gaussian with mean and variance as predicted by the linear theory (see also Tab. B.2).

where I_1 and I_2 are rotationally invariant quantities equal to

$$I_1 = \lambda_1 + \lambda_2 + \lambda_3 \tag{4.23}$$

$$I_2 = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 \tag{4.24}$$

(Sheth et al. (2013)). We remark that q, differently from e and p, is not coupled with δ and the distribution of q^2 is drawn from a $\chi^2_5(\sigma)$ and so $\langle q^2\rangle \sim \sigma^2$ (Sheth and Tormen (2002), where q is labelled as r). The distributions of δ and q calculated on all grid points, using different smoothing scales can be seen in 4.8.

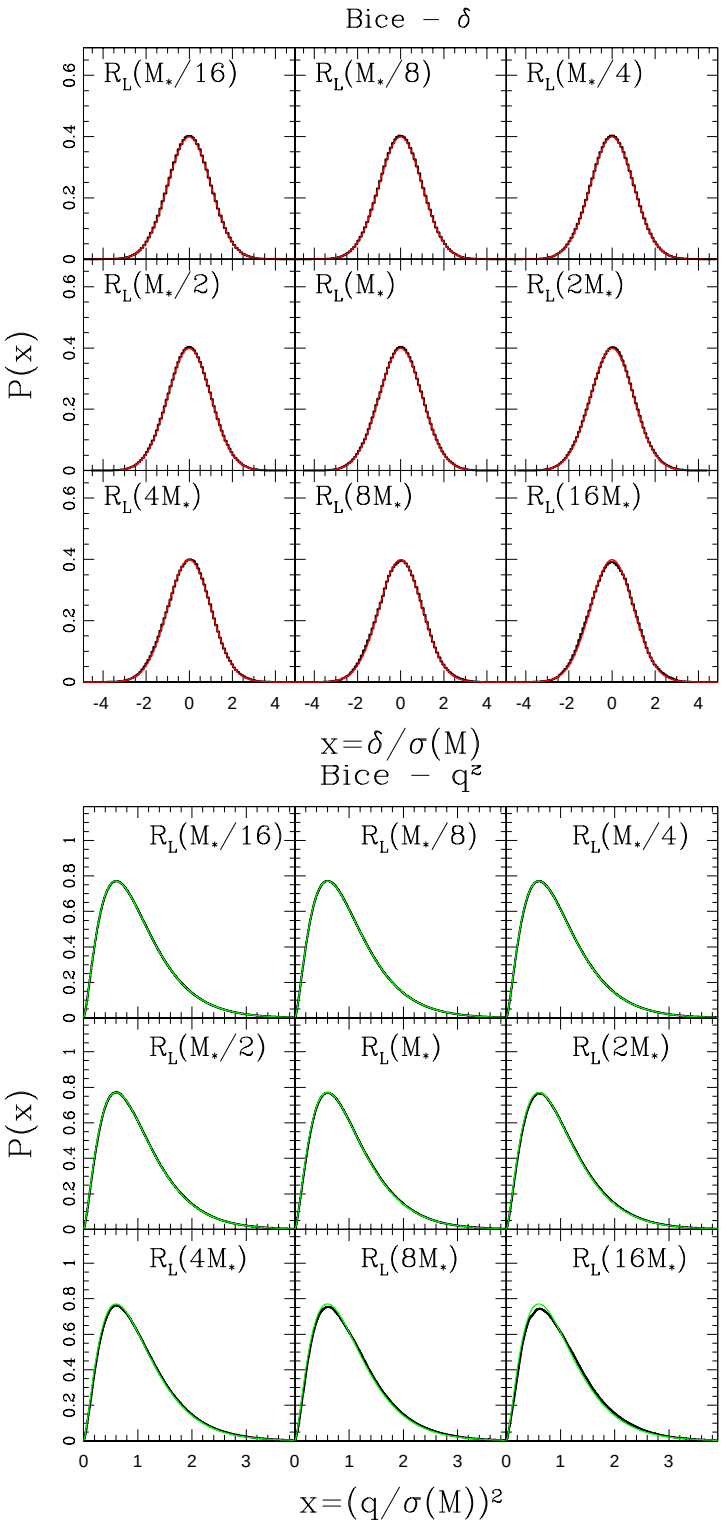

Figure 4.8: Distribution of $\delta/\sigma(M)$ (top) and $q/\sigma(M)$ (bottom) calculated on all grid points with different smoothing scales. δ distribution is described by a Gaussian (red curve), whereas q^2 is drawn by a χ^2 with five degrees of freedom (green curve).

4.3 Distribution of formation times for relaxed and non relaxed halos

The most important halo properties like the concentration and the halo structure are largely known to depend on the halo assembly history. The definition of a parameter quantifying the formation time of a halo is therefore very important. A typical approach defines the formation time of a halo as the redshift at which the halo main progenitor has assembled a fraction f of the final halo mass, $m_{MP}(z) > fM_0$, where the main progenitor is defined as the most massive progenitor of the most massive progenitor of the most massive progenitor... and it is therefore different from the most massive progenitor at a certain redshift (see also Giocoli et al. (2012a)). Given the importance of the formation redshift on relating halo parameters, it is useful to properly estimate its distribution as a function of the halo fraction f, M_0 and z_0. A scaled variable has been provided in order to absorb the dependence on M and f:

$$w_f = \frac{\delta_c(z_f) - \delta_c(z_0)}{\sqrt{\sigma(fM_0) - \sigma(M_0)}} \tag{4.25}$$

where $\delta_c(z)$ is the Spherical Collapse threshold extrapolated to the present time, $\sigma(M)$ is the mass variance smoothed on a scale $R = (3M/4\pi\bar{\rho})^{1/3}$ with $\bar{\rho}$ equal to background density. Different authors (Lacey and Cole (1993); Nusser and Sheth (1999)) estimate the shape of the distribution for $f \geq 1/2$ in the form

$$p_f(w_f) = \Big(\frac{1}{f} - 1\Big)\,\mathrm{erfc}\,\Big(\frac{w_f}{\sqrt{2}}\Big) + \Big(2 - \frac{1}{f}\Big)\sqrt{\frac{2}{\pi}}\exp\Big(-\frac{w_f^2}{2}\Big) \tag{4.26}$$

and Giocoli et al. (2012b) show that this equation fits also for smaller values of f. In this section we analyze separately the distribution of the formation time for both relaxed and non relaxed halos. We first show how skewness and kurtosis evolve with f and with different combinations of relaxation criteria. At a later stage we show how the distribution of formation times can be model out using a Weibull distribution once we take into account opportunely selected relaxed halos.

4.3.1 Importance of relaxation on the distribution of formation times

We consider the relaxation criteria and the halo masses presented in Section 4.1.3 and we study the distribution of relaxed and non relaxed halos using fourteen different values of the mass fraction f: 0.01, 0.02, 0.04, 0.1, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8 and 0.9.

Given a relaxation criterion (or a combination of more criteria), we firstly analyze how the distribution of formation times move away from a Gaussian distribution. To do so, we investigate the distribution of skewness and kurtosis as a function of the mass fraction f. We show the results in Fig.4.9 using relaxed (blue), non relaxed (red) and all halos (green). Different panels show results for different relaxation criteria, identified with numbers as

in Section 4.1.3. We show also the results of a Jarque-Bera test (full circles), testing the normality of the distribution using both skewness and kurtosis. When imposing the second relaxation criterium (i.e. the total energy must be negative; top right panel), skewness and kurtosis are strongly reduced for relaxed halos; whereas, the effect of the first two criteria is less important. When considering the first three criteria together (bottom left panel) both skewness a kurtosis are very similar to zero for $f \leq 0.6$. It is noticeable that at f=0.5, a value frequently used in literature, the distribution of relaxed halos is very similar to a Gaussian, with a small asimmetry due to the left constraint in zero. When including the last two criteria, what obtained is not significantly different (bottom right panel).

In the next section we study the distribution of formation times for relaxed halos selected using together all five relaxation criteria.

Distribution of relaxed halos

We consider all five relaxation criteria and we show the distribution of relaxed halos in the top plot of Fig. 4.10. As we discussed in Section 4.1.3, as long as f is small, the distribution is similar to a Gaussian (blue curve) with the same mean and the same variance of the distribution. As f grows, the average value of w_f decreases (i.e. when a larger fraction of assembled mass is assumed in the definition of the formation time, the halo forms later) and the distribution departs from a Gaussian cause of the left constraint present at zero. The green curve shows a fit with a Weibull distribution. By definition, a two-parameter Weibull is null for values lower than 0, whereas for positive values the probability density function is equal to

$$W(w_f; k, \lambda) = \frac{k}{\lambda}(\frac{w_f}{\lambda})^{k-1} e^{-(\frac{w_f}{\lambda})^k} \tag{4.27}$$

where $k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. We show that the Weibull distribution seems to fit qualitatively well for f larger than about 0.4 and, in this range, its residuals are smaller and more symmetric respect to the Gaussian ones (bottom plot).

Fig. 4.11 shows (in logarithmic scale) the results of a best fit procedure for the parameters of the Weibull distribution. It is remarkable that both the (logarithmic) scale and shape parameter scale linearly when $0.2 \leq f \leq 0.6$; this range includes the value $f = 0.5$, frequently used in literature. Therefore, in this range the parameters of the Weibull follow a similar power law, with exponent equal to about -0.25. At lower and higher f the scaling is still linear in logarithmic units but the slope is different.

Distribution of non relaxed halos

Fig. 4.12 shows the distribution of non relaxed halos. This distribution diverges from a Gaussian also for small values of f cause of a long tail towards very high w_f. When considering high f, it is difficult to model out the distribution with a two-parameter distribution

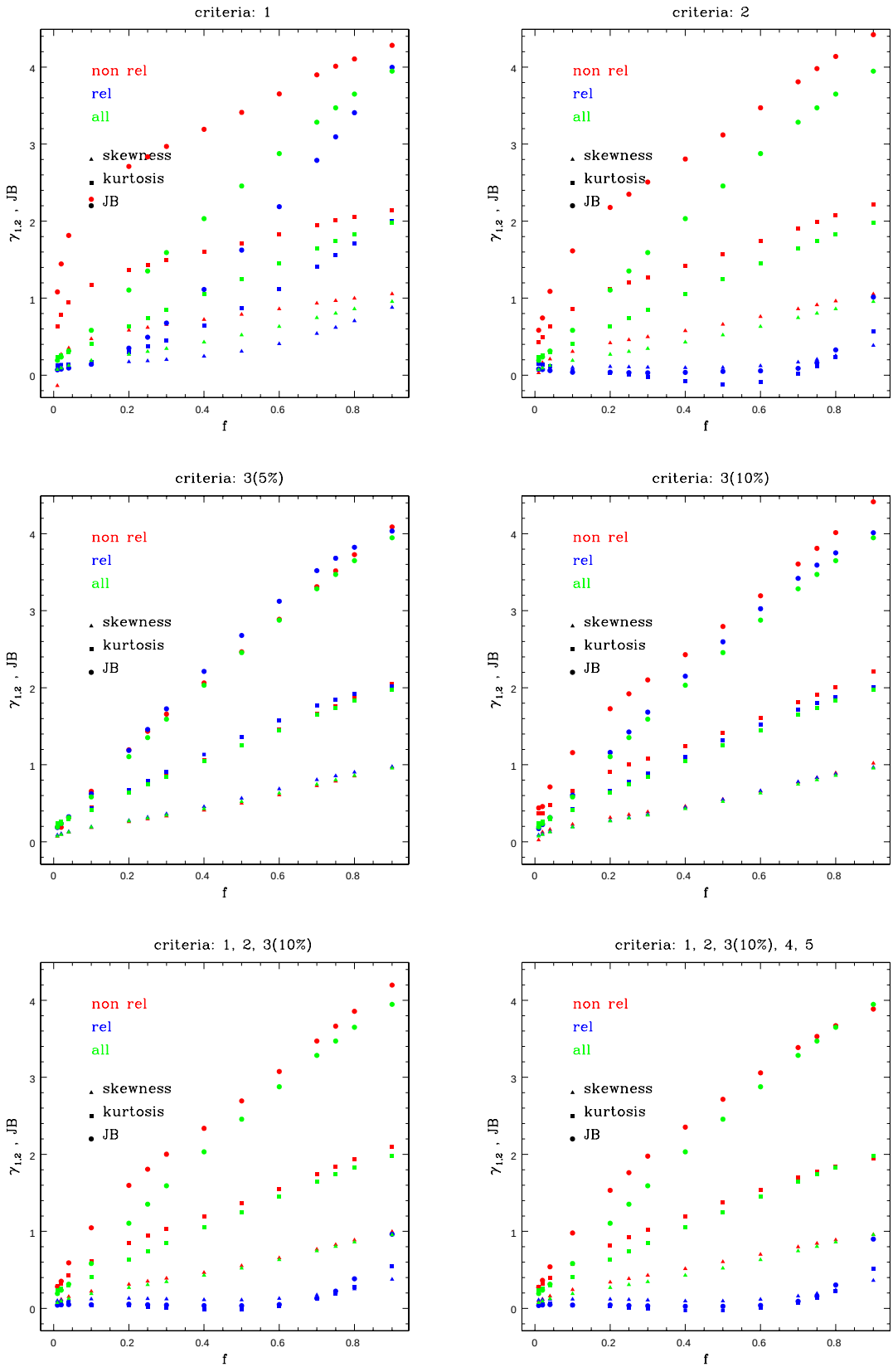

Figure 4.9: Skewness (triangles), kurtosis (squares) and a JB test (circles) as a function of the mass fraction f. We show results for relaxed (blue), non relaxed (red) and all halos (green). The six panels show results for different relaxation criteria, as defined in Section 4.1.3.

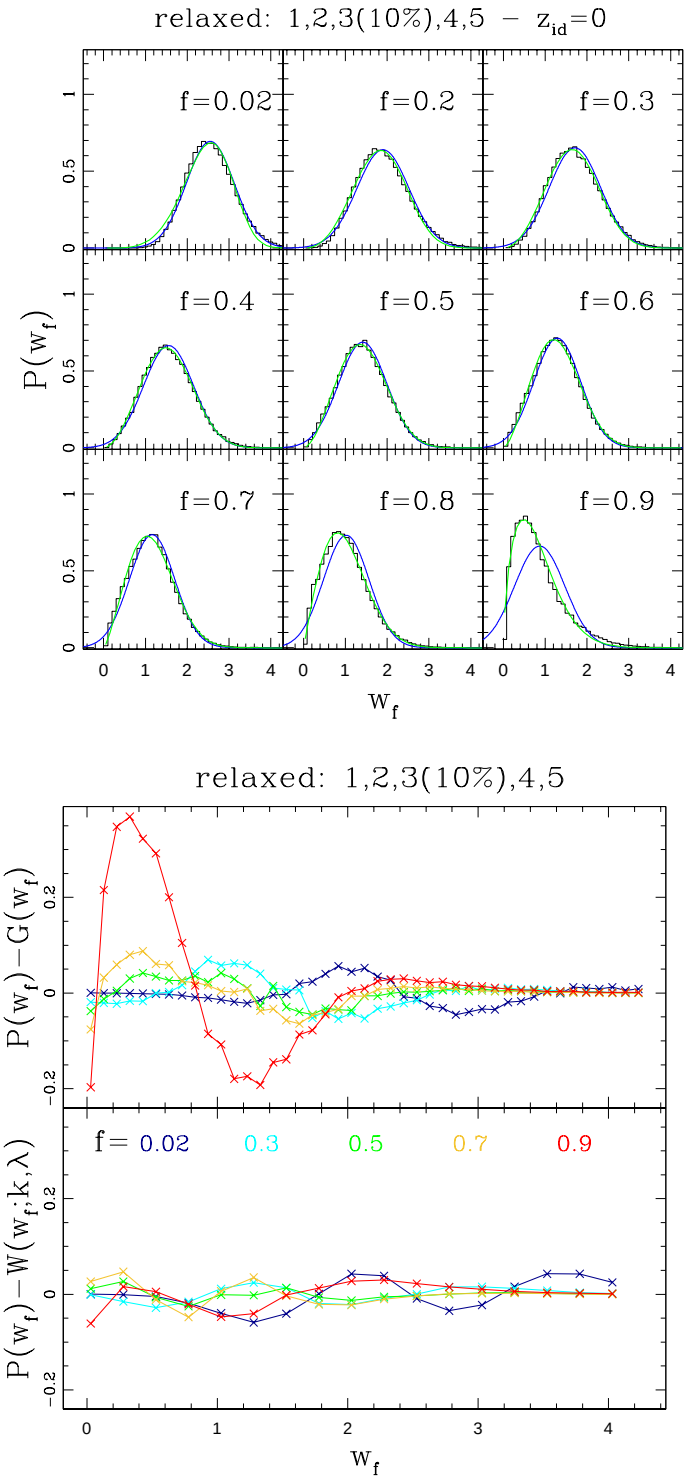

Figure 4.10: Top plot: Comparison of the distribution of formation times w_f of relaxed halos, a Gaussian (blue) and a Weibull (green) distribution. We consider nine different values of the mass fraction f and we select halos employing all five relaxation criteria. Bottom plot: residual of a Gaussian (top panel) with the same mean and the same variance of the distribution and residual of the best fit of a Weibull (bottom panel). We show results for five different values of the mass fraction f.

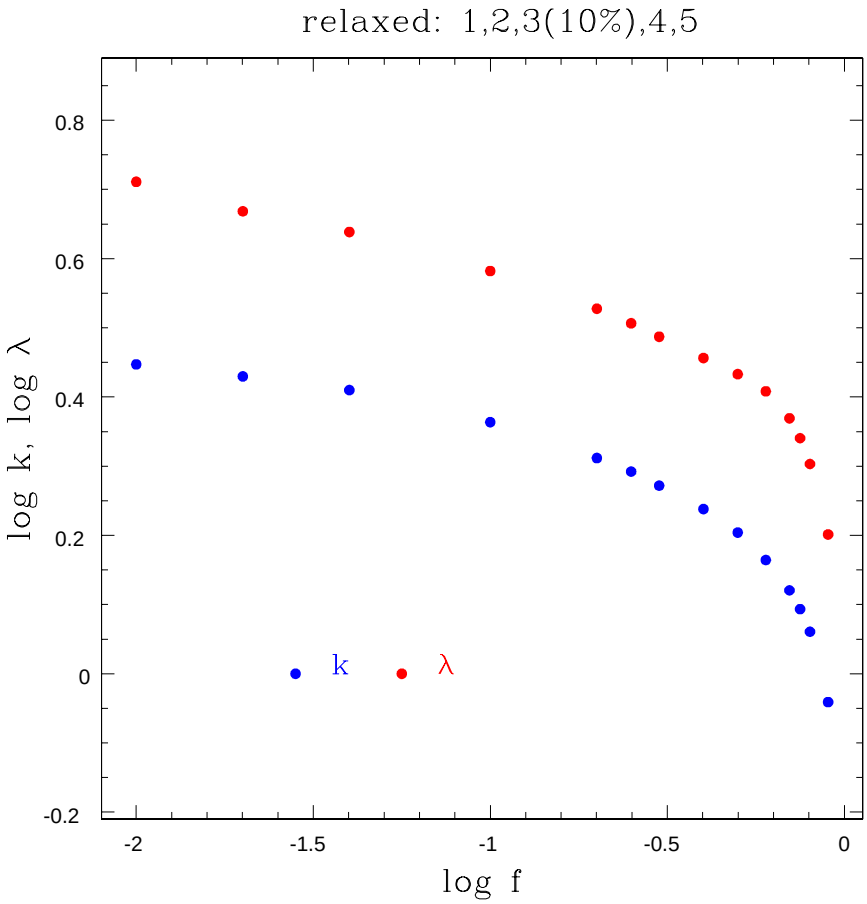

Figure 4.11: Logarithmic distribution of the shape parameter k (blue) and the scale parameter λ (red) as a function of the (logarithmic) mass fraction f. We take into account fourteen different values of f.

function. In fact the shape of the distribution at low w_f becomes strongly irregular and, moreover, it strongly depends on the chosen combination of relaxation criteria. We find similarities between the distribution of relaxed and non relaxed halos only when considering the highest f. In fact, at $f = 0.9$ both distribution are qualitatively well described by a Weibull.

4.4 Summary

In this Chapter we showed how we identified halos in our simulations and the adopted criteria in the selection of halo mass bin and relaxed halos. At a later stage we described how the parameters of the ellipsoidal collapse model have been calculated from the eigen-

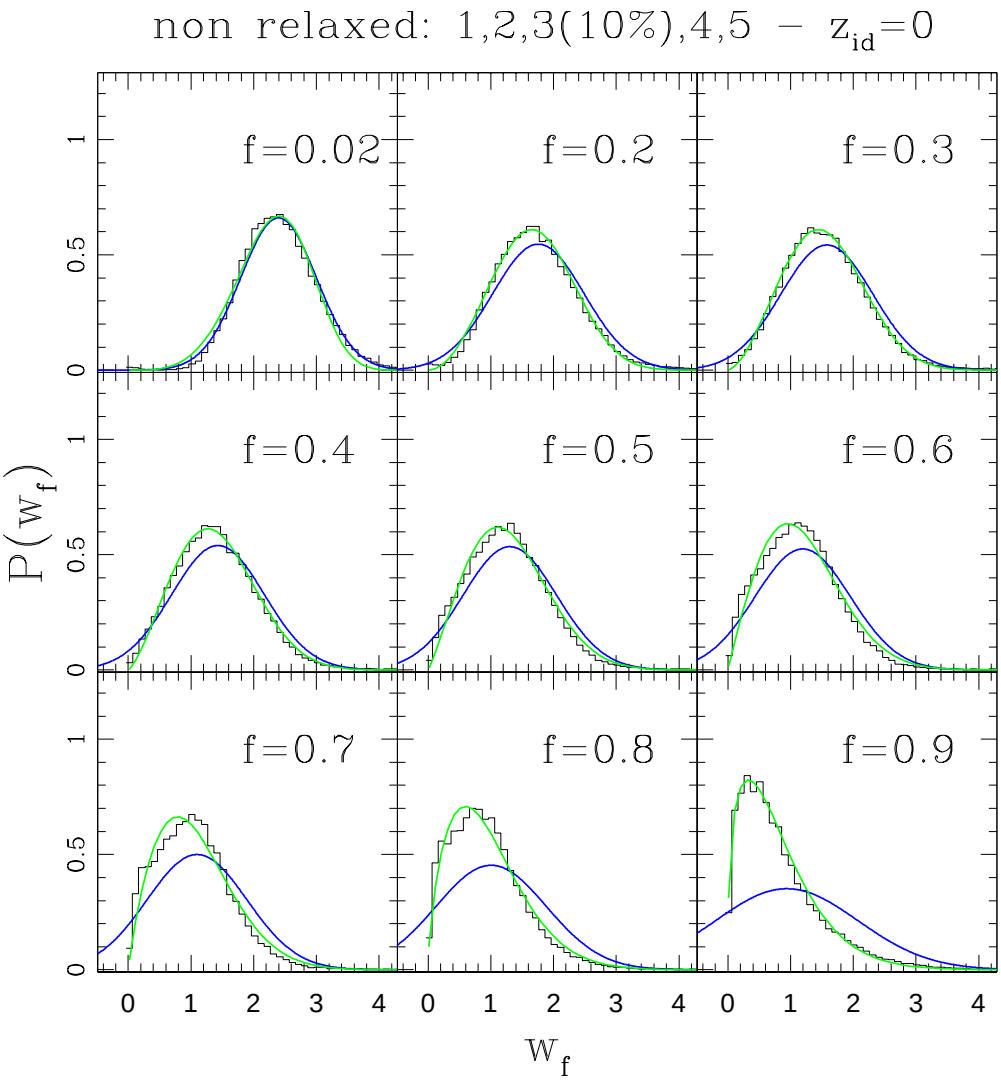

Figure 4.12: Top plot: Comparison of the distribution of formation times w_f of non relaxed halos, a Gaussian (blue) truncated at five σ and a Weibull (green) distribution. We consider nine different values of the mass fraction f and we select halos employing all five relaxation criteria. Bottom plot: residual of a Gaussian (top panel) with the same mean and the same variance of the distribution and residual of the best fit of a Weibull (bottom panel). We show results for five different values of the mass fraction f.

values of the deformation tensor. We then analyzed the halo formation time, opportunely rescaled, and we showed that their distribution is easier to model out when considering only relaxed halos. More specifically, we found that this distribution is well approximated by a Weibull, a distribution with two free parameters.

CHAPTER 5

The location of the halo formation

5.1 Correlation functions

The study of the halo formation can be addressed in different ways. In the *peaks formalism* (Bardeen et al. (1986), Bond and Myers (1996)), gravitationally bound objects form in correspondence of local maxima (i.e. peaks) of the linear density field smoothed on the Lagrangian halo scale. This formalism has been also used to study the clustering properties of halos (Desjacques (2008), Desjacques and Sheth (2010)) and their dependence on the formation time (Dalal et al., 2008). Although a check of the peaks formalism on a object-by-object basis have been done by some authors (Porciani et al. (2002), Ludlow and Porciani (2011)), this approach has been tested mainly on a statistical way.

In this chapter we firstly explain how we identify peaks on the initial conditions of simulations. We then investigate the correlation function between protohalo centers and peaks of δ smoothed on various scales. In this way we consider not only peaks close to the halo seeds but we can also investigate the peak distribution within a much larger region. We also sjow how the correlation function changes when considering constraints on the peaks height. Finally, we analyze the correlation function of protohalo centers with peaks and valleys on the distribution of the traceless shear parameter q. Our purpose is to understand how the constraint that the q distribution has a null gradient affects the formation of halos.

5.1.1 Peak identification on the DM distribution at the initial conditions

Peaks (valleys) are points where the scalar field characterizing a certain parameter and smoothed on certain scales has a local maximum (minimum). In what follows we are interested on peaks of the distribution of two parameters: δ and the traceless shear q. The

choice of the smoothing scales is related to halo mass bin (see Sec. 2 for details). Our algorithm works as follows: we center at each grid point and we calculate the value of a certain field smoothed on a certain scale. We then compare this value with those obtained smoothing the same field centering on the six adjacent grid points. Given a direction, if the value is larger (smaller) of those found on the two adjacent grid points, we have a maximum (minimum) along that direction. Finally, for each grid point we store three symbols, one for each direction with the following meaning:

- X: the value on a grid point is larger than the values on the two adjacent grid points along a certain direction.
- N: the value on a grid point is smaller than the values on the adjacent grid points along a certain direction.
- C: the value of the parameter grows along the positive direction.
- D: the value of the parameter decreases along the positive direction.

We call (local) peaks the points that are maxima along all three directions and we call (local) valleys the minima. The number of peaks and valleys of δ and q for each smoothing scale is visible in Tab. C.1 and Tab. C.2. In the same table we also show the number of peaks obtained after imposing different constraints on the δ height. More details on this point will be provided in the next section. We notice that the number of peaks decreases as the smoothing scale increases and that peaks and valleys of δ are present in almost equal quantity, as expected from symmetry considerations.

Fig. 5.1 shows the distribution of δ smoothed on nine different scales. Given a smoothing scale, we compare the distribution of δ on all grid points (black histograms) and the distribution of δ identified as peaks. Whereas the first distribution has zero mean, the mean of the distribution of δ_{pk} is shifted towards larger δ's.

5.1.2 Correlation functions

Whereas in most works (Ludlow and Porciani (2011)) peaks are selected in proximity of protohalos, in what follows we consider peaks identified everywhere. This approach allows us to analyze the peaks distribution within larger regions around the protohalo centers and in principle it permits to study the clustering of peaks as a function of the smoothing scale and the peaks height.

One way to approach this analysis is to calculate the cross correlation function between quantities of interest (i.e. between centers and peaks/valleys). The correlation function has been calculated as follows. Given a protohalo center and peaks on a certain scale, we bin their distances over spherical shells built around the protohalo center. Finally, we divide the number of counts within each shell (N_{counts}) by counts obtained using a random distribution of peaks and centers. For each mass bin and smoothing scale, the random

number of counts within a spherical shell of volume dV is:

$$N_{random} = \frac{dV N_{pk} N_h}{V_u} \tag{5.1}$$

where V_u and dV are respectively the volume of the box and that of the spherical shell in units of $(h^{-1}Mpc)^3$, N_{pk} is the total number of peaks of a certain quantity smoothed on that scale and N_h is the number of protohalos in the chosen mass bin. Therefore, the correlation function can be written in the form:

$$1 + \chi = \frac{N_{counts}}{N_{random}} \tag{5.2}$$

We choose shells to be equally spaced with logarithmic step equal to 0.2 grid points. The wrap necessary to respect the bound conditions allows us to have a maximum distance along each coordinate of half the length of the box side. In order to have a better resolution at small distances and a larger halo statistics we use different simulations when inspecting correlation function for different halo masses. In what follows, we use halos identified at four different redshifts: $z_{id} = 0, 1, 2$ and 4. To make the comparison easier we consider the universal mass parameter ν as defined in 4.1.2.

5.1.3 Correlation function between centers and overdensity peaks

A selection of correlation functions between protohalo centers of mass and overdensity peaks are shown in Fig.5.2-5.5. More correlations for different scales and different halo masses are shown in appendix D. More specifically, each plot shows the correlation function between centers of protohalos within a certain mass bin and peaks smoothed on the characteristic Lagrangian halo scale together with other four smoothing scales: two smaller and two larger of the Lagrangian one. In each page the left plot shows the correlation functions without any restriction on the peak height, whereas the right plot show four panels, corresponding to peaks larger than certain thresholds: $\delta_C(z_{id})$, $2\delta_C(z_{id})$, $3\delta_C(z_{id})$ and $\delta_C(z_{id}) + 0.4q$. The first three thresholds correspond to multiples of the spherical collapse threshold (i.e. they correspond to one, two and three σ of the typical scale forming halos at the identification redshift). The fourth threshold is an approximation of the moving barrier emerging in the excursion sets formalism when considering ellipsoidal collapse (Sheth et al. (2001), Despali et al. (2013), Sheth et al. (2013)). Although our different approach, we use this last threshold to investigate how the shear affects the correlation function between centers and δ peaks. Being q always positive, it always increases the peaks height necessary to a halo for collapse. This point will be investigated deeply in Ch. 6 where we leave the peaks approach and we switch to the analysis of Lagrangian parameters (i.e. parameters smoothed on the Lagrangian scale of halos after centering on the protohalo centers). We notice that:

- Correlation functions tend to zero outside the Lagrangian radius, consistently with theoretical expectations. However, the distance at which this occurs changes with mass.

- When the constraint on the peak heights is low, the correlation functions are mostly positive inside the halo Lagrangian radius (R_L).
- When no constraints on the peak heights are taken into account (left figures), the correlation function at the smallest distances is maximum for smoothing scales equal to the Lagrangian radius (green curve) and immediately smaller (cyan). In the innermost region, the correlation is lowest when δ is smoothed on the largest scale (red curve); on the other hand, in proximity of the Lagrangian radius the lowest correlation function corresponds to the smallest scales. As ν increases, the correlation function in the innermost region grows for scales equal to and lower than the Lagrangian scale.
- When a weak constraint is imposed on the peak heights (top left panel of the right figures), the correlation function in the innermost regions always grows for moderately large scales and the Lagrangian scale tends to stand above the other ones when ν is large enough.
- When strong constraints on the peak height are taken into account (top right and bottom left panels of right figures), only peaks for δ smoothed on the smallest scales survive within the Lagrangian radius. In the framework of peak theory, most of these peaks correspond to halo progenitors formed at higher redshift.
- When considering the effects of shear (bottom right panel of right figures) at small distances, the correlation function for peaks corresponding to scales equal and smaller of the Lagrangian scale (blue, cyan and green curve) is high, whereas the same correlation for peaks on larger scales is not defined to a lack of peaks. So, this threshold seems to better select peaks involved in the halo formation.

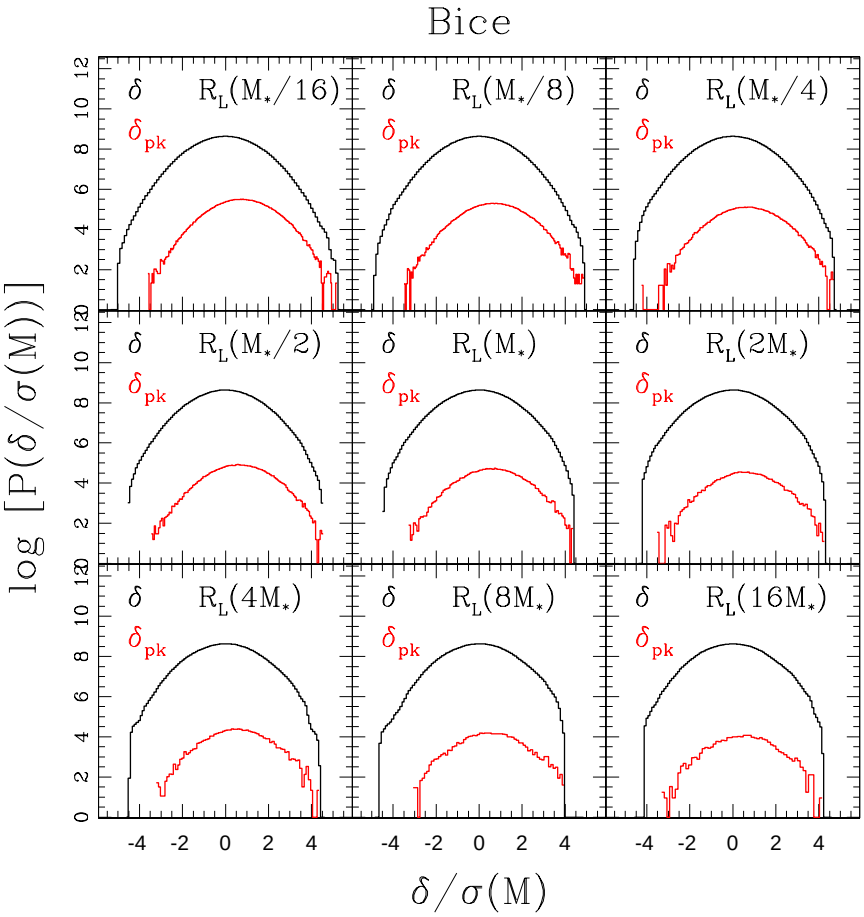

Figure 5.1: Distribution of δ smoothed with nine different scales on all grid points (black) and on grid points identified as peaks (red). These distributions have been obtained with Bice simulation. The smoothing scales corresponds to the Lagrangian scales of halos identified at $z_{id} = 0$ with masses as in parentheses.

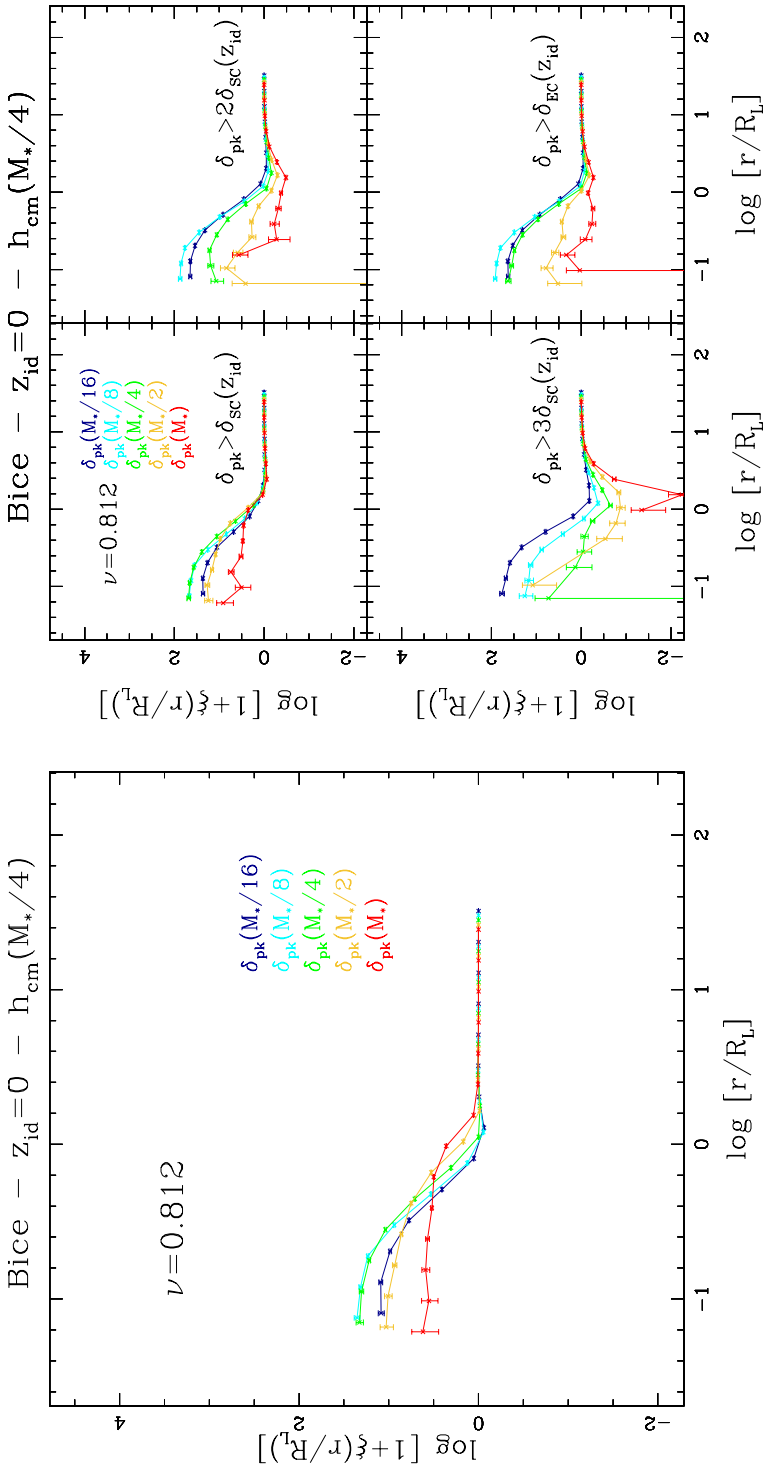

Figure 5.2: Correlation functions between centers of mass of protohalos with $\nu = 0.812$ and δ peaks on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 0$. Halos have been identified with Bice simulation. Left figure: all peaks; right figure: peaks larger than four different thresholds.

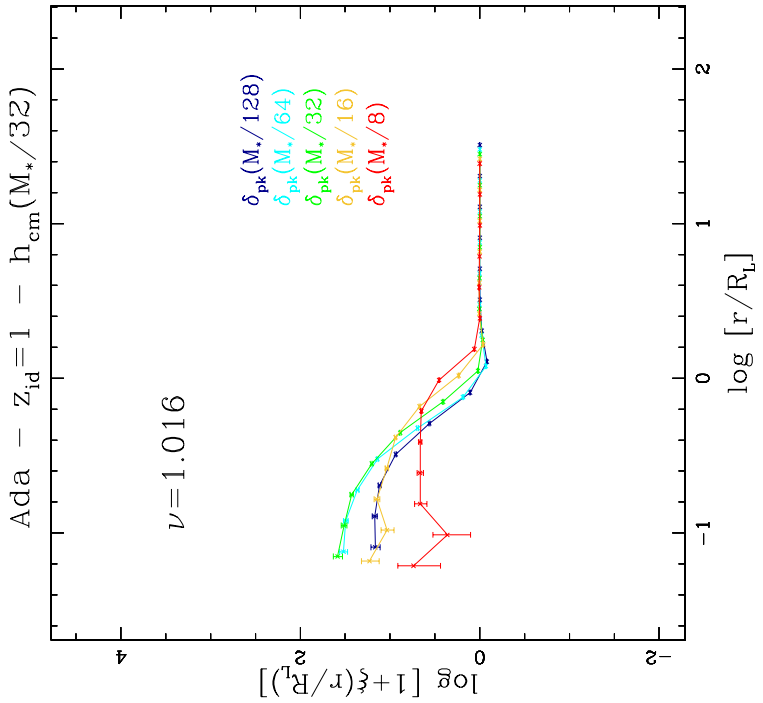

Figure 5.3: Correlation functions between centers of mass of protohalos with $\nu = 1.016$ and δ peaks on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 1$. Halos have been identified with Ada simulation. Left figure: all peaks; right figure: peaks larger than four different thresholds.

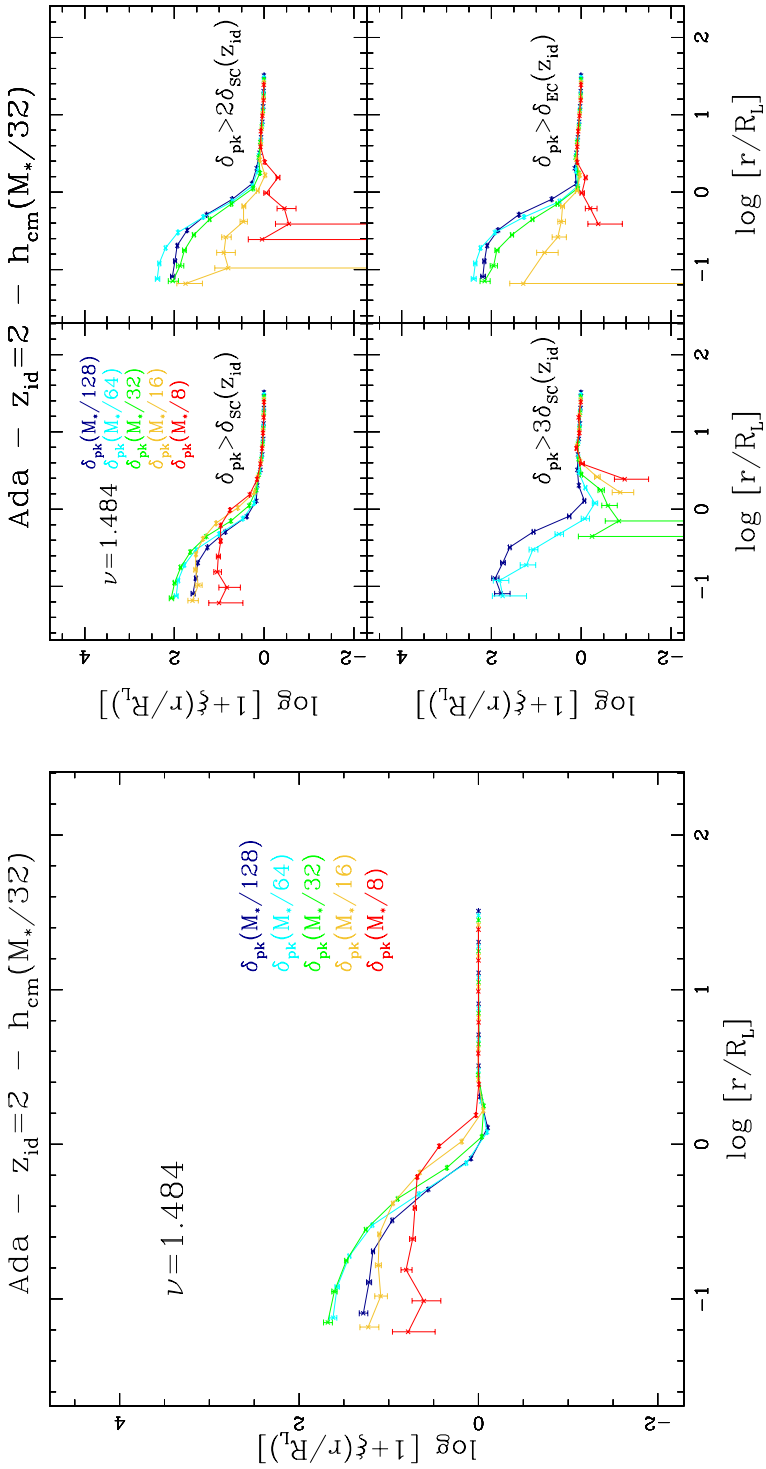

Figure 5.4: Correlation functions between centers of mass of protohalos with $\nu = 1.484$ and δ peaks on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 2$. Halos have been identified with Ada simulation. Left figure: all peaks; right figure: peaks larger than four different thresholds.

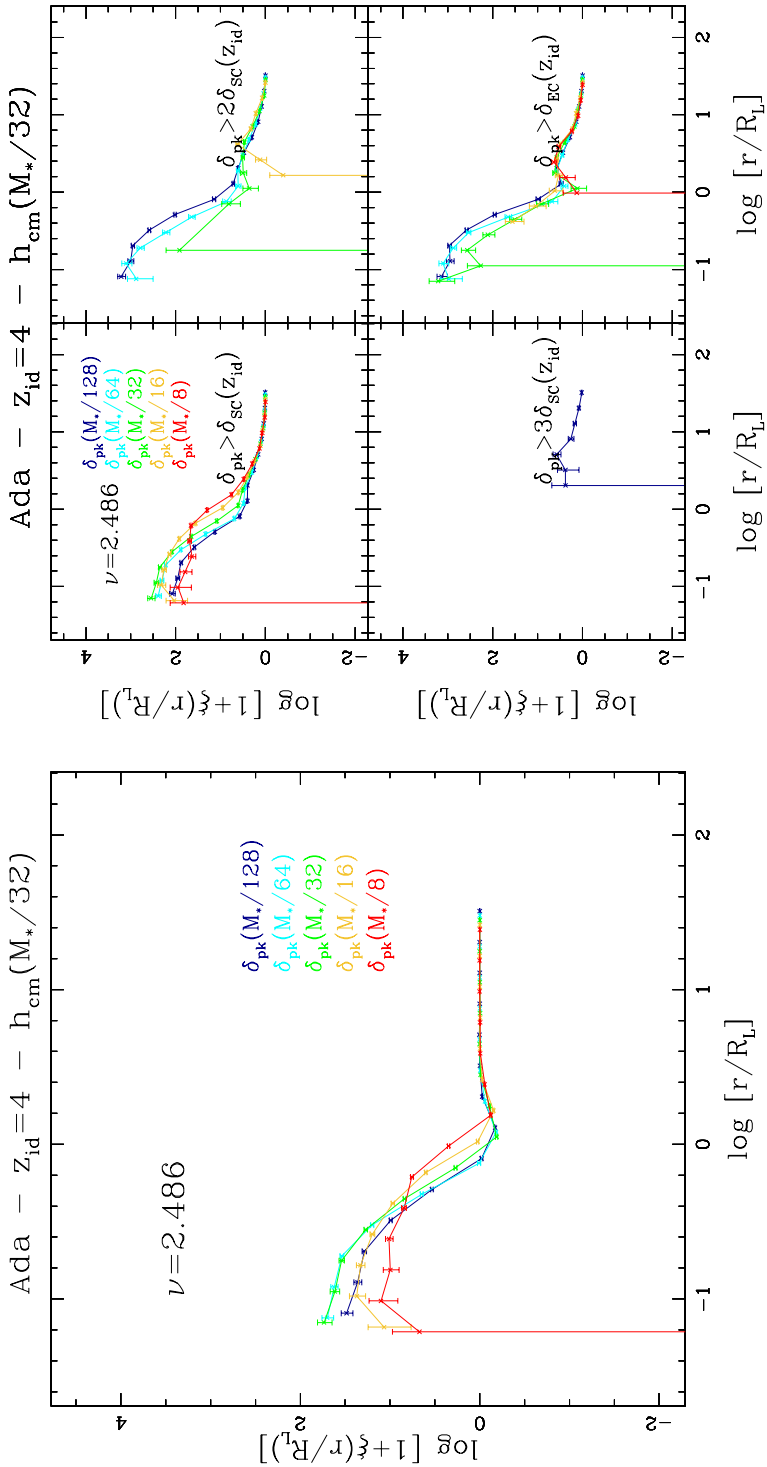

Figure 5.5: Correlation functions between centers of mass of protohalos with $\nu = 2.486$ and δ peaks on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 4$. Halos have been identified with Ada simulation. Left figure: all peaks; right figure: peaks larger than four different thresholds.

We summarize the results at small and large distances in Fig. 5.6. In left figure we choose a distance equal to about the 13% of the Lagrangian radius and we show the value associated to the correlation function at that distance. The right figure has been built in the same way, but taking into account larger distances (about $1.3R_L$). The nine panels correspond to different values of ν and show results obtained employing different thresholds (more details in the figure caption). In general, the correlation function at small distances (left panel) with peaks on scales equal ($R(M_h)$, green points) to and immediately smaller (cyan) than the Lagrangian scale is always largest when considering no constraints on the peaks height (case a) and when considering a threshold equal to $\delta_{SC}(z)$ (case b). On the other hand, the correlation with peaks on largest scales is always the lowest and it quickly disappears when increasing the threshold height. Indeed, if peaks on the largest scales for high δ survive in the innermost regions, they would be identified as halos of larger mass. When considering a larger threshold (case c) the correlation function with peaks on the Lagrangian scale (green) is positive only for smallest halos. We can understand that point considering that these halos have assembled a large fraction of the mass at high redshift, when fluctuations were smaller. Moreover, small mass halos needs a larger overdensity to collapse because of tidal effects The last threshold (case e) largely removes the correlation with peaks on largest scales and, at the same time, keeps high the correlation with peaks on scales equal to and larger than the Lagrangian one. So, when taking into account a combination of δ and shear in the definition of the threshold for peaks, we are able to select peaks related to the halo formation history. However, this combination should be more accurately calibrated. This will be the purpose of a future analysis. The right figure of 5.6 shows that the correlation function at high distances ($1.3R_L$) is low or negative at all scales when considering high thresholds (case c and d). Indeed, the presence of high peaks in this region would affect the formation of the halo. Higher correlations are present only for peaks on smallest scales. When considering a lower threshold (case b) we see a positive correlation for peaks on largest scales. They are probably the seeds of future formation of higher mass halos. A future step of this analysis will focus on the study of the peak clustering, trying to extend this approach on a object-by-object analysis.

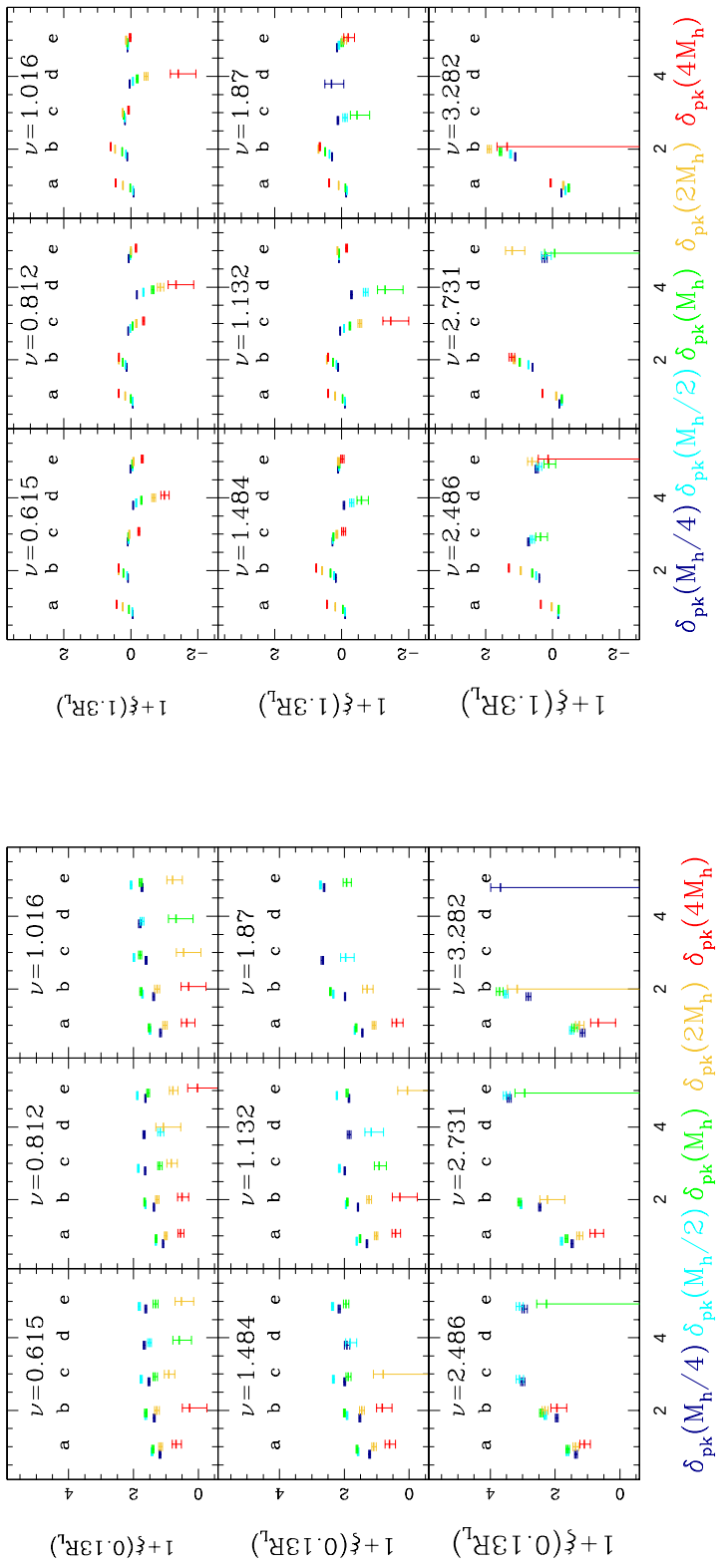

Figure 5.6: Value of the correlation function between protohalo centers and δ peaks at a distance of $0.13R_L$ (left) and $1.3R_L$ (right). The notation of letter is the following: a) all peaks; b) $\delta > \delta_{SC}$; c) $\delta > 2\delta_{SC}$; d) $\delta > 3\delta_{SC}$; e) $\delta > \delta_{SC} + 0.4q$

5.1.4 Correlation function between centers and peaks and valleys of q

In order to investigate how the shear distribution affects the halo formation, we calculate the correlation function between protohalo centers and peaks and valleys of the q distribution. As we discussed earlier, the shear tends to increase the threshold for collapse. This effect is expected to be higher at smaller masses i.e. small halos need a higher density to keep their mass together against the tidal forces. Whereas in the previous section we used q to raise the threshold of collapse, here we are mainly interested on its gradient i.e. we want to investigate how the fact that q is a peak or a valley affects the halo formation.

Fig. 5.7-5.10 show the correlation function of protohalo centers of mass with peaks (q_{pk}) and valleys (q_{vl}) of q. Each plot shows four panels with a notation similar to previous plots. However, in this case the collapse threshold does not relate to δ peaks but to peaks and valleys of the q distribution. In practice, we consider all peaks and valleys of q together with peaks and valleys of q corresponding to a value of δ larger than a certain threshold. We notice that:

- When considering peaks of q (right figures) with $\delta > \delta_{SC}(z_{id})$ (top left panels), we find the highest correlation function at small distances when peaks of q are smoothed on scales larger than the Lagrangian one. On the contrary, when considering valleys of q, the highest correlation is for peaks on smallest scales.

- When considering both peaks and valleys of q with high thresholds (top right and bottom left panels), we find that the highest correlation function at small distances tends to move toward smaller scales.

- When considering high thresholds and distances around the Lagrangian radius, nearly all scales show a negative correlation.

- If we take into account bottom right panels ($\delta > \delta_{SC}(z_{id})$, the correlation function for peaks and valleys of q tends to be similar and, at least at small ν, select scales equal to and smaller than the Lagrangian scale.

In general, when considering smoothing scales larger than the Lagrangian scale, only peaks of q are allowed to be close to the protohalo centers. However, when the threshold of collapse increases, only the smallest scales survive in proximity of centers. When considering this result together with correlations showed in the bottom left panels, we can say that the value of q tends to be more important than the fact that q is a peak or a valley.

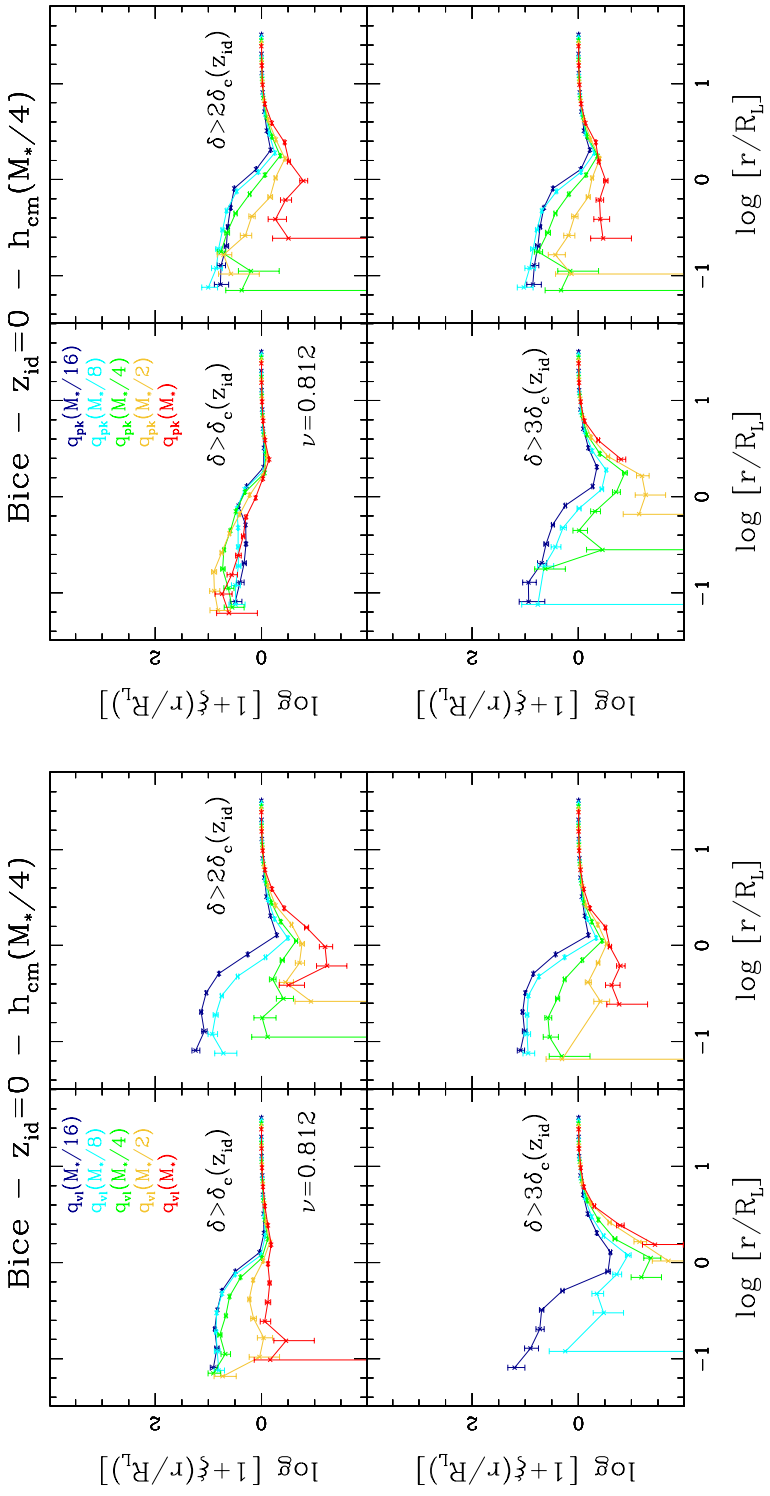

Figure 5.7: Correlation functions between centers of mass of protohalos with $\nu = 0.812$ and q peaks (right) and valleys (left) on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 0$. Halos have been identified with Bice simulation. We consider for different threshold for the value of δ.

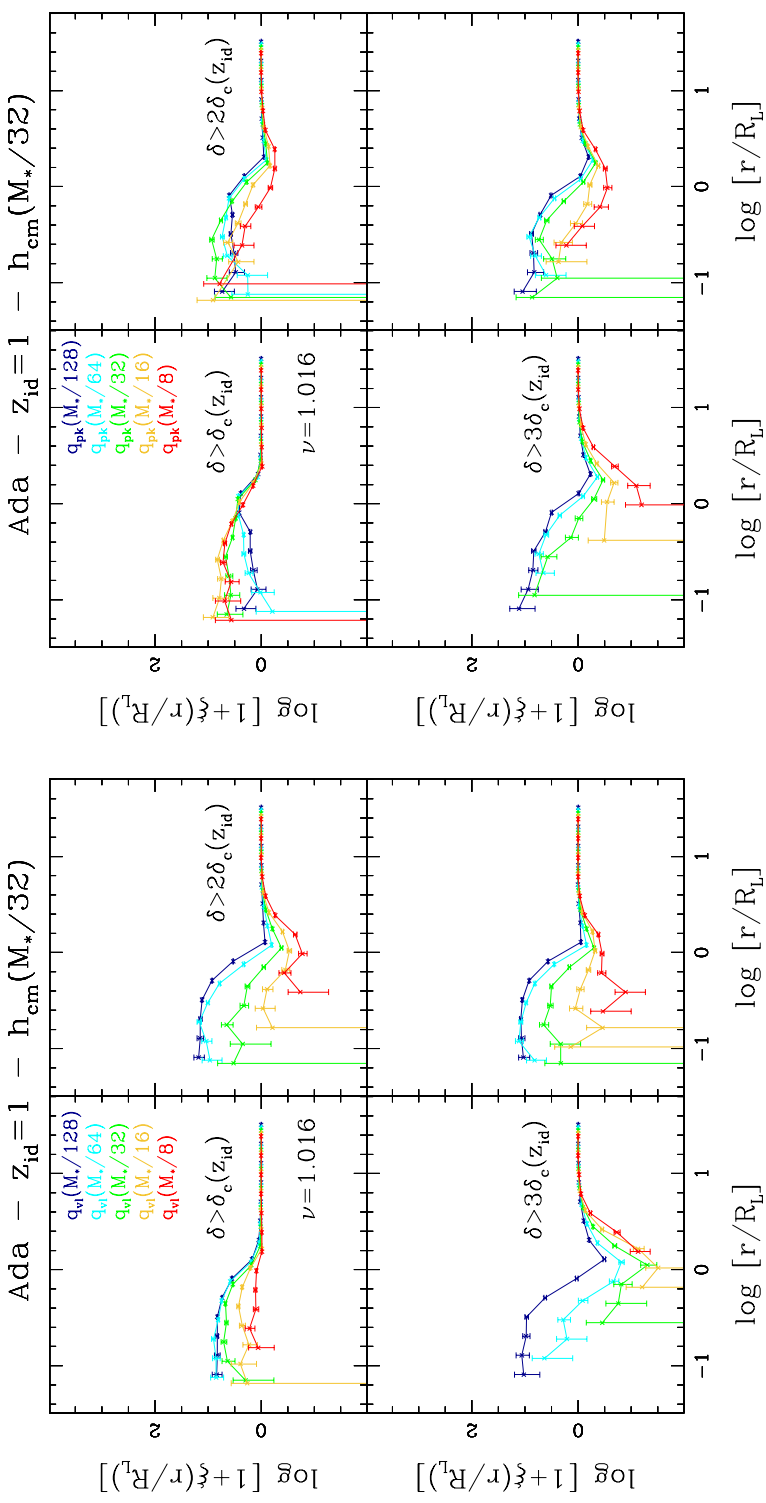

Figure 5.8: Correlation functions between centers of mass of protohalos with $\nu = 1.016$ and q peaks (right) and valleys (left) on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 1$. Halos have been identified with Ada simulation. We consider for different threshold for the value of δ.

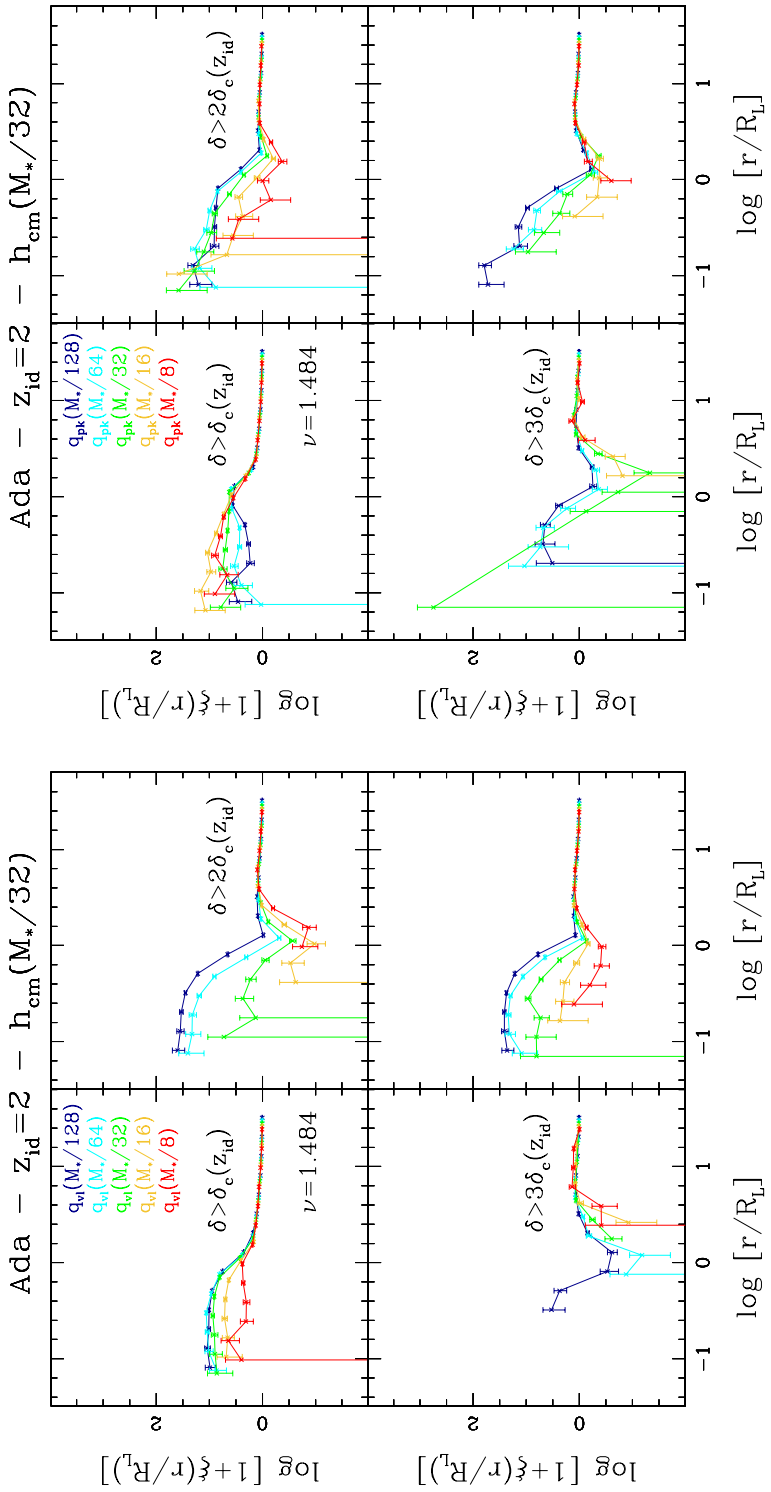

Figure 5.9: Correlation functions between centers of mass of protohalos with $\nu = 1.484$ and q peaks (right) and valleys (left) on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 2$. Halos have been identified with Ada simulation. We consider for different threshold for the value of δ.

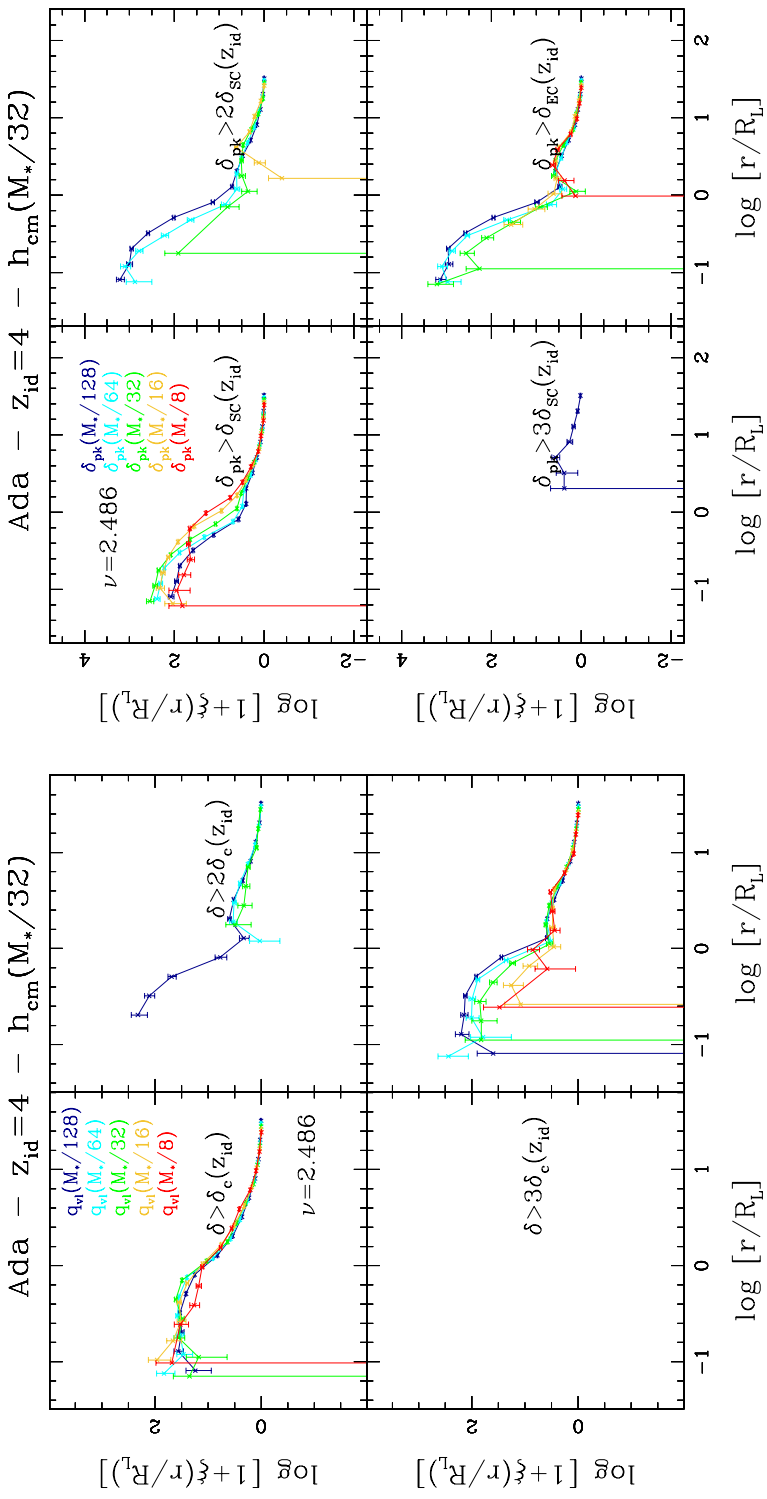

Figure 5.10: Correlation functions between centers of mass of protohalos with $\nu = 2.486$ and q peaks (right) and valleys (left) on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 4$. Halos have been identified with Ada simulation. We consider for different threshold for the value of δ.

5.2 An alternative way to locate halo seeds

In this Section we sketch an alternative to the peak model for halo formation. Sec. 5.1 of Chapter 5 shows that there is a positive correlation function at small distances between centers of protohalos of a certain mass and peaks for the Lagrangian overdensity field smoothed on the Lagrangian scale. However, this statistical characterization is far to be perfect and in a object-by-object analysis the correspondence between peaks and centers breaks down, especially for small halos. This is partially due to the fact that shear highly affects the formation of small halos and so, the complex formation history of some objects can lead to a lack of peaks within the Lagrangian radius (see Ludlow and Porciani (2011) for a object-by-object analysis). In Section 5.1 we investigated the correlation function using peaks and valleys of the Lagrangian shear q. A possible alternative approach concerns the study of peaks in some other fields (e.g., some combination of the overdensity and shear field); this will be part of a future analysis. In what follows we present another way to inspect the halo seeds; it needs to set the peak approach aside and look for a better guiding principle.

5.2.1 Dipole as a characterization of the halo formation

In perturbation theory emerges a dipole term that is missing in spherical and ellipsoidal collapse. This is due to the fact that these models explicitly describe collapse around the center of mass of protohalos, whereas it could be more appropriate to assume that protohalos form from regions in which the initial dipole vanishes. In practice, we can focus on points where velocity field converges. Given a point acting as a gravitational center, we can describe the halo formation as a collapse of concentric shells, where each shell is by definition constituted by points falling simultaneously on the gravitational center. In general, given a sphere, the dipole depends on the position of the mass center compared with the position of the geometric center. When dipole is not zero, a spherical surface falls on the mass center with a non isotropic velocity; therefore, the position of the sphere can be shifted until the dipole is null. Ultimately, when neglecting the effect of higher multipoles, a point with a null dipole is a point where the collapse of the sphere is isotropic.

The analysis of the halo seeds employing this approach has not been yet investigated and will be part of a work close to be published. We provide a numerical contribution in order to check appropriately theoretical predictions.

In our work, dipole has been calculated in the following way: we center on a certain point at the initial conditions and we consider a sphere with a certain radius. We consider all particles within that sphere and we calculate their center of mass. At a later stage, we calculate the distance between the center of mass and the geometric center of the sphere and we use this distance as a measure of the dipole.

Now we need a method to test the importance of dipole in the characterization of the halo seeds. Therefore, we consider protohalos of a certain mass and we check if dipoles calculated in proximity of the protohalo center of mass heads toward the center. Indeed,

the condition on the eigenvalues of the dipole gradient assures that shifting a sphere in the direction of dipole D, the length of D decreases. We can express the same concept saying that points with D=0 are stable. Given a protohalo, we consider a coordinate system centered on the protohalo mass center with axes parallel to the sides of the simulation box. After that we move of a certain quantity along the positive and negative direction of each of the three axes, identifying six points. We center on each point and we calculate the center of mass within a sphere enclosing a mass equal to the protohalo mass. We can now calculate the components of the dipole along the three axis in the form:

$$\epsilon_{r_i} = \frac{\bar{r}_{i;P} - \bar{r}_{i;CM}}{\Delta r_i} \tag{5.3}$$

where $\bar{r}_{i;P}$ are the coordinates of the shifted points, $\bar{r}_{i;CM}$ are the coordinates of the protohalo center of mass and Δr_i is the distance between those points along each coordinate. The left plot of Fig. 5.11 shows results for protohalos of mass $M_*/8$ and a shift along the three axes equal to $R_L/10$, where R_L is the protohalo Lagrangian radius. Given a shift along one coordinate, the dipole heads towards the mass center along that coordinate but remains unchanged on the other two. The right plot of Fig. 5.11 has been built on the same way but dipole has been measured within spheres enclosing a larger mass (i.e. M_*). Although the dipole along the direction of the displacement points again towards the protohalo mass center, its value is much smaller. This is plausible because of large scale mass outside the protohalo region.

At a later stage we want to find the points at which the dipole is minimum and compare those points with the protohalo mass center. First of all we select a spherical region of given radius around the protohalo center and we consider a three dimensional grid. We center on each grid point and we consider a sphere of radius equal to the Lagrangian radius of the protohalo. We then calculate the value of dipole for each sphere and we select the center of sphere at which the dipole is minimum. We decide to consider spherical regions of radius equal to $0.4R_L$ in order to avoid the possibility of identifying points not directly related to the formation of the halo. Top left panels of left and right plots in Fig.5.12 show two examples of protohalos obtained tracing back to the initial conditions the particles of halos with mass $4M_*$ identified at $z_{id} = 0$. We show the protohalo particles (black dots) projected on three planes together with the closest overdensity peak (red square), the minimum dipole (blue circle) identified within a region of $R = 0.4R_L$ and the position of centers of mass and moving centers at the identification redshift $z_{id} = 0$ and at the initial conditions. When the halo has a regular shape (left figure), the closest peak of δ and the minimum dipole are both close to the protohalo center; on the other hand, when the protohalo shape is not regular (right figure), both points tend to be far from the protohalo center. In the other three panels we consider the particles of the main progenitors at different redshifts. Let's consider the most irregular protohalo (right plots). When the shape of the main progenitor protohalo becomes more regular, both the dipole and the peak for δ smoothed on the scale of the protohalo mass tend to be close to the protohalo center of mass.

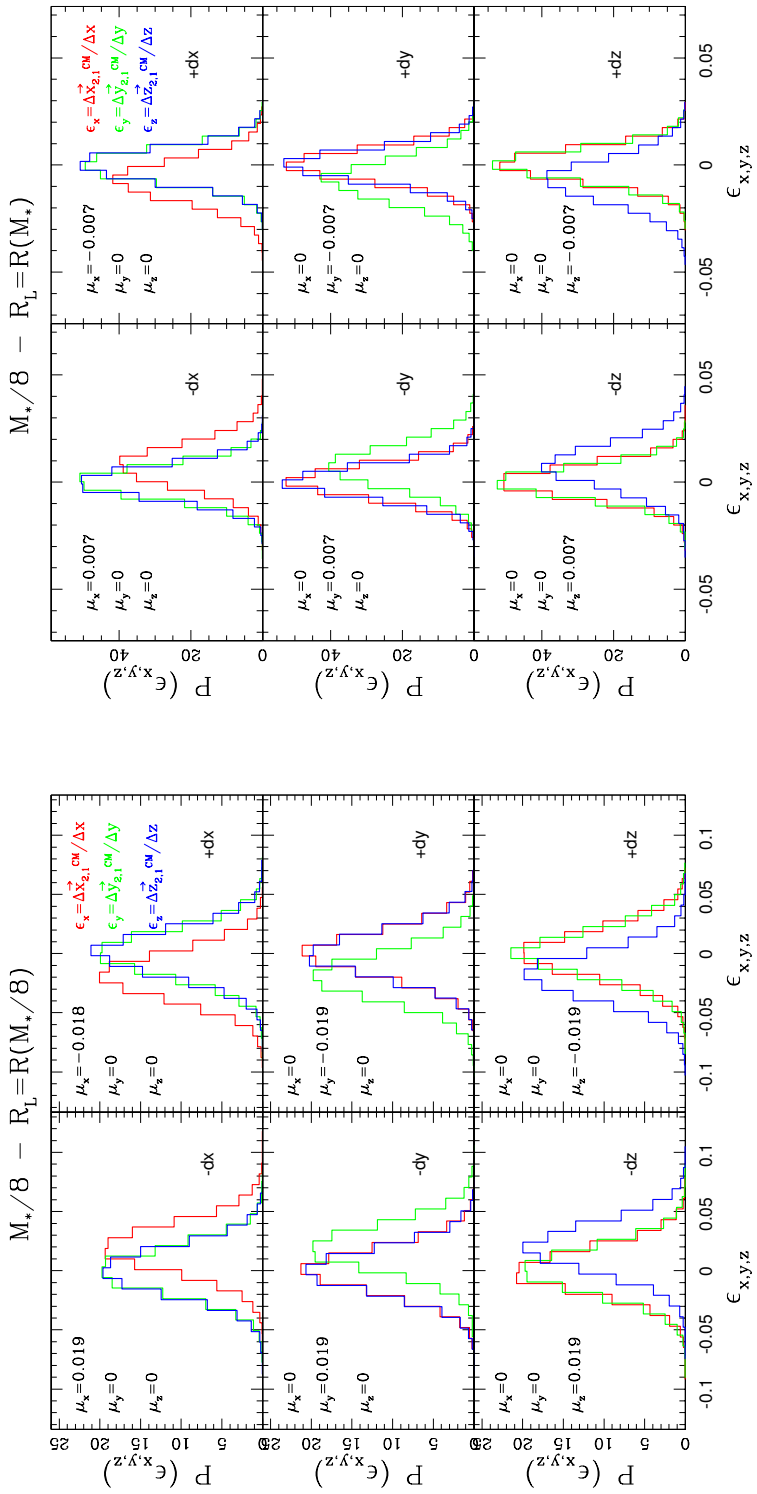

Figure 5.11: Dipole of spheres moved to a distance of $R_L/10$ from the protohalo mass center along negative (left) and positive (right) directions of the three coordinates.

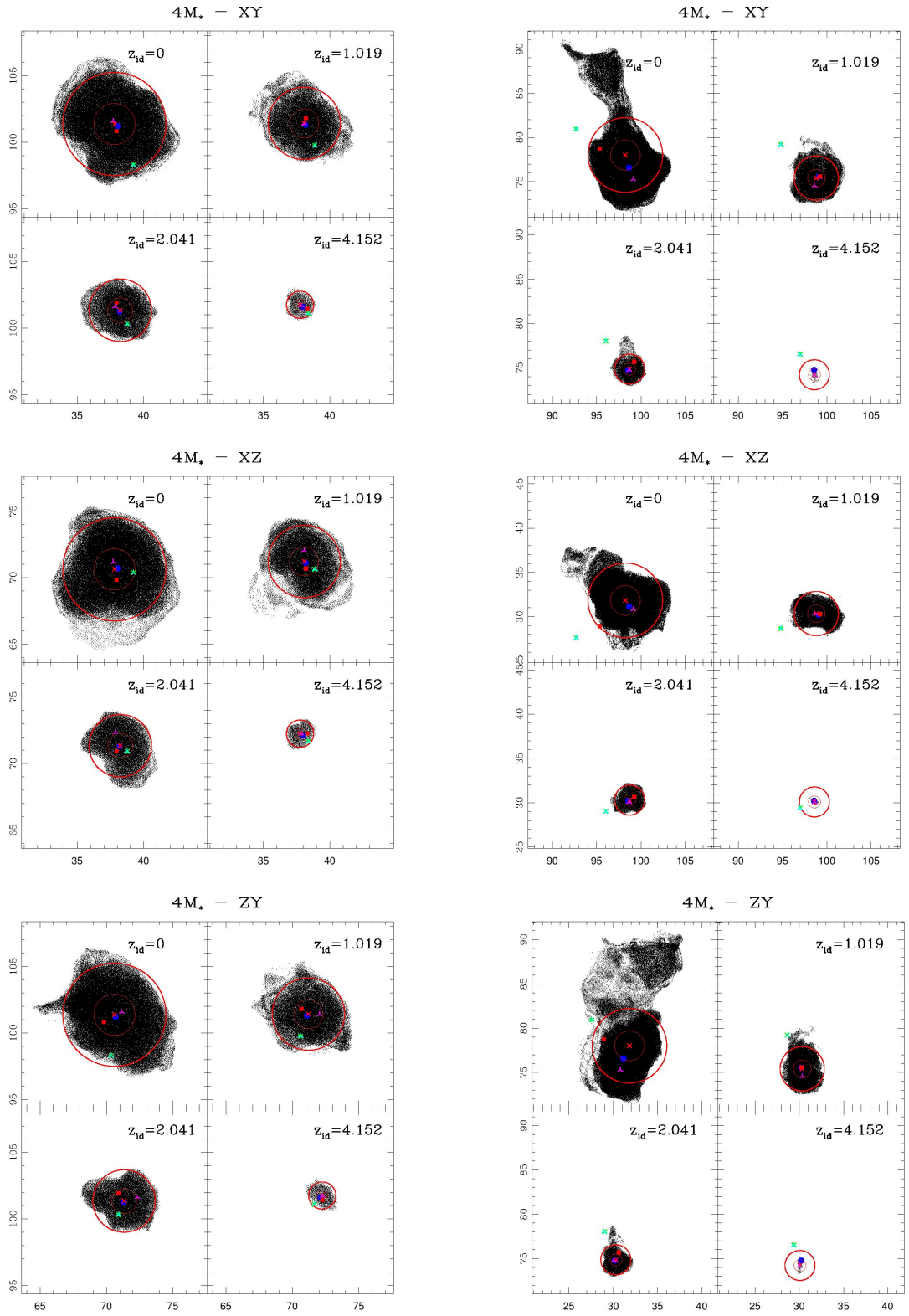

Figure 5.12: Distribution of protohalo particles of two halos (left and right plots) identified at $z_{id} = 0$ (top left panels) together with particles of halo main progenitors identified at three higher redhsifts: $z_{id} = 1$ (top right); $z_{id} = 2$ (bottom left); $z_{id} = 4$ (bottom right). The innermost circumference enclose the region where the minimum dipole has been selected; whereas, the outermost circumference has a radius equal to the Lagrangian radius. Red square: closest δ peak; blue circle: minimum dipole; red (magenta) cross: mass (moving) center at the initial conditions; light (dark) green cross: mass (moving) center at the identification redshift.

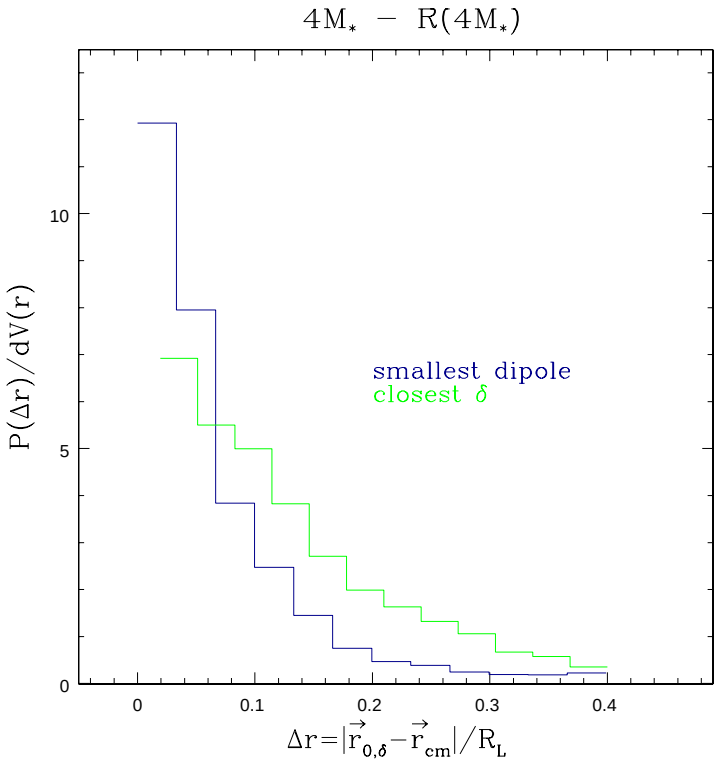

Figure 5.13: Distribution of distances between the centers of protohalos with mass equal to $4M_*$ and both the closest δ peaks (green curve) and the minimum of dipole (blue curve).

So, we have good hints that dipole characterizes well the main bunch of particles forming halos. At a later stage we would understand whether or not the dipole approach works better than the peak one. This part is far from being completed. However we have good hints that the dipole characterization is more suited to characterized halo formation. As an example, we consider Fig.5.13 where we show the distribution of distances between the protohalo centers and the closest peaks for δ smoothed on the Lagrangian halo scale together with distances of the minimum dipole. In both cases we consider only quantities identified within a spherical region around the protohalo mass center of radius equal to $0.4R_L$. We show how the minimum of dipole seems to characterize better the halo center of mass. However, whereas δ peaks can be in principle identified everywhere, the minimum dipole has to be compared within a specific region. So, although this positive confirmation, in future we will investigate a more in-depth analysis.

5.3 Summary

In this chapter we investigated the correlation function between protohalo centers and peaks and valleys of the δ and q distribution, when smoothed on different scales. We defined peaks (valleys) of a certain quantity the grid points larger (smaller) than the six adjacent grid points. We firstly inspected the correlation function of protohalo centers ans peaks of the δ distribution, after imposing different thresholds on the height of peaks. When no constraints on the peak heights are taken into account, the correlation function at the smallest distances is maximum for smoothing scales equal to and immediately lower than the Lagrangian scale, whereas the correlation is lowest when δ is smoothed on the largest scale. As ν and the threshold increase, the correlation function in the innermost region grows for scales equal to and lower than the Lagrangian scale. In general, when imposing a weak constraint on the peak heights, the correlation function with peaks of δ smoothed on the Lagrangian scale tends to stand above the other ones when ν is large enough. Moreover, we showed that peaks associated with the halo patches are well selected when considering the contribution of shear (derived in the excursion set framework) as a constraint on the peak height. At a later stage we found that protohalo centers and valleys of q smoothed on scales equal or smaller than the Lagrangian one show a high correlation function when no other constraints is taken into account; on the other hand the correlation function with peaks of q shows an opposite behaviour. However, we showed that considerations on the value of q are more important than the fact that q is a valley or a peak.

In the second section we investigated an alternative way to characterize the formation of dark matter halos. Indeed, in perturbation theory emerges a dipole term that is missing in spherical and ellipsoidal collapse model. We investigated the module and direction of this quantity in proximity of protohalo mass centers and we show that in average the dipole heads towards the mass center. Moreover, we show that dipole seems to characterize the protohalo center of mass better than *delta* peaks.

CHAPTER 6

The importance of shear in the halo formation

In chapter 5 we investigated the correlation function between protohalo mass centers and peaks and valleys in the δ and q distribution. Among other things, we showed how a halo tends to form close to δ peaks filtered on scales equal to (and immediately lower than) the Lagrangian scale. We highlighted also how the null gradient (i.e. peaks and valleys) of q affects the halo formation: when weak contraints on the peak height are taken into account, the cross correlation of protohalo centers with local minima of the q distribution is higher than the correlation with local maxima when considering scales equal to and lower than the Lagrangian scale. However, this effect seems to depend on the value of q and not on the fact that q is a valley (i.e. a valley of q is generally lower than a peak and so the effect of shear is also lower).

In this chapter we leave the peak approach in order to investigate the correlation between Lagrangian parameters δ_L and q_L (i.e. δ and q smoothed on the halo scale after centering on the protohalo center) and the formation redshift z_f. In the first section we describe the distribution of different Lagrangian parameters and we compare our results with theoretical expectations and previous works. At a later stage, we investigate correlations. It is well known that the external shear increases the collapse threshold, making the collapse harder to happen. Therefore, given a halo identified at a certain redshift, we expect that δ_L extrapolated linearly to z_{id} is larger when q_L is larger. Some discussion of why this correlation matters for assembly bias in Castorina and Sheth (2013). However, our is the first study of Lagrangian shear and delta at different redshifts. Recent works have also highlighted that δ_L correlates also with the formation time: halos that form earlier (i.e. a patch that reaches the collapse threshold at earlier times) has a larger δ_L at the identification time compared to halos forming in more recent times. We inspect the above quoted correlations for different identification redshifts and we try to understand if

there is a direct correlation between δ_L and the parameters q_L and z_{50} or, on the contrary, if the correlation is due to their mutual dependence.

6.1 Protohalos and Lagrangian parameters

One of the aim of this work is the study of the distribution of protohalo Lagrangian properties. We define protohalos as follows. Let's consider a halo identified at z_{id} within a bin of characteristic mass M. The protohalo is obtained by tracing the halo particles back to $z = 99$, when the density field can be considered to be linear. We then calculate the elements of the deformation tensor in correspondence of protohalo patches. The deformation tensor at each point is defined as the second derivatives of the gravitational potential Φ. In the Zel'dovich approximation this corresponds to the first derivatives of the initial displacements,

$$\zeta_{ij} = -\frac{\partial^2 \Phi}{\partial x_i \partial x_j} = -\frac{\partial \psi}{\partial x} \tag{6.1}$$

where displacements have been obtained by interpolating the particles on the grid with a tsc interpolation. We calculate the deformation tensor at each grid point by convolving the field with a Top-Hat filter via multiplication in the Fourier space. We choose smoothing scales corresponding to the characteristic bin masses M shown in Tab. 4.1. Since $\sigma^2(M) << 1$ at $z = 99$, the smoothing scales can be calculated as

$$R_L = \sqrt[3]{\frac{3M}{4\pi\rho_0}} \tag{6.2}$$

where ρ_0 is the initial density. Tab. 6.1 shows R_L in units of $h^{-1}Mpc$ together with the number of grid points spanned by these scales in the different simulations.

We then consider the center of mass of protohalos. Whereas the deformation tensor has been defined in correspondence of grid points, the protohalo centers can be found everywhere. We then interpolate the deformation tensor elements at the exact position of the centers of mass using again a tsc interpolation. In this way we avoid resolution problems arising when considering, for example, the closest grid point.

We now calculate the eigenvalues of the deformation tensor smoothed on the Lagrangian scales ($\lambda_{1;L} \geq \lambda_{2;L} \geq \lambda_{3;L}$) and, from these, the parameters of the ellipsoidal collapse model. We refer to these parameters as the Lagrangian parameters. They are: the trace of the deformation tensor,

$$\delta_L = \lambda_{1;L} + \lambda_{2;L} + \lambda_{3;L}, \tag{6.3}$$

the ellipticity,

$$e_L = \frac{\lambda_{1;L} - \lambda_{3;L}}{2\delta_L}, \tag{6.4}$$

Smoothing scales										
	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$
$M(10^{12}h^{-1}M_\odot)$	0.39	0.78	0.156	0.312	0.625	1.25	2.45	4.9	9.8	19.6
$R[h^{-1}Mpc]$	0.48	0.6	0.76	0.95	1.2	1.51	1.90	2.40	3.02	3.80
$R_A[gp]$	7.78	9.81	12.36	15.57	19.62	24.72	31.14	39.24	49.43	
$R_B[gp]$				7.78	9.81	12.36	15.57	19.62	24.72	31.14
$R_C[gp]$							7.78	9.81	12.36	15.57
$R_D[gp]$										7.78

	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
$M(10^{12}h^{-1}M_\odot)$	39.2	78.4	156.8	313.6	627.2
R [Mpc]	4.79	6.04	7.6	9.6	12.08
$R_A[gp]$					
$R_B[gp]$	49.43				
$R_C[gp]$	19.62	24.72	31.14	39.24	49.43
$R_D[gp]$	9.81	12.36	15.57	19.62	24.72

Table 6.1: Lagrangian scales in units of Mpc and corresponding number of grid points for the different simulations. Capitol letters specify the different simulations: A=Ada; B=Bice; C=Cloe; D=Dora.

and the prolateness,

$$p_L = \frac{\lambda_{1;L} + \lambda_{3;L} - 2\lambda_{2;L}}{2\delta_L}. \tag{6.5}$$

When $\delta_L > 0$, we have $e_L \geq 0$ and $|p_L| \leq 0$.

6.2 Statistics of Lagrangian parameters

In this Section we firstly analyze the distribution of the Lagrangian eigenvalues of the deformation tensor together with the Lagrangian parameters of the ellipsoidal collapse model. At a later stage we define the shear parameter q and we show how it affects the collapse threshold. Recent works (Robertson et al. (2009), Despali et al. (2013), Sheth et al. (2013)) investigated some of these distributions but considered only halos identified at $z = 0$. Here we extend this analysis by considering also larger identification redshifts. We can therefore inspect a wider range of ν and compare results obtained from different identification times.

6.2.1 Lagrangian eigenvalues

Firstly, we take into account the eigenvalues of the deformation tensor. Fig. 6.1 shows the distribution of $\lambda_{i;L}$, rescaled to the halo identification redshift z_{id}, as a function of (the logarithm of) ν. We notice that almost all halos have at least two positive eigenvalues. On average, 39.6% of patches have $\lambda_{1,2;L} > 0$ and $\lambda_{3;L} < 0$, whereas the 60.2% have $\lambda_{1,2,3;L} > 0$; less than 0.2% have only one positive eigenvalue. We compare these fractions

at $z_{id} = 0$ in the mass range of Despali et al. (2013) using only Bice simulations (in order to avoid the bias due to the different ν at which each simulation comes into play) and we find results very similar to theirs (about 30% of positive triplets and 70% of combinations with only one negative eigenvalue). The fraction of triplets with three (blue) and two (magenta) positive eigenvalues as a function of ν are shown in Fig. 6.2, where we plot separately the different identification redshifts. We notice that the number of halos with three positive eigenvalues grows as ν increases i.e. the contraction along the direction of λ_3 of small mass halos is slowed down more than for massive ones. When ν is slightly higher than 1 ($log\nu \simeq 0.02 - 0.04$), the number of protohalos with three and two positive eigenvalues is the same. Moreover, when the identification redshift gets higher, the fraction with three positive eigenvalues is generally smaller, especially at low ν.

6.2.2 Lagrangian distribution of the EC parameters

We now consider the distribution of the parameters which determine the evolution in the EC model.

In Fig. 6.3 we compare δ_L (rescaled to z_{id}) with the predictions of the spherical collapse model, as a function of ν. We notice that δ_L is always larger than the SC threshold, except for the highest ν. Moreover, δ_L decreases as ν grows. This is consistent with the predictions of the ellipsoidal collapse model, according to which low mass halos needs a larger overdensity to oppose the tidal effects. An estimate of the collapse barrier in the framework of the Bond and Myers (1996) ellipsoidal collapse model was provided in Sheth et al. (2001). They derived the dependence of the collapse threshold on ellipticity and prolateness of the shear field in the form:

$$\frac{\delta_{EC}(e,p)}{\delta_{SC}} = 1 + \beta\Big[5(e^2 \pm p^2)\frac{\delta^2_{EC}(e,p)}{\delta^2_{SC}}\Big] \tag{6.6}$$

where δ, e and p can vary from one position to another and the parameters β and γ are estimated by fitting on the dynamical model. Recent values are $\beta = 0.412$ and $\gamma = 0.618$ (Desjacques, 2008). Top plot of Fig. 6.3 shows a comparison with the ellipsoidal collapse barrier (cyan curve) using a rough estimates of $\langle\delta e\rangle$ and $\langle\delta p\rangle$ (respectively $\sigma/\sqrt{5}$ and 0). These estimates have been calculated by Sheth et al. (2001) by using results of Doroshkevich (1970) and by averaging the distribution of e and p over all position of the Gaussian field. Therefore, the barrier can be rewritten as

$$\frac{\delta}{\delta_{SC}(z)} = 1 + \beta\nu^{-2\gamma}. \tag{6.7}$$

We can move the dependence on redshift from the spherical collapse threshold to the fluctuation field, and write:

$$\frac{\delta(z)}{\delta_{SC}(z=0)} = 1 + 0.25\sigma(z)^{1.23}. \tag{6.8}$$

When comparing our results with the EC threshold, our values are sistematically lower than predictions at both small ν and high identification redshifts. The discrepancies at low masses partly depends on the fact that δe diverges from rough theoretical predictions in this mass range (top plot of Fig. 6.4). Fig.6.3 shows also a comparison with the "Sheth et al." barrier (green curve)

$$B_{SMT}(\sigma, z) = \sqrt{a}\delta_{SC}(z)\Big[1 + \beta\Big(\frac{\sigma^2}{a\delta_{SC}^2(z)}\Big)^{\gamma}\Big], \tag{6.9}$$

where $\beta = 0.5$, $\gamma = 0.6$ and the barrier converges to a value $\sqrt{a_{ST}}\delta_{SC} \simeq 0.84\delta_{SC}$. This formula has been obtained by modifying the Ellipsoidal Collapse barrier in order to account for the halo abundances in the GIF simulations (Sheth et al. (2001)). We find that our distribution is always included between these two barriers at large ν. When dividing this distribution by $\sigma(M)$ we remove most of the mass dependence (bottom plot of Fig.6.3). At $z_{id} = 0$, we find $\delta/\delta_{SC} - 1 \simeq 0.27\sigma$ with an rms scatter of 0.19σ. A comparison with previous works shows a good agreement with results of Robertson et al. (2009) in both mean ($0.48/\delta_{SC} = 0.28$) and rms ($0.3/\delta_{SC} = 0.18$), but values higher than those of Despali et al. (2013) (mean of 0.2 and rms of 0.12). These discrepancies are probably due to the different methods employed in the calculation of the Lagrangian parameters. In fact, Despali et al. (2013) averaged over the protohalo shapes, whereas Robertson et al. (2009) and our work assume that protohalos are spherical. Moreover, Despali et al. (2013) used EO derived quantities, whereas we base our work on the SO identification algorithm.

Fig. 6.5 shows skewness γ_1 and kurtosis γ_2 of the distribution of the Lagrangian density contrast δ_L, as a function of ν. We show results for both all (TOP) and relaxed (BOTTOM) halos, with different colors meaning different halo identification redshifts. We notice that, for both halo selections, skewness is nearly mass invariant, whereas kurtosis shows a slightly decreasing trend, especially at small ν. Moreover, kurtosis is lower when considering relaxed halos; the scatter for both parameters is also lower.

EC model also predicts that e and p are independent on the halo mass when scaled as $\delta e/\sigma(M)$ and $\delta p/\sigma(M)$. We show these distributions in bottom plots of Fig. 6.4 as a function of (the logarithm of) ν. The mean values and the rms of $\langle \delta_L e_L \rangle/\sigma(M)$ (respectively, 0.509 and 0.16) are a bit higher than theoretical predictions ($1/\sqrt{5} \simeq 0.447$ and 0.14) but in the case of $\langle \delta_L p_L \rangle/\sigma(M)$ the accordance is very good (0 and 0.21 against the predictions of 0 and 0.22).

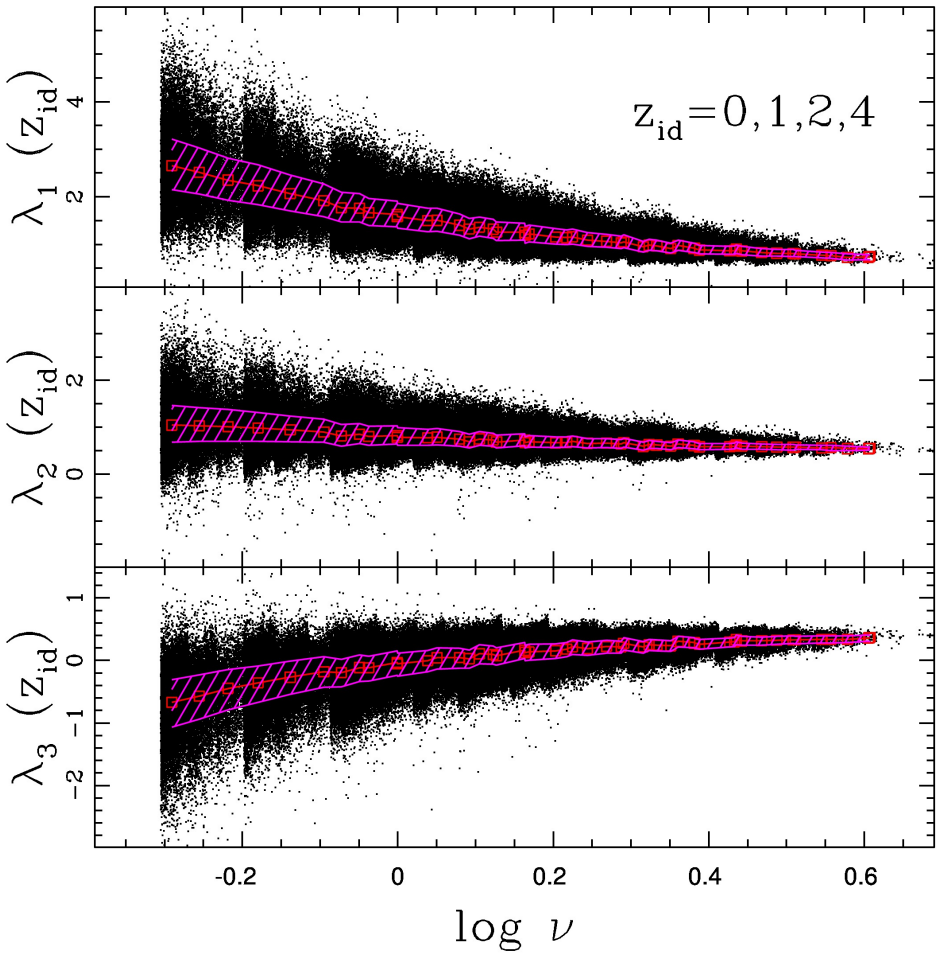

Figure 6.1: Distribution of Lagrangian $\lambda_{1;L}$ (TOP), $\lambda_{2;L}$ (MEDIUM) and $\lambda_{3;L}$ (BOTTOM) as a function of (the logarithm of) ν. Lagrangian values are averaged over spheres of radius equal to the halo characteristic Lagrangian scale $R_L = (3M/4\pi\bar{\rho})^{1/3}$. We plot $\lambda_{i;L}$'s after rescaling their value to the halo identification redshift z_{id}. We show in red the median of the distribution, whereas the magenta shaded region encompass values between the first and the third quartile. Most protohalos have more than one positive eigenvalue.

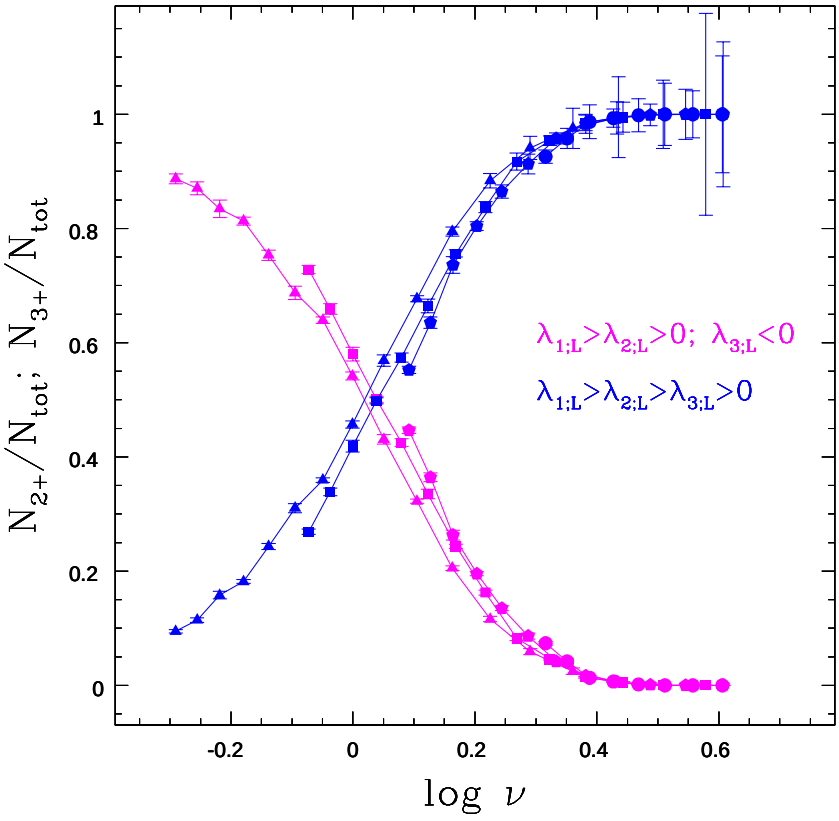

Figure 6.2: Fraction of triplets with two (blue) and three (magenta) positive eigenvalues, as a function of (the logarithm of) ν. We show separately the results for the different identification redshifts: $z_{id} = 0$ (triangles), 1 (squares), 2 (pentagons) and 4 (circles). The number of positive triplets grows as ν increases but decreases with z_{id} (at least at not too large ν).

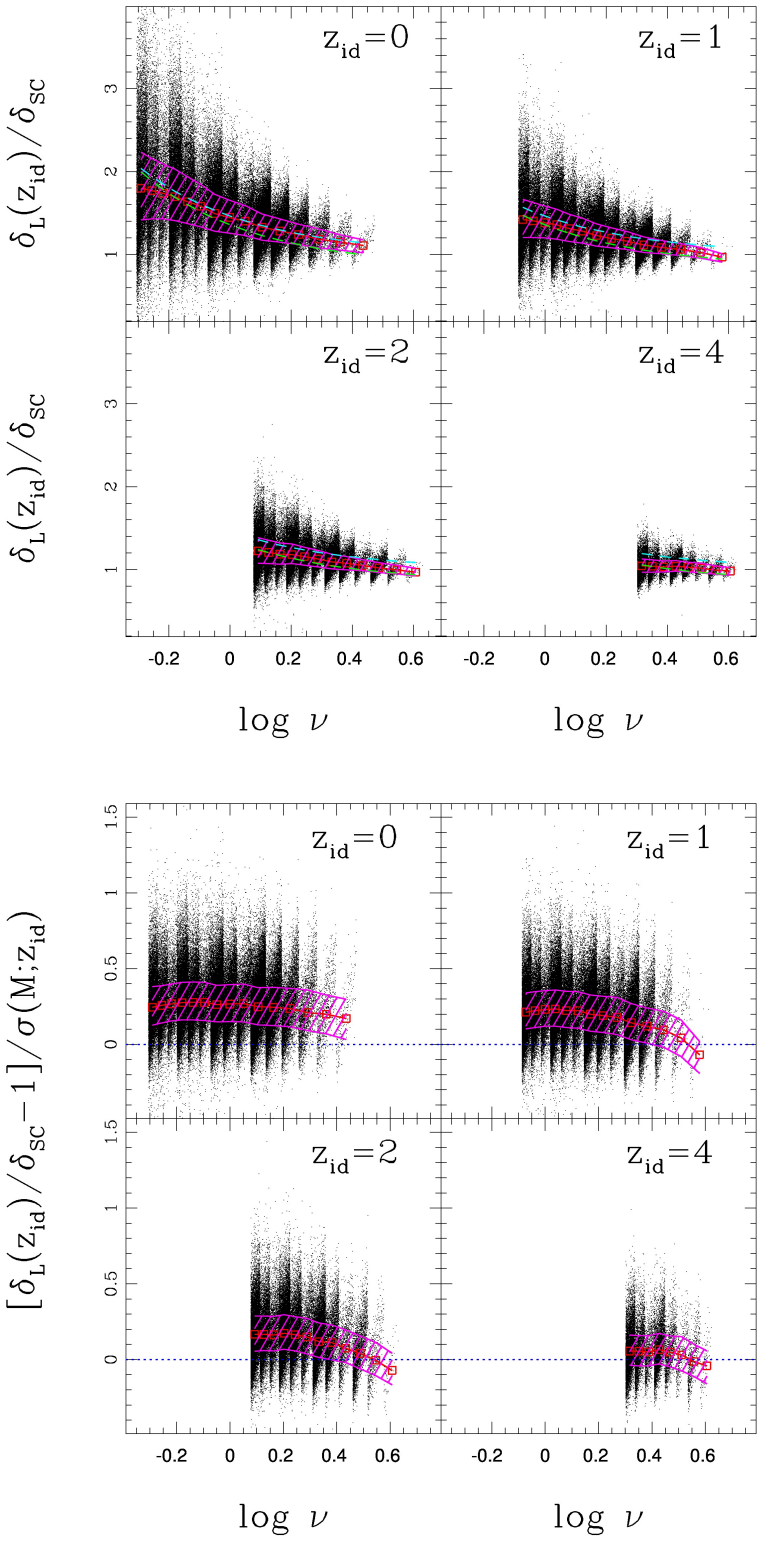

Figure 6.3: Top plot: distribution of δ_L as a function of (the logarithm of) ν for halos identified at the four different redshifts. The Ellipsoidal Collapse barrier (cyan) works well at $z_{id} = 0$ but it always overestimates our results at larger z_{id}. In general, the median of the distribution is always comprised between the EC and the "Sheth et al." barrier. Bottom plot: distribution of δ_L rescaled with $\sigma(M)$ in order to remove the mass dependence. We show median (red) and the first and third quartile (magenta).

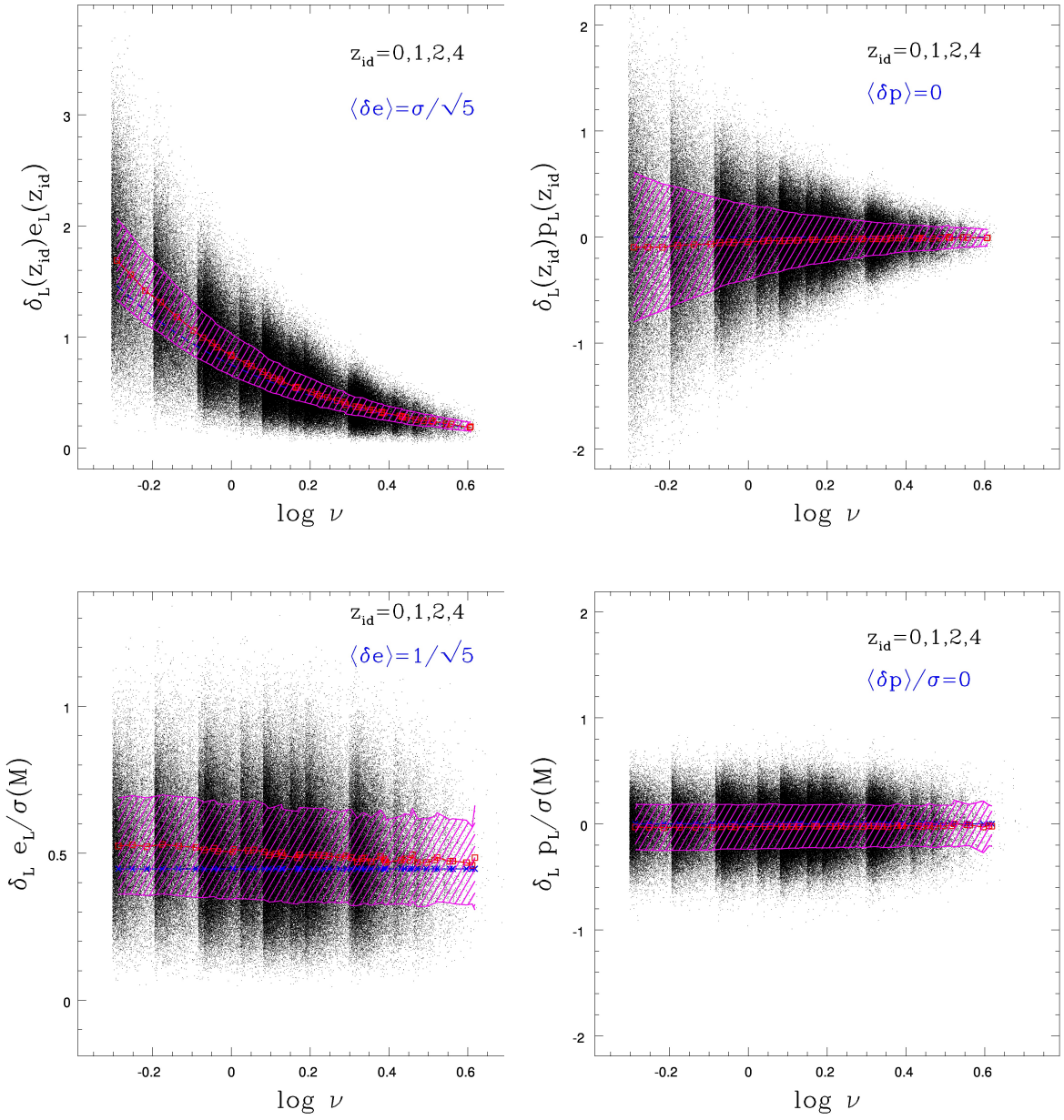

Figure 6.4: Top plots: Distribution of Lagrangian $\delta_L e_L$ (left) and $\delta_L p_L$ (left) as a function of (the logarithm of) ν. Parameters have been rescaled to the identification redshift. Bottom plots: Same distributions but rescaled by dividing by $\sigma(M)$ in order to remove much of the mass dependence. In all four plots we show the median (red) and the region between the first and the third quartile (shaded magenta region) and mean values (blue curves) arising in the EC framework by averaging the distribution of e and p over all position of the Gaussian field.

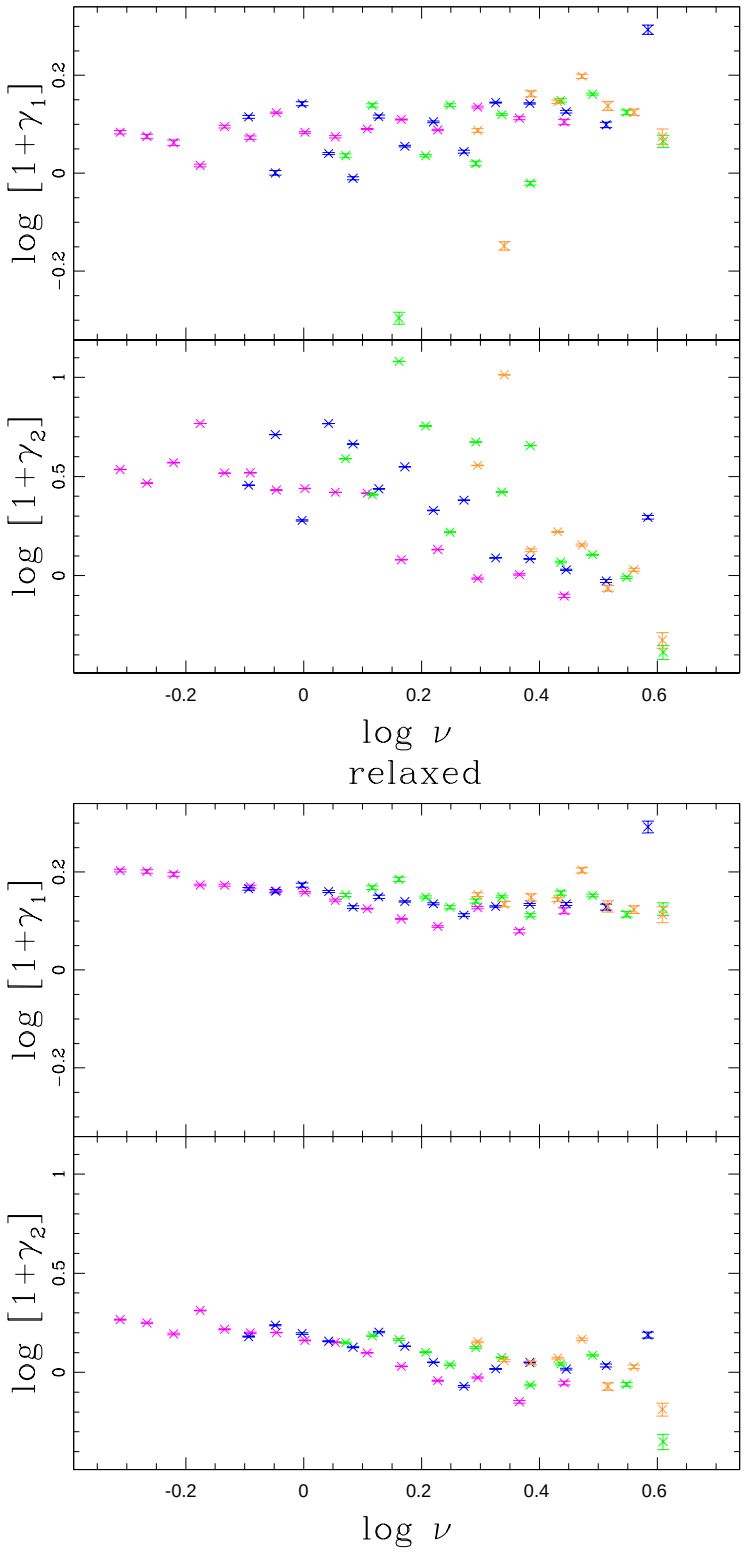

Figure 6.5: Skewness (γ_1; top panel) and kurtosis (γ_2; bottom panel) of the δ_L distribution as a function of ν. We compare results for all (top plot) and relaxed (bottom plot) halos. Different colors correspond to different identification redshifts: $z_{id} = 0$ (magenta), 1 (blue), 2 (green) and 4 (orange). In both cases, skewness is nearly mass invariant, whereas kurtosis shows a slightly decreasing trend, especially at small ν. Moreover, kurtosis is lower when considering relaxed halos and the scatter of both parameters is also lower.

6.2.3 Lagrangian distribution of q_L

Although e and p are largely employed to describe the shape of the gravitational potential, some alternatives have the advantage of not depending on a ratio of eigenvalues. In this work we employ the square root of the traceless shear q, arising from a perturbative analysis of the ellipsoidal collapse model. It can be expressed using the parameters of the ellipsoidal collapse model

$$q^2 = I_1^2 - 3I_2 = \delta^2(3e^2 + p^2) \tag{6.10}$$

where I_1 and I_2 are rotationally invariant quantities equal to

$$I_1 = \lambda_1 + \lambda_2 + \lambda_3 \tag{6.11}$$

$$I_2 = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 \tag{6.12}$$

(Sheth et al. (2013)). We remark that q, differently from e and p, is not coupled with δ and the distribution of q^2 is drawn from a $\chi^2_5(\sigma)$ and so $\langle q^2\rangle \sim \sigma^2$ (Sheth and Tormen (2002), where q is labelled as r). In what follows we consider the Lagrangian traceless shear associated with the protohalo patches (q_L). Top plot of Fig. 6.6 shows that q_L still scales approximately like $\sigma(M)$. Moreover, the distribution of $(q_L/\sigma(M))^2$ is still a $\chi^2_5(\sigma)$ after scaling with

$$\Big(\frac{q_L(z_{id})}{\beta\sigma(M, z_{id})}\Big)^2\sigma(M, z_{id})^\alpha \tag{6.13}$$

where $\alpha \simeq 0.16$ and $\beta \simeq 0.95$.

We can inspect the importance of shear on the collapse threshold using a barrier B depending on the traceless shear parameter q_L, rather than the ellipsoidal collapse parameters e_L and p_L:

$$B(q_L) = \delta_c\Big(1 + \sqrt{q_L^2(z_{id})/q_c^2}\Big) \tag{6.14}$$

where q_c decreases when the shear becomes more important. This simple model provides an easy way to see how effects associated with non spherical collapse depends on both ν and z_{id} and permits a comparison with previous works. Fig. 6.6 shows the distribution of $\Big[\delta_L - B(q)\Big]/\sigma(M)$ as a function of $1/\nu$. At $z_{id} = 0$ we notice that the scaling by $\sigma(M)$ removes most of the mass dependence and a reasonable value of q_c^2/δ_c^2 is ~ 4, whereas previous works found $q_c^2/\delta_c^2 \sim 6$ (Despali et al. (2013), Sheth et al. (2013)). The importance of shear decreases (q_c increases) as z_{id} increases. Moreover, given a identification redshift, we find that at the smallest ν (high $1/\nu$) the distribution depends weakly by mass; on the contrary, as ν grows the effects of shear becomes less important and the distribution quickly decreases. The presence of a trend with ν means that the actual scaling is not perfectly proportional to q_L.

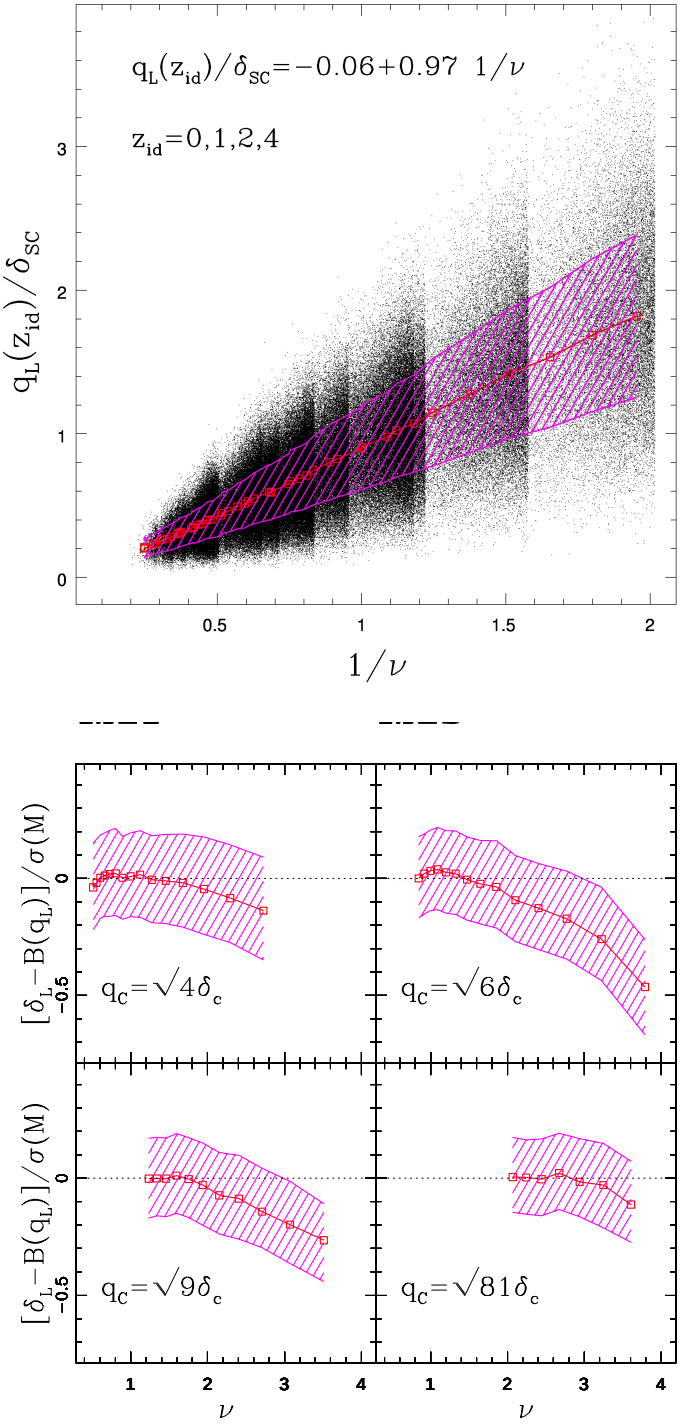

Figure 6.6: Top plot: distribution of q_L/δ_{SC} as a function of $1/\nu$. The parameters has been rescaled to the identification redshift. We show the median (red) and the values between the first and the third quartile (shaded magenta). Bottom Plot: distribution of $\Big[\delta_L(z_{id})-B(q_L(z_{id}))\Big]/\sigma(M,z_{id})$ as a function of $1/\nu$ at the different identification redshifts. We notice that the importance of shear decreases as z_{id} increases.

6.3 Lagrangian correlations

In Section 6.2.3 we showed how Lagrangian overdensity depends on the shear parameter q_L. The existence of such a correlation for halos identified at the present time is well known and it implies that if shear is high a protohalo needs a large overdensity to keep itself together (Sheth et al. (2001), Sheth and Tormen (2002), Despali et al. (2013)). However, the presence of substantial scatter around the median in Fig. 6.6 suggests that other quantities other than q_L determine the value of δ_L (Sheth et al. (2013)). Recent works have highlighted that the initial overdensity correlates with the formation redshift: higher overdensity means that halos form earlier. Nevertheless, it is not clear if the correlation between shear and overdensity is direct or if it is a consequence of the correlation between shear-formation time and formation time-overdensity. In order to clarify this point we study the correlation between Lagrangian δ and both formation time (at fixed shear) and shear (at fixed formation time). In what follows, the formation time of a halo is defined as the redshift at which half of the final halo mass has been assembled into a main progenitor i.e. z_{50} (Lacey and Cole (1993)).

In order to remove the dependence of correlations on scatter, we calculate normalized quantities by subtracting the mean and dividing by the rms:

$$\Delta = \frac{\delta_L - \langle \delta_L \rangle}{\sigma(\delta_L)} \tag{6.15}$$

$$\zeta = \frac{z_{50} - \langle z_{50} \rangle}{\sigma(z_{50})} \tag{6.16}$$

$$\theta = \frac{q_L - \langle q_L \rangle}{\sigma(q_L)} \tag{6.17}$$

Therefore the relation between between Δ and both normalized shear and formation time can be written in the form:

$$\Delta = a_\zeta \zeta + b_\theta \theta \tag{6.18}$$

where a_ζ and b_θ can be estimated with a least square fit. Our aim is therefore to find the values of a_ζ and b_θ as a function of the halo mass and identification redshift.

We firstly investigate the correlation between Δ and ζ at fixed shear. We subtract from Eq. 6.18 the same equation at fixed shear, obtaining:

$$\Delta - \langle \Delta|\theta \rangle = a_\zeta(\zeta - \langle \zeta|\theta \rangle) + b_\theta(\theta - \langle \theta|\theta \rangle) \tag{6.19}$$

$$= a_\zeta(\zeta - \langle \zeta|\theta \rangle) \tag{6.20}$$

We therefore obtain two new variables:

$$\Delta - \langle \Delta \Big| \theta \rangle \tag{6.21}$$

$$\zeta - \langle \Delta \Big| \theta \rangle \tag{6.22}$$

by subtracting from Δ and ζ the correlation of both variables with the third variable θ. We plot results for $z_{id} = 0$ in left plot of Fig.6.7. We notice the presence of a tail of points at high z_{50} and low masses. At the lowest mass ($nu = 0.52$) these points are about 7% of the total and their origin can be related to the non gaussianity of the z_{50} distribution (Lacey and Cole (1993)) as well as misleading identifications of the halo finder. When considering only relaxed halos (left plot of Fig.6.7), this tail disappears and the slopes of the least square fit are slightly lower (right plot of Fig.6.7). We show how the different relaxation criteria affects the coefficient a_ζ in left plot of Fig. 6.8. The first two criteria slightly affects the correlation only at the lowest masses, raising (the first criterion) or lowering (the second criterion) the slope a_z. The third criterion tends to lower the slope at all masses. However, a_z shows a similar trend with mass whether or not a relaxation criterion is imposed. Right plot of Fig. 6.8 shows the same distribution for all halos considering together all identification redshifts. We notice that the coefficient a_z reaches a maximum around $\nu \sim 2$ (corresponding to a mass of about $32M_* \simeq 16 \times 10^{15} M_\odot$) and then decreases down to $a_\zeta \sim 1.5$. At larger ν the error on the slopes becomes very large because of the small halo statistics.

We now investigate in the same way the correlation between Δ and θ at fixed formation time. We subtract from Eq. 6.18 the same equation at fixed formation time, obtaining:

$$\Delta - \langle \Delta | \zeta \rangle = b_\theta (\theta - \langle \theta | \zeta \rangle)$$

that relates the two variables

$$\Delta - \langle \Delta \Big| \zeta \rangle \tag{6.23}$$

$$\theta - \langle \Delta \Big| \zeta \rangle \tag{6.24}$$

. Fig. 6.9 shows that the correlations at $z_{id} = 0$. When considering all halos (left) the slope of the least square fit is a bit lower than for relaxed halos (right). However, the trend with mass is again very similar. Fig. 6.10 shows the coefficient b_θ for all halos and all identification redshifts. At low ν the correlation between the two variables is high but strongly decreases going towards large ν.

We can compare the importance of the two correlations by calculating:

$$\Delta - \langle \Delta | \theta \rangle = c_{\theta\zeta} (\Delta - \langle \Delta | \zeta \rangle) \tag{6.25}$$

where

$$c_{\theta\zeta} = \frac{a_\zeta}{b_\theta} \frac{\zeta - \langle \zeta | \theta \rangle}{\theta - \langle \theta | \zeta \rangle}. \tag{6.26}$$

Left plot of Fig. 6.11 shows this correlation at $z_{id} = 0$, whereas the right plot shows the distribution of the coefficient $c_{\theta\zeta}$. In order to have a more reliable estimate of the correlation between the two variables, we average the slope of the least square fit with the slope of the same fit obtained by inverting the axes. The error on the slope is now given by half the difference between the two slopes.

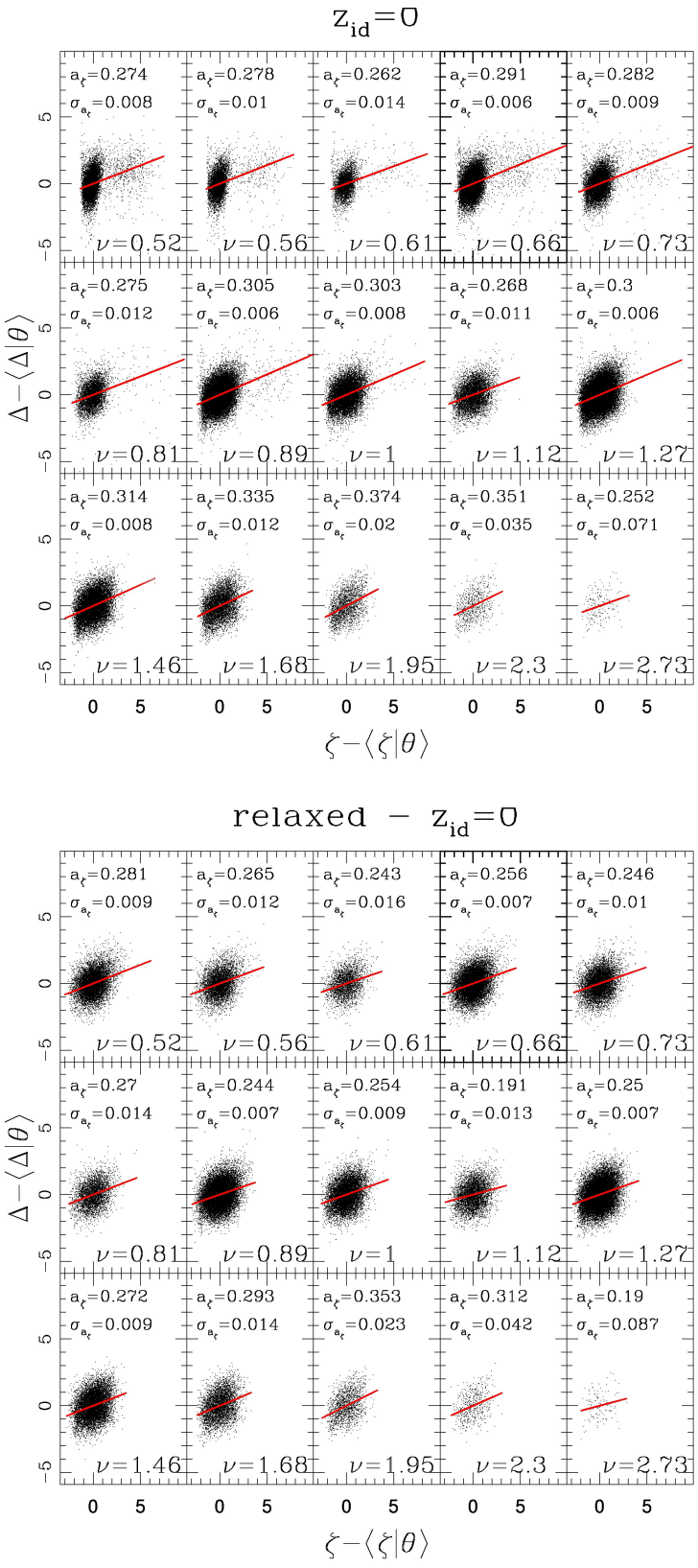

Figure 6.7: Correlation between between θ and Δ at fixed ζ for halos identified at $z_{id} = 0$. We consider both all (top) and relaxed (bottom) halos. σ_{a_ζ} measures the slope uncertainty. When imposing relaxation the tail of points at high z_{50} disappears.

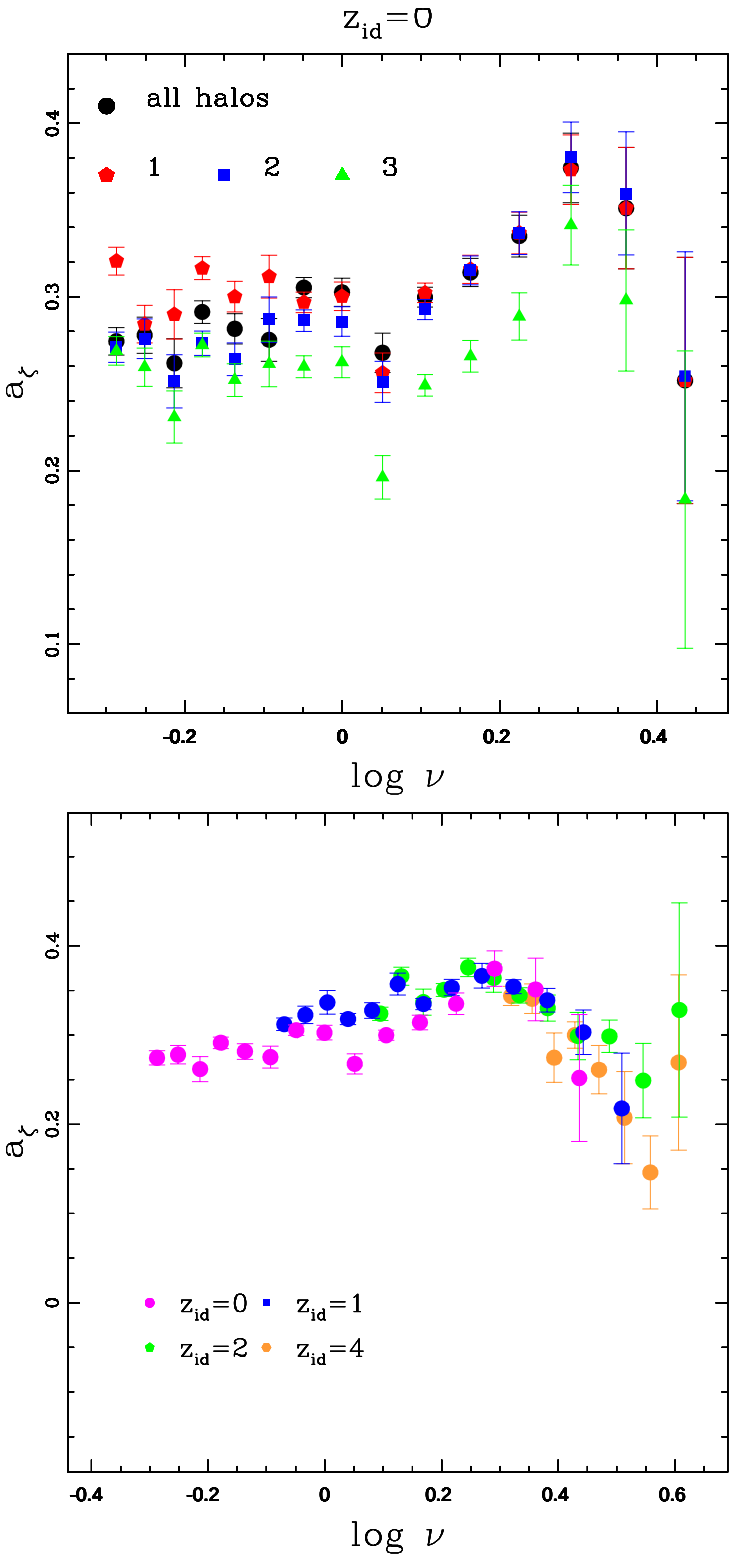

Figure 6.8: Top plot: Distribution of the slope a_ζ of the least square fit between Δ and ζ at fixed θ, as a function of ν. We consider only halos identified at z_{id} and we show results for both all and relaxed halos. We consider separately three relaxation criteria (the meaning of numbers is defined in Sec. 4.1.3). Whereas the first two relaxation criteria affects only low mass halos, the third one lower the slope at all masses. Bottom plot: Distribution of a_ζ as a function of ν for all halos and all identification redshifts.

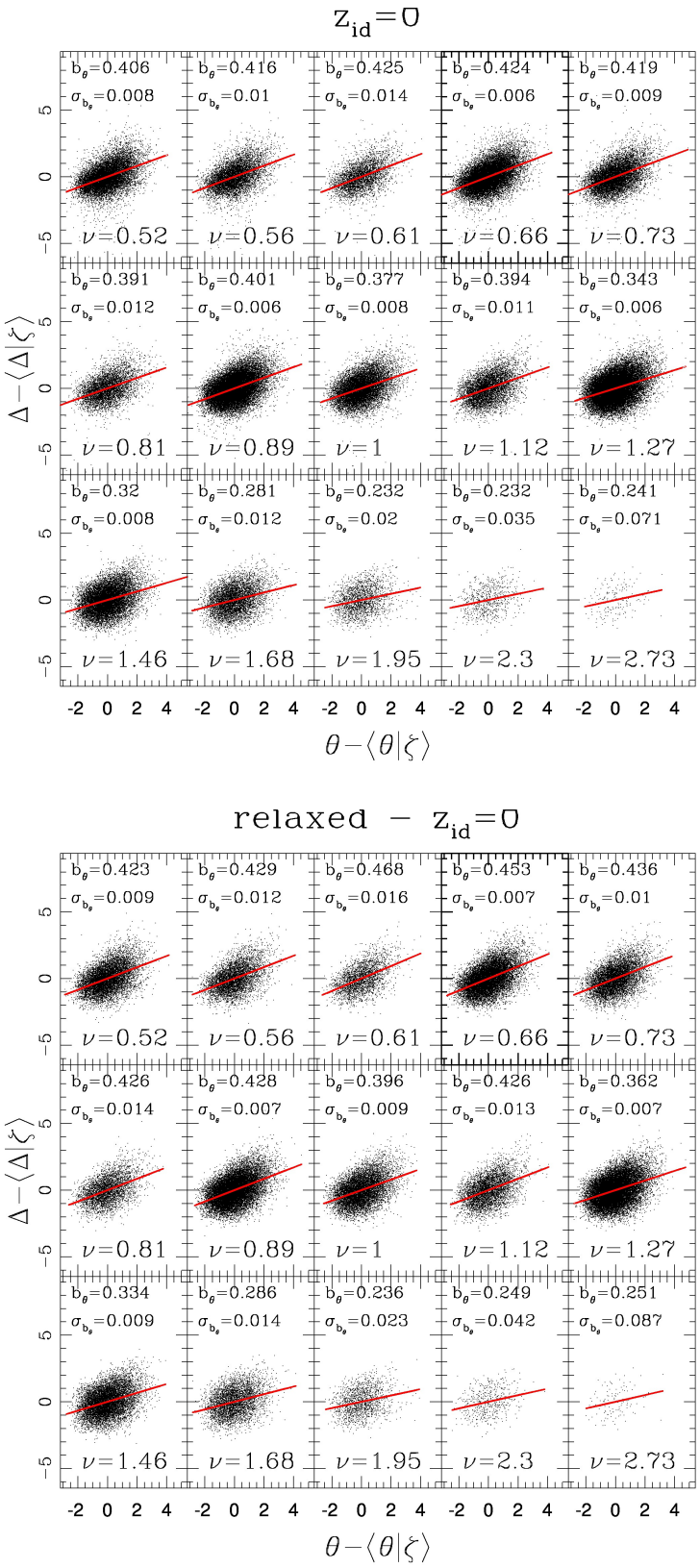

Figure 6.9: Correlation between Δ and θ at fixed ζ for halos identified at $z_{id} = 0$. We show a comparison between all (left) and relaxed (right) halos. σ_{b_θ} measures the slope uncertainty.

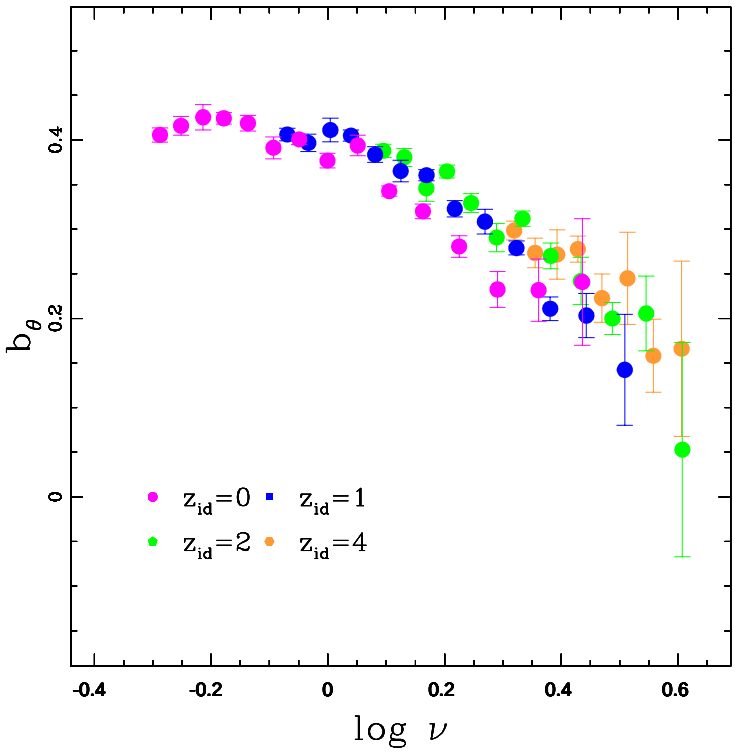

Figure 6.10: Slopes of the least square fit between Δ and θ at fixed ζ as a function of the halo mass for all halos and all identification redshifts. We find $b_\theta \sim 0.4$ at low masses and then the slope of the least square fit decreases going towards larger ν.

So, we find that shear is more important than z_{50} on determining δ_L at low masses. When ν grows, the correlation with shear decreases faster than the correlation with the formation time and at the highest ν both correlations are low.

6.4 Summary

In this chapter we would understand the importance of shear in the formation of halos by inspecting Lagrangian parameters of the ellipsoidal collapse model together with the traceless shear parameter q. Moreover we investigated the correlation of shear, overdensity and formation time.

Lagrangian parameters are obtained by centering on the protohalo mass center and by averaging over spheres with radius equal to the Lagrangian halo scale. In this analysis we considered four halo identification redshifts, spanning a large range of ν. We found that,

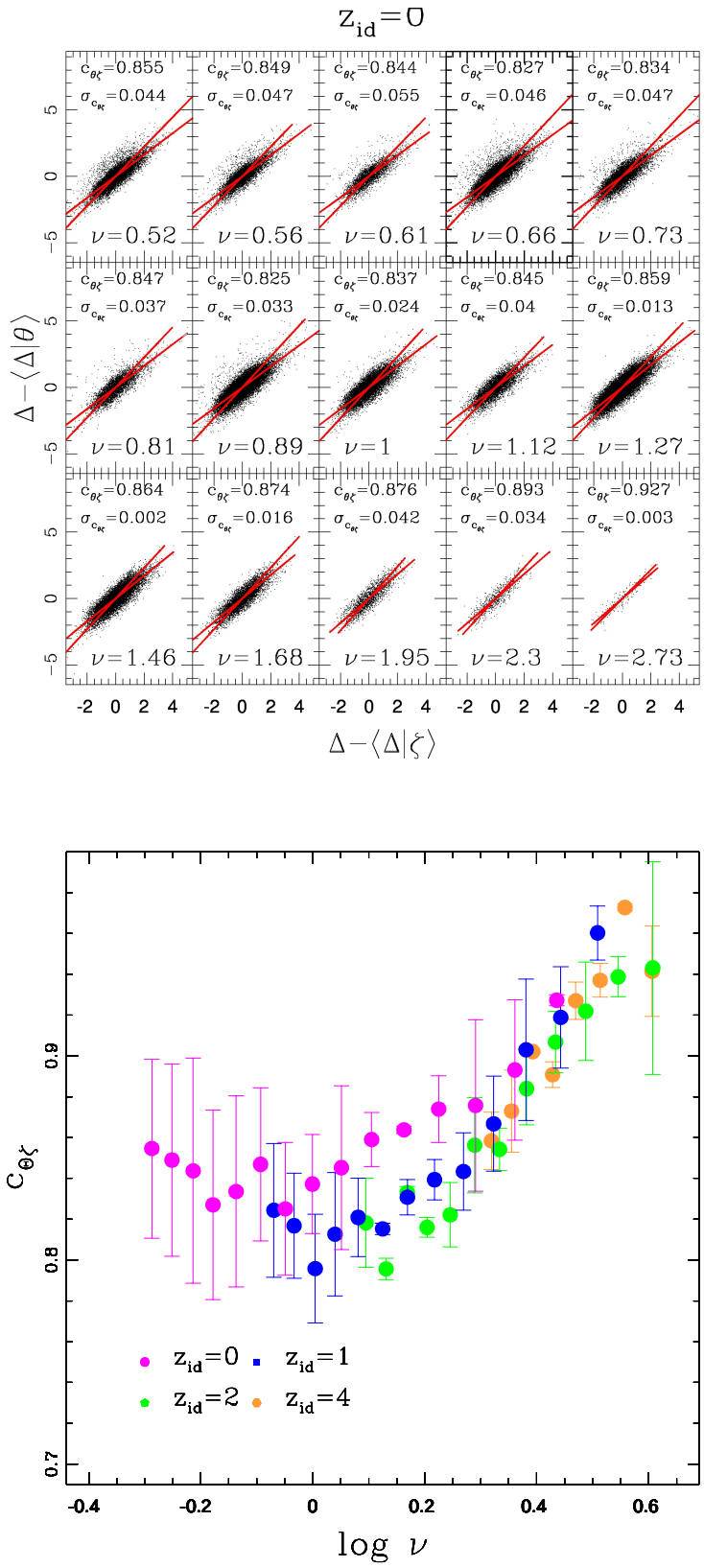

Figure 6.11: Mean values of the two slopes of the least square fit of Δ at fixed ζ and Δ at fixed θ for all halos and identification redshift. The error bar is half the difference of the two slopes.

although only two eigenvalues are generally positive at small masses, most protohalos have three positive eigenvalues. The distribution of δ_L/δ_C is approximately equal to $(1 + 0.28\sigma(M))$ with a rms of nearly 0.18. These values are larger than those found by Despali et al. (2013) but are in good agreement with Robertson et al. (2009), who used a method similar to ours in order to calculate Lagrangian parameters (i.e. they assume a spherical shape for protohalos). We find a nearly mass independent distribution for $\delta e/\sigma(M)$ and $\delta p/\sigma(M)$. Moreover, we find that the distribution of δ_L is Gaussian and q_L^2, when opportunely rescaled, is a χ^2 with five degrees of freedom.

We then investigated the correlation between Lagrangian δ and q and formation time. We would understand if the well known shear-overdensity correlation is the result of the formation-time correlation i.e. if shear and δ_L correlate cause of a correlation between shear-formation time and formation time-overdensity. This is the first time that the analysis of these correlations has been investigated at redshift larger than zero. We found a direct correlation between overdensity and both Lagrangian shear and formation time. This means that the correlation of shear-overdensity is not due to the correlation of formation time-overdensity. We found also that the first correlation depends strongly on mass (with a decreasing trend), whereas for the second one the mass dependence is much weaker, at least at $log\nu < 0.5$.

CHAPTER 7

Properties of Lagrangian and Eulerian profiles

In Chapter 6 we considered the Lagrangian parameters δ_L and q_L i.e. we centered on the protohalo mass centers and we took into account parameters smoothed on the Lagrangian scale of the halos. Here we consider profiles. The main question we wish to address is one which arises naturally in the excursion set and peaks models of halo formation: namely, how does the slope of the initial profile within the Lagrangian radius correlate with its height on larger scales. This sort of correlation is often called Assembly bias, although in the context of halo formation it is also sometimes called non-local or stochastic bias. We look to see how the initial shear (the stochastic variable) impacts halo concentrations and large scale bias.

In Section 7.1 we build Lagrangian profiles of both linear q and δ around the protohalo centers and we investigate how the inner and the outer linear density profile correlate with Lagrangian shear and halo formation time, as a function of the universal mass ν. In Section 7.2 we consider Eulerian profiles built around both the center of mass of final halos and the center of mass of the halo particles traced back to larger redshifts. We show how the correlation with Lagrangian shear and formation time affects these profiles. Finally, we study the evolution of profiles with time: we compare the Eulerian profiles with the predictions of Bernardeau (1993) and we study the evolution of the inner region when considering profiles built around both the particle mass center and the center of mass of the main progenitor.

7.1 Lagrangian profiles

Given halos at different masses and identification redshifts, we trace the halo particles back to the initial conditions and we consider protohalos. To build profiles we interpolate

the values of δ and q smoothed on different scales to the location of protohalo centers. The smoothing scales are the Lagrangian protohalo scale (R_L) together with twelve additional scales: six larger and six smaller than the Lagrangian one. In this way we can investigate a region from $0.25R_L$ to about $7.6R_L$. (More details on protohalos and the choice of smoothing scales can be found in Sec. 2). Profiles are then rescaled to $z = 0$ using the scale factor. Fig. 7.1 shows (logarithmic) linear δ profiles of protohalos related to four different halo identification redshifts. Each profile has been normalized by dividing by the mass variance related to that halo mass ($\sigma(M_L)$). In this way we can delete much of the dependence on power spectrum. We notice that at fixed z_{id}, for example $z_{id} = 0$, δ at Lagrangian radius ($log(R/R_L) = 0$) grows as ν grows (top left plot of Fig.7.1). This is consistent with what showed in the Chapter 6 when analyzing $\delta/\delta_{SC} - 1$ as a function of ν (top left panel of the right plot in Fig.6.3) (see also Despali et al. (2013) and Robertson et al. (2009)). Given two δ_L (say $\delta_{L,1}$ and $\delta_{L,2}$) corresponding to different ν (ν_1 and ν_2) and two different mass variance σ_1 and σ_2, right plot of Fig. 6.3 shows that, at given z_{id}, we have:

$$\frac{\left(\delta_{L,1}/\delta_{SC} - 1\right)}{\sigma_1} \simeq \frac{\left(\delta_{L,2}/\delta_{SC} - 1\right)}{\sigma_2} \tag{7.1}$$

and so:

$$\frac{\delta_{L,1}}{\sigma_1} = \frac{\delta_{L,2}}{\sigma_2} + (\frac{1}{\sigma_1} - \frac{1}{\sigma_2})\delta_{SC} \tag{7.2}$$

. If $\nu_1 < \nu_2$, we have $\sigma_1 > \sigma_2$ and the first member of this equation is larger than the second member, as we showed in profiles. We can measure the average steep of the inner profiles by measuring the different height of profiles between the innermost point and the Lagrangian radius. When considering $z_{id} = 0$, we find

$$\log\left[1 + \delta(0.25R_L)/\delta_{SC}\right] - \log\left[1 + \delta(R_L)/\delta_{SC}\right] \simeq 0.2 \tag{7.3}$$

that means $\frac{1+\delta(0.25R_L)}{1+\delta(R_L)} \simeq 1.6$, nearly invariant with ν but slightly decreasing as the identification redshift grows. On the contrary, the outer profiles becomes steeper as ν increases.

We now investigate how the Lagrangian δ profiles correlate with formation time and Lagrangian shear.

Given a halo mass and a identification redshift, we consider four bins in the halo formation time z_{50} so that each bin contains nearly the same number of halos. We then average profiles within each bin and plot the four profiles together in Fig. 7.2). We see that the binning in formation time strongly affects the inner protohalo region so that higher formation times correspond to higher δ. This correlation becomes less strong at scales larger than the Lagrangian radius and it nearly disappears at very large scales (except for low ν, where the profile goes up again at the largest scales). Moreover, as the identification redshift grows, the correlation between δ and z_{50} seems to become less strong. The top plots of Fig. 7.2 show a comparison at $z_{id} = 0$ between all (left) and relaxed (right) halos (more details on relaxation criteria can be found in Sec. 4.1.3). Although the correlation

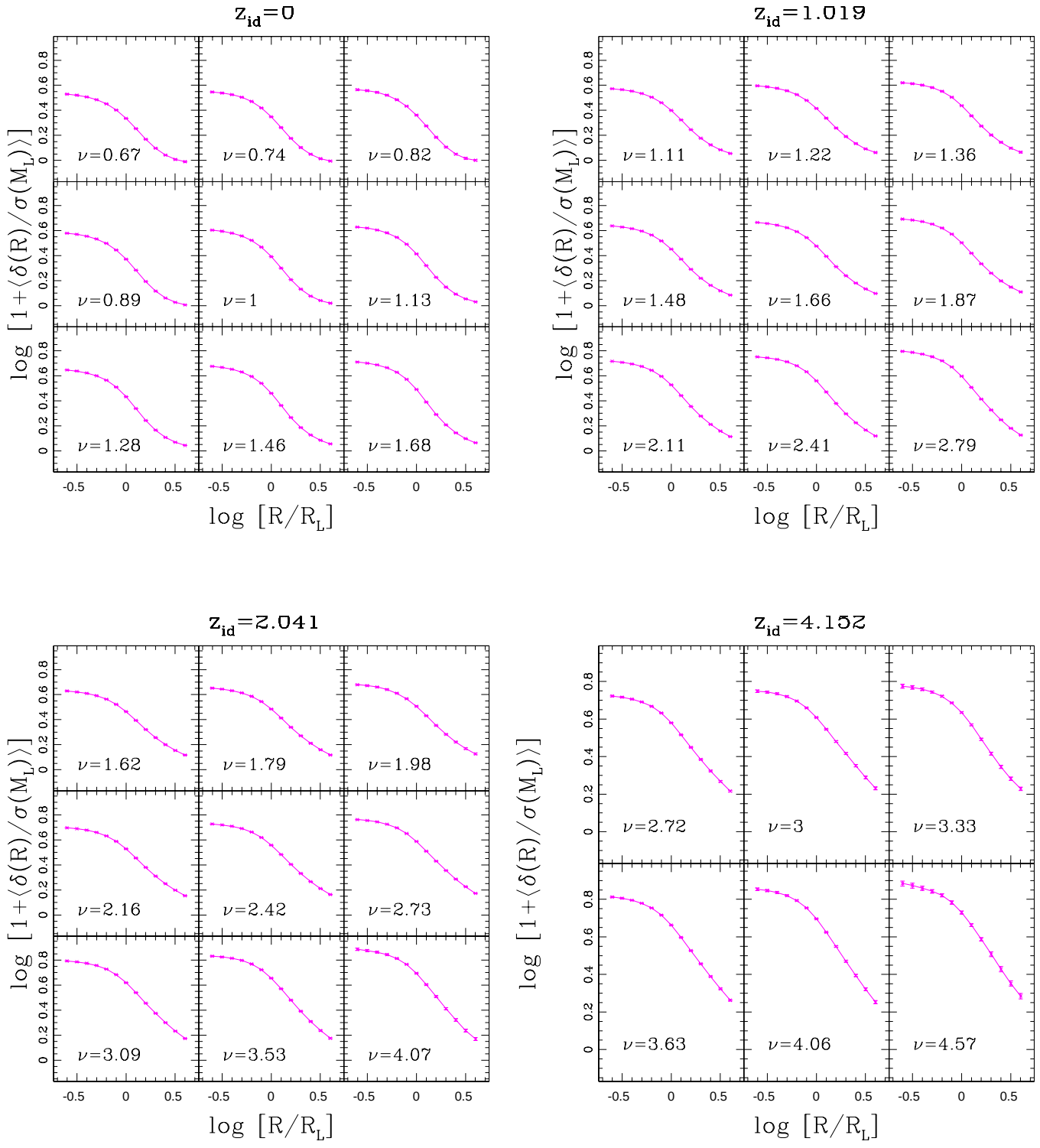

Figure 7.1: Lagrangian δ profiles of protohalos at different ν and different halo identification redshifts. δ has been normalized by the mass variance in order to avoid much of the dependence on the power spectrum. Profiles are rescaled at $z = 0$.

does depend very slightly on relaxation, the difference between innermost profiles at high and low z_{50} is a bit smaller when considering relaxed halos. This is mainly due to the first criterion, that removes halos that during their formation history acquire more mass than the final halo mass and then lose it in recent times. These halos happen to have a larger formation time and affect the correlation. Unless otherwise specified, from here on out we consider only relaxed halos.

We now consider profiles binned in q_L. Similarly to the previous case, we consider four bins and we average profiles within each bin. We notice that the correlation of δ profiles is similar at both high and small smoothing scales i.e. when q_L is higher, the profile is also higher at all scales. However, this correlation decreases going towards larger ν. Moreover, the correlation in the innermost region is always less strong than that obtained binning in formation time. At very high z_{id} the correlation is less clear, because of a low halo statistics.

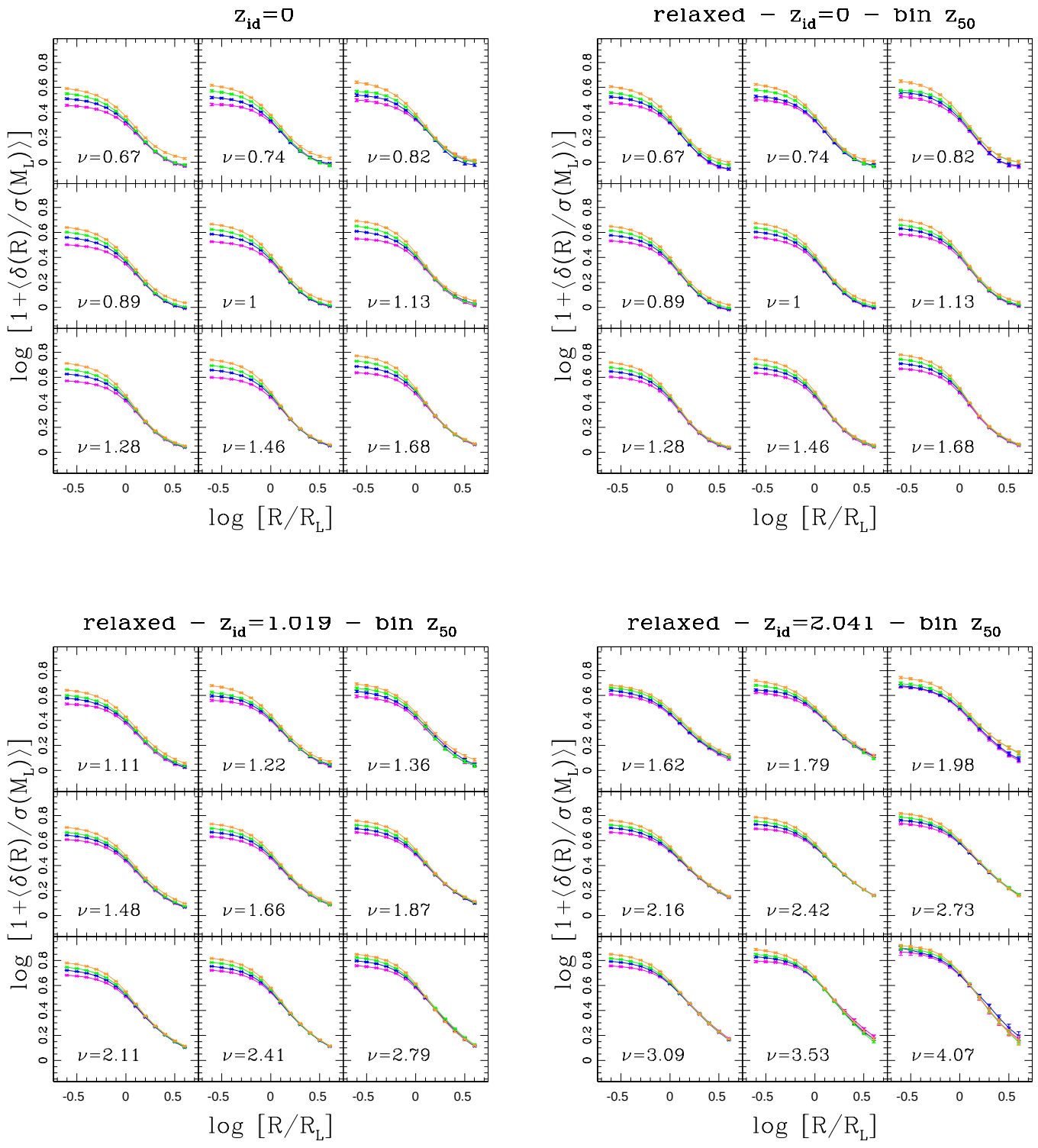

Figure 7.2: Lagrangian δ profiles of protohalos at different ν and different halo identification redshifts. We bin in z_{50} so that each bin contains nearly the same number of halos. Going from low to high z_{50}, the colors are the following: magenta, blue, green and orange.

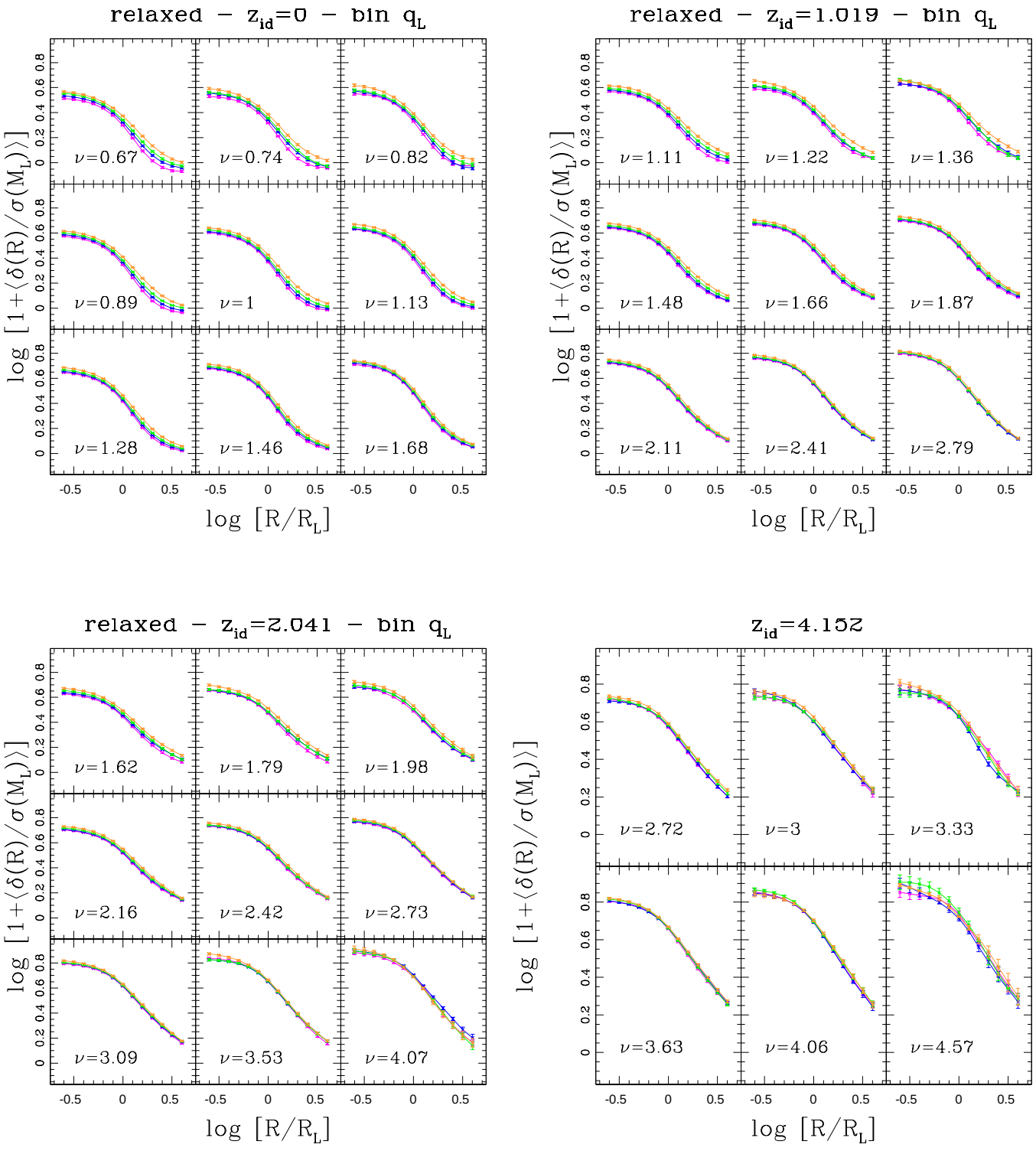

Figure 7.3: Lagrangian δ profiles of protohalos at different ν and different halo identification redshifts. We bin in q_L so that each bin contains nearly the same number of halos. Going from low to high q_L, the colors are the following: magenta, blue, green and orange.

7.2 Eulerian profiles

Whereas in the previous section we took into account linear profiles, we now consider Eulerian ones. We consider the same halos at the snapshot corresponding to their identification redshift and we build profiles around the center of mass, as follows. We calculate the cumulative density within spheres of equally-spaced logarithmic comoving radii up to 10 times the virial radius. For each simulation we decide to cut the inner region of profiles at two times the softening for computational motivations (softening for the different simulations is listed in 3.1). Fig. 7.4 shows Eulerian profiles in comoving coordinates for different z_{id}.

7.2.1 Comparison between Eulerian profiles and Lagrangian quantities

Non linear prediction of the outer halo profile

Lagrangian profiles can be used to predict qualitatively the outer region of the final halo profiles. In principle we could rescale Lagrangian profiles at the time of the halo identification redshift using the growth factor and see if Lagrangian and Eulerian profiles match well. However, when the perturbations become nonlinear ($\delta \simeq 1$) the evolved linear profiles are clearly not a good approximation of the final shape. In the spherical collapse model a relation between δ_L and δ_{NL} was provided by Bernardeau (1993). Given a linear profile, its non linear evolution at a certain redshift assumes the form:

$$1 + \delta_{NL} = \left(1 - \frac{D(z)\delta_L^{ic}(< R_L)}{\delta_C}\right)^{\delta_C} \tag{7.4}$$

where δ_L^{ic} is δ_L at the initial conditions, $D(z)$ is the growth factor and the exact value of δ_C (the critical value for collapse) depends on the background cosmology. We assume $\delta_C = 1.675$. The nonlinear scale is obtained from the Lagrangian one using:

$$r_{NL} = R_L \left(1 - \frac{D(z)\delta_L^{ic}(< R_L)}{\delta_C}\right)^{-\frac{\delta_C}{3}} \tag{7.5}$$

Given these equations we compare the halo density profile at a certain z with the evolved linear one, bearing in mind that the equation 7.4 works only for $\delta = D(z)\delta_L^{ic}$ smaller than the critical value δ_C predicted by the linear collapse model i.e. it describes profiles only in the outer halo regions. We plot three different profiles: the outer cumulative density halo profile at a certain z_{id}; the linear profile rescaled at the same z; the non linear parametric profile built from the linear one. We can see our results in Fig. 7.5, 7.6 and 7.7. The top left figures show these comparison for halos of three different masses identified at $z_{id} = 0$: $M_*/8$ ($\nu = 0.74$), M_* ($\nu = 1$) and $16M_*$ ($\nu = 1.68$). We show that the outermost region of final halo profiles (magenta curve) at high ν is consistent with non linear predictions (dashed red curve). The other three panels show a comparison with the outermost profile

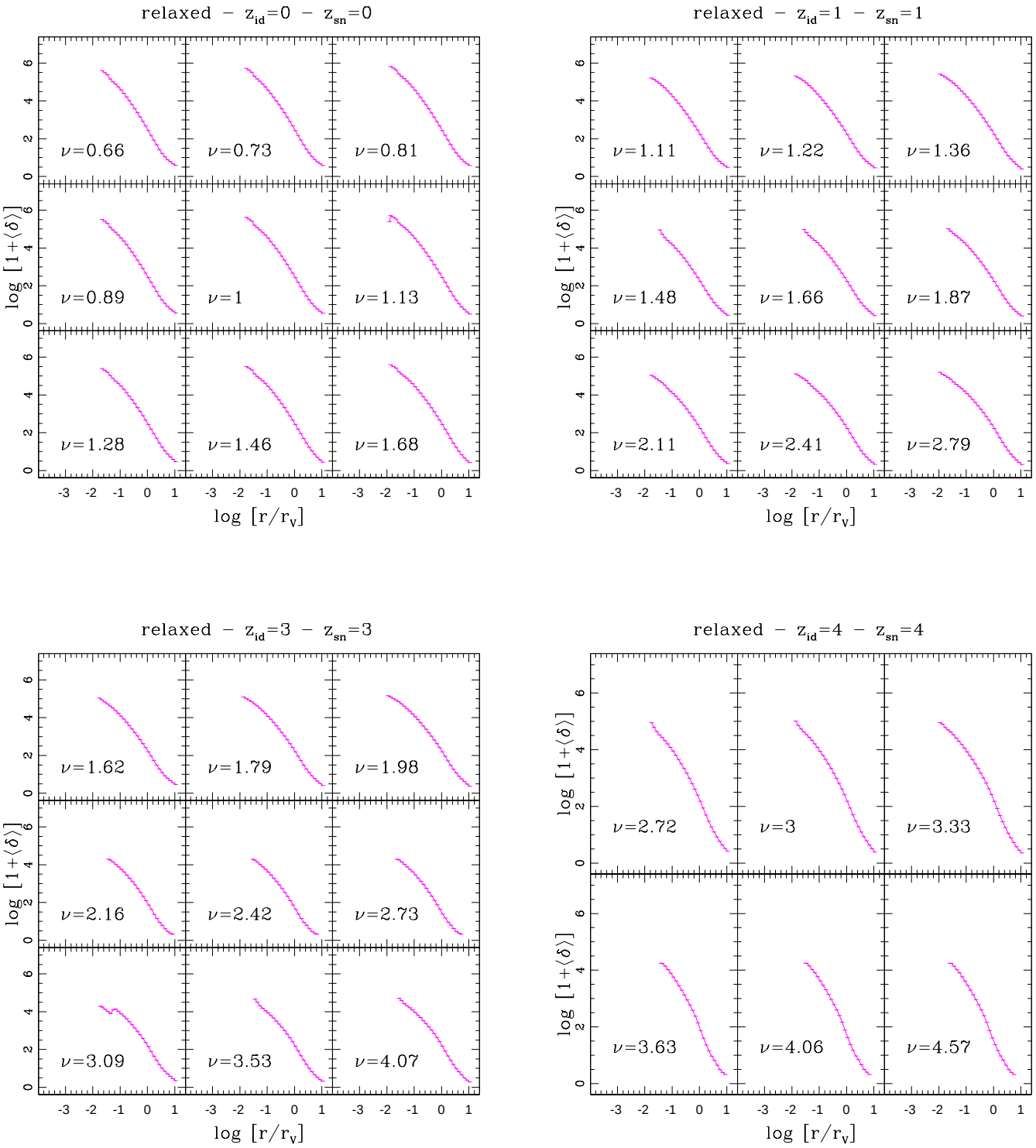

Figure 7.4: Cumulative density profiles in comoving coordinates of halos identified at four different redshifts. The scales go from two times the softening scale up to ten virial radii.

obtained by tracing the halo particles back to three earlier redshifts: $z_{sn} = 1$ (top right), $z_{sn} = 2$ (bottom left) and $z_{sn} = 4$ (bottom right). As z_{sn} increases, the non linear prediction match better the particle profiles, especially at high ν. At high z_{sn} and high radii the three profiles coincide i.e. the halo particle profile is very well reproduced by the linear profile.

7.2.2 Correlation of Eulerian profiles with Lagrangian shear and formation time

We want now to understand how the Lagrangian shear and the formation z affect the final Eulerian halo density profile. In order to have a clear overall view of this correlation we proceed as follow. Given a identification redshift and a mass bin, we first investigate the correlation with shear at given formation time: we consider two bins of z_{50} and then we bin again using four bins of the Lagrangian shear q_L. We then repeat exactly the same procedure, but considering small and large q_L and then binning again in formation time. In Fig. 7.8, 7.9 and 7.10 we show results for profiles of halos with three different characteristic masses ($M_*/16$, M_* and $16M_*$) identified at $z_{id} = 0$ together with their particle profiles at three more redshifts: $z_{sn} = 1$ (top right plot), $z_{sn} = 2$ (bottom left plot) and $z_{sn} = 4$ (bottom right plot). Top panels of each plot show the correlations with q at given shear, whereas in the bottom panels we show the correlation with shear at given formation time. All profiles have been divided by the total averaged one. When considering profiles at the halo identification redshift (top left plots), nearly all profiles match at a certain radius. This corresponds to the halo virial radius at which all halos have been identified with the same density, by definition of the Spherical Overdensity algorithm (see 4.1. In top right plots, profiles within the virial radius are still ordered both in z_{50} and q_L so that higher z_{50} and higher q_L correspond to higher profiles. However, the contribution of formation time and shear is not the same: when fixing low or high z_{50} (top left and top right panels) we show that profiles are respectively lower and higher than the total averaged one. On the contrary, when fixing low or high q_L (bottom left and bottom right panels), profiles can be still found both below and above the total averaged profile, depending on the formation time. So, the correlation with formation time is more important than the correlation with Lagrangian shear. At scales larger than the virial radius the ordering of profiles is generally reversed, as expected. Indeed, profiles of different height must have the same overdensity at the virial radius (by definition) and so at larger scales higher profiles become lower and vice versa. At very high scales, the ordering of binned profiles in bottom panels reverses again. This effect is probably related with halo clustering and will be part of a further analysis in future works. When considering particle profiles traced back at higher redshifts (top right and bottom panels) the ordering within the virial radius is preserved and the ratio of the profile height is enhanced. At large radii the contribution of shear becomes more important and the relative height of profiles resemble that found in Lagrangian profiles: at small ν, it is larger when binning in q_L and smaller when binning in z_{50}, whereas at large ν the difference is less marked.

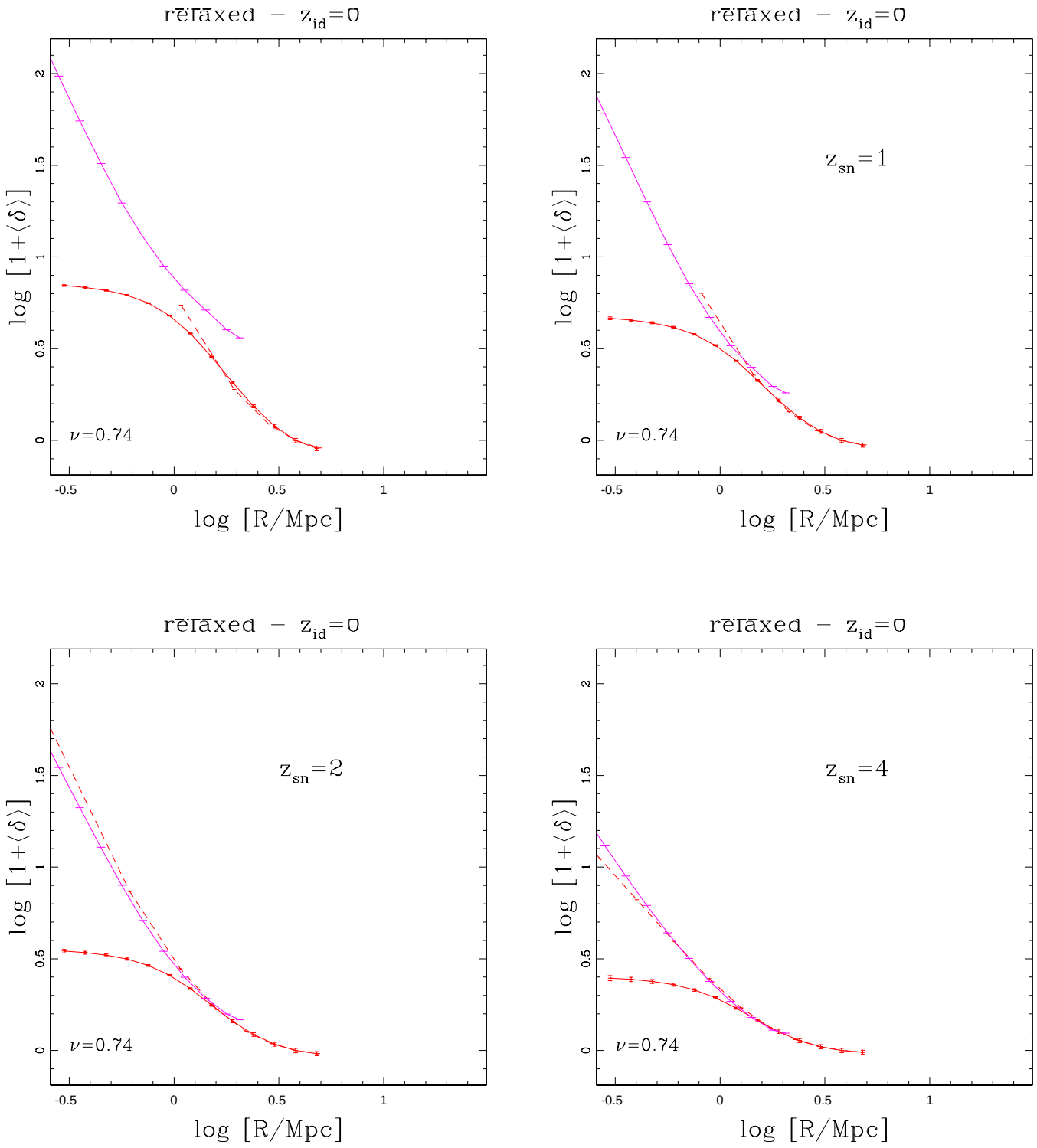

Figure 7.5: Top left panel: comparison between the outermost Eulerian profile (solid magenta curve), the non linear prediction (dashed red curve) and the linear profile (red curve) for relaxed halos of mass $M_*/8$ identified at $z_{id} = 0$ ($\nu = 0.74$)). The other three panels show the same comparison after tracing the halo particles back to different redshift ($z_{sn} = 1, 2, 4$) and rescaling the linear profile and non linear prediction to the same redshifts.

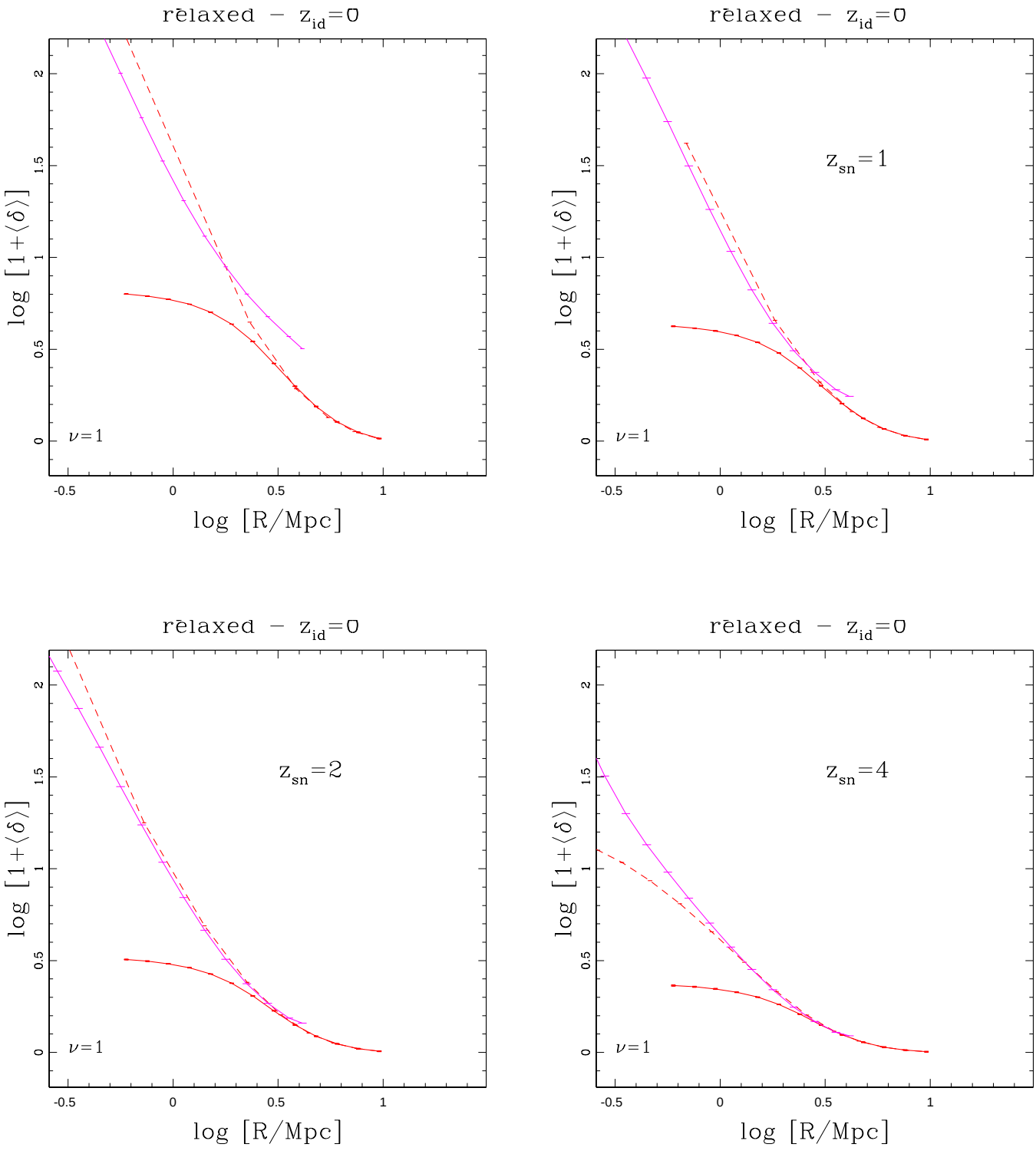

Figure 7.6: Top left panel: comparison between the outermost Eulerian profile (solid magenta curve), the non linear prediction (dashed red curve) and the linear profile (red curve) for relaxed halos of mass M_* identified at $z_{id} = 0$ ($\nu = 1$.)). The other three panels show the same comparison after tracing the halo particles back to different redshift ($z_{sn} = 1, 2, 4$) and rescaling the linear profile and non linear prediction to the same redshifts.

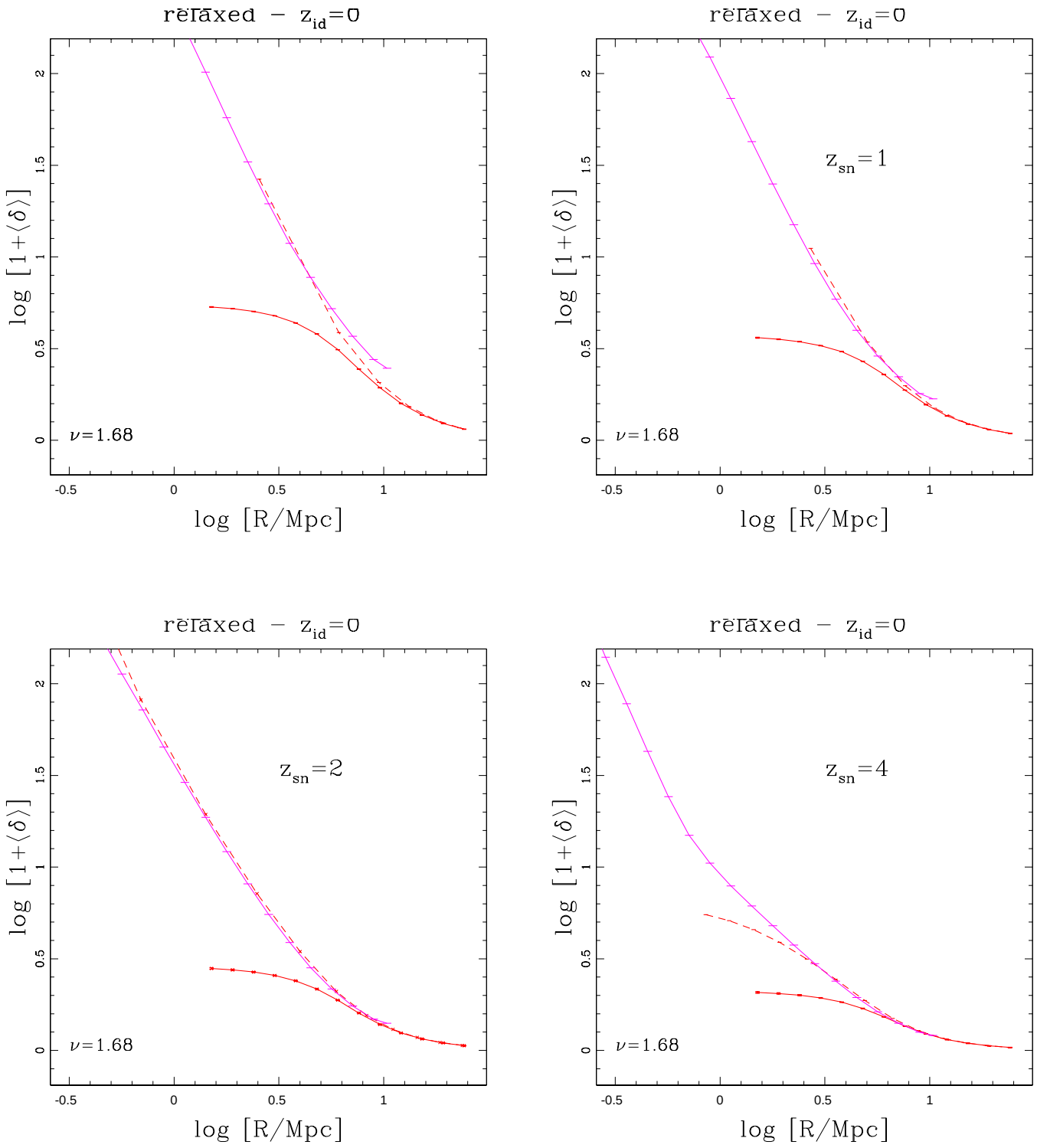

Figure 7.7: Top left panel: comparison between the outermost Eulerian profile (solid magenta curve), the non linear prediction (dashed red curve) and the linear profile (red curve) for relaxed halos of mass $16M_*$ identified at $z_{id} = 0$ ($\nu = 1.68$)). The other three panels show the same comparison after tracing the halo particles back to different redshift ($z_{sn} = 1, 2, 4$) and rescaling the linear profile and non linear prediction to the same redshifts.

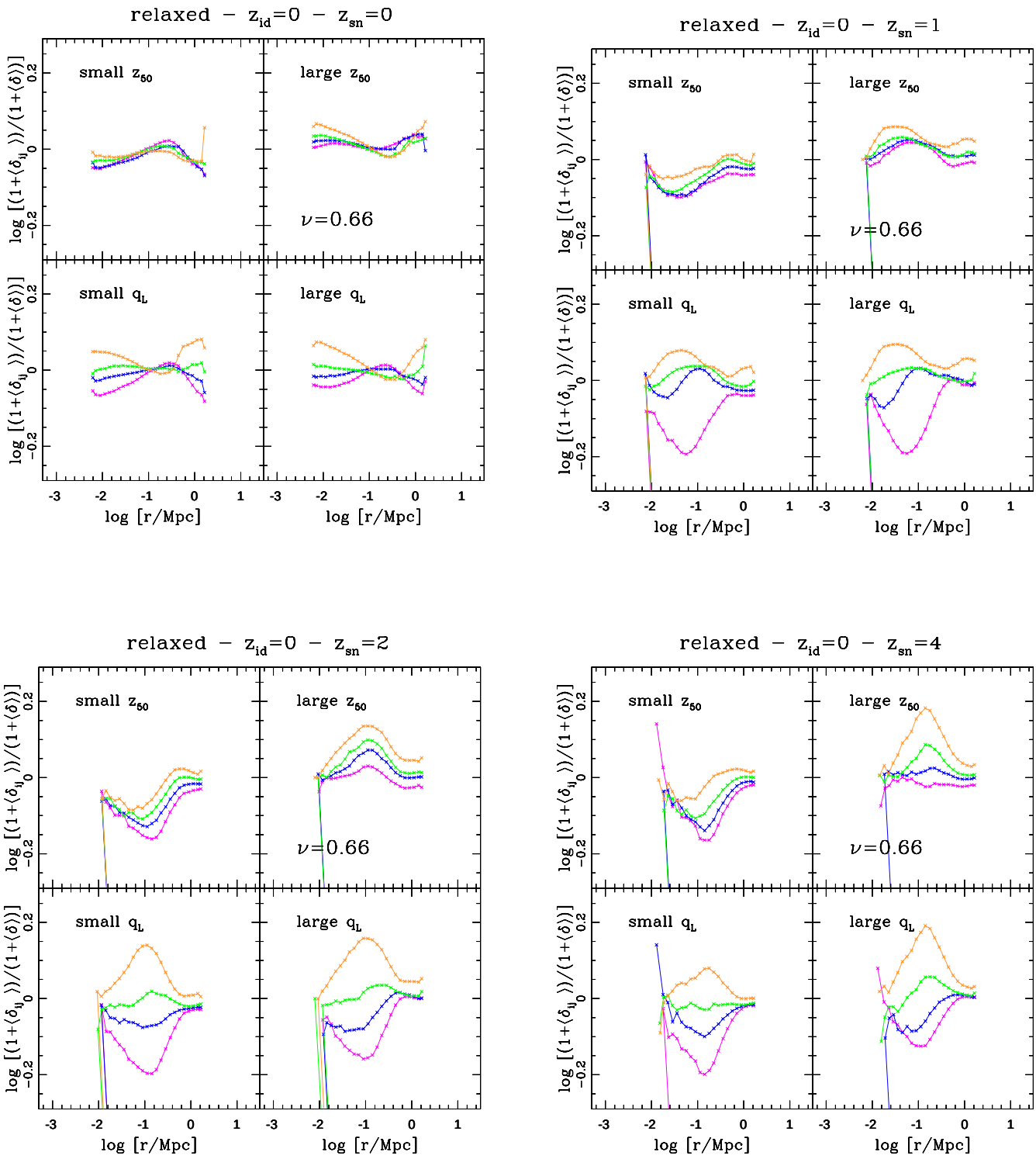

Figure 7.8: Correlations of Eulerian profiles with shear at given formation time (top panels) and with formation time at given shear (bottom panels). We consider halos with mass $M_*/16$ identified at $z_{id} = 0$. Top left plot show profiles at the identification redshifts, whereas the other plots consider particle profiles at larger z: $z_{sn} = 1$ (top right plot), $z_{sn} = 1$ (bottom left plot) and $z_{sn} = 1$ (bottom right plot).

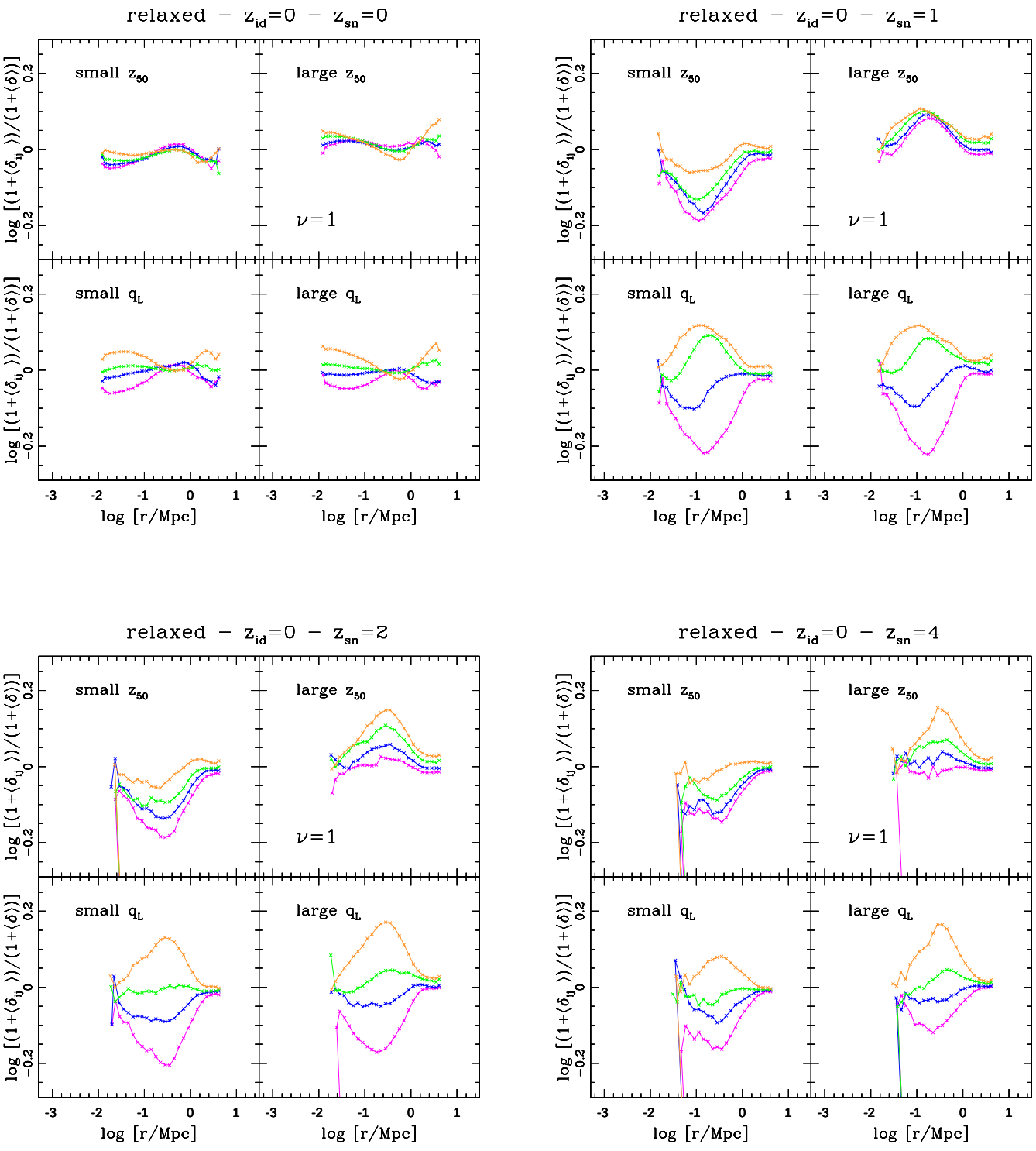

Figure 7.9: Correlations of Eulerian profiles with shear at given formation time (top panels) and with formation time at given shear (bottom panels). We consider halos with mass M_* identified at $z_{id} = 0$. Top left plot show profiles at the identification redshifts, whereas the other plots consider particle profiles at larger z: $z_{sn} = 1$ (top right plot), $z_{sn} = 1$ (bottom left plot) and $z_{sn} = 1$ (bottom right plot).

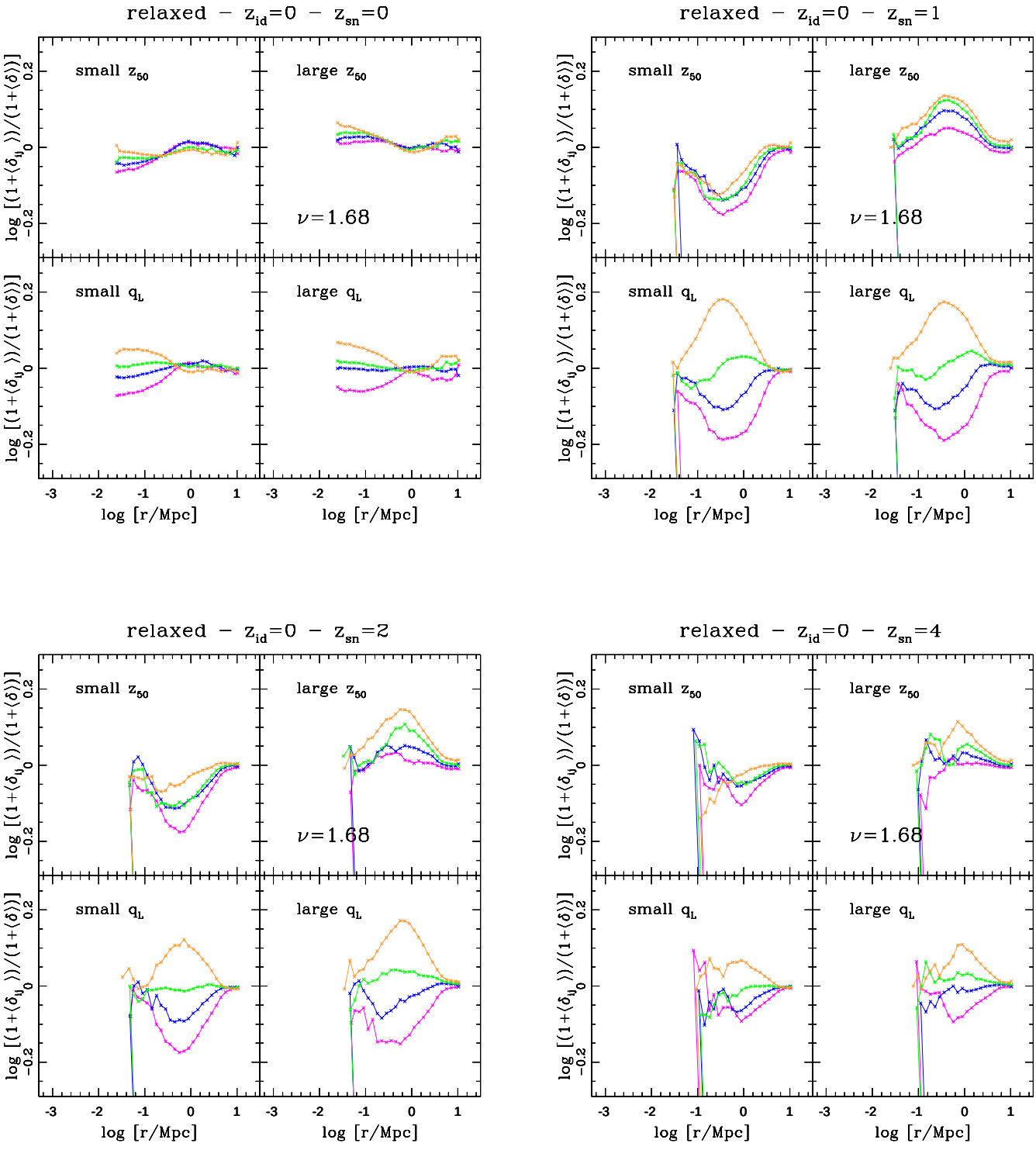

Figure 7.10: Correlations of Eulerian profiles with shear at given formation time (top panels) and with formation time at given shear (bottom panels). We consider halos with mass $16M_*$ identified at $z_{id} = 0$. Top left plot show profiles at the identification redshifts, whereas the other plots consider particle profiles at larger z: $z_{sn} = 1$ (top right plot), $z_{sn} = 1$ (bottom left plot) and $z_{sn} = 1$ (bottom right plot).

7.2.3 Evolution of the halo particle profile

In Sec. 7.2.1 we compared the outermost halo profile with non linear predictions of Bernardeau (1993). In this last section we inspect the height of the innermost halo profile at different redshift. To do so, we consider again halos identified at a certain redshift (say $z_{id} = 0$) and the halo particles traced back to higher redshifts $z_{sn} = 1$, $z_{sn} = 2$ and $z_{sn} = 4$. When plotting together profiles at different redshifts in comoving coordinates (left panel of Fig. 7.11, the central region of different profiles happens to be lower as z increases. In the standard scenario for the halo formation we know that, after a first phase of rapid growth, the density inside the central region of the halo should stay more or less unaltered. So, we plot the same profiles converting from comoving to proper coordinates:

$$r_P(z_{sn}) = \frac{R}{D(z_{sn})} \tag{7.6}$$

$$\delta_P(z_{sn}) = \delta * D(z_{sn})^3 \tag{7.7}$$

where $r_P(R)$ and δ_P (δ) are the radius and the overdensity expressed in proper (comoving) coordinates. We show the new profiles in the right panel of Fig. 7.11. We notice that the central region is very similar. However, the match is not perfect. A possible explanation is related to the locations around which halos have been built. Indeed, although the outermost profile does not depend very much on the choice of the profile center, the central region does. Fig. 7.12 shows particles of two halos identified at $z_{id} = 0$ with mass M_*. Different panels show the particles when traced back to three higher redshifts. Black dots are the halo particles, whereas red dots are the particles of the main progenitor at that redshift. We notice that, especially at high z, the location of particle center can be found in less dense region and so the central profiles can be lower than expected. We build the same profiles centering around the mass centers of main progenitors and we plot together profiles built at different redshifts. Fig. 7.13) shows that the central region of those profiles in proper coordinates superposes very well.

7.3 Summary

In this chapter we investigated both Lagrangian and Eulerian profiles and we related their properties at different redshifts. Firstly we studied the correlation of linear δ profiles with both shear and formation time and we found that, when these quantities are higher, the linear profile is higher as well. However, binning in formation time leads to a strong correlation only at scales larger than Lagrangian radius, whereas binning in q affects both small and large scales. Afterwards, we took into account halo profiles. We showed that non linear predictions of Bernardeau (1993) are consistent with the outermost halo profiles, at least when tracing the halo particle profile at high enough redshifts. We then inspected the correlation of halo profiles with Lagrangian shear and formation time. We showed that this correlation is higher when binning in z_{50}. When considering particle profiles built at higher

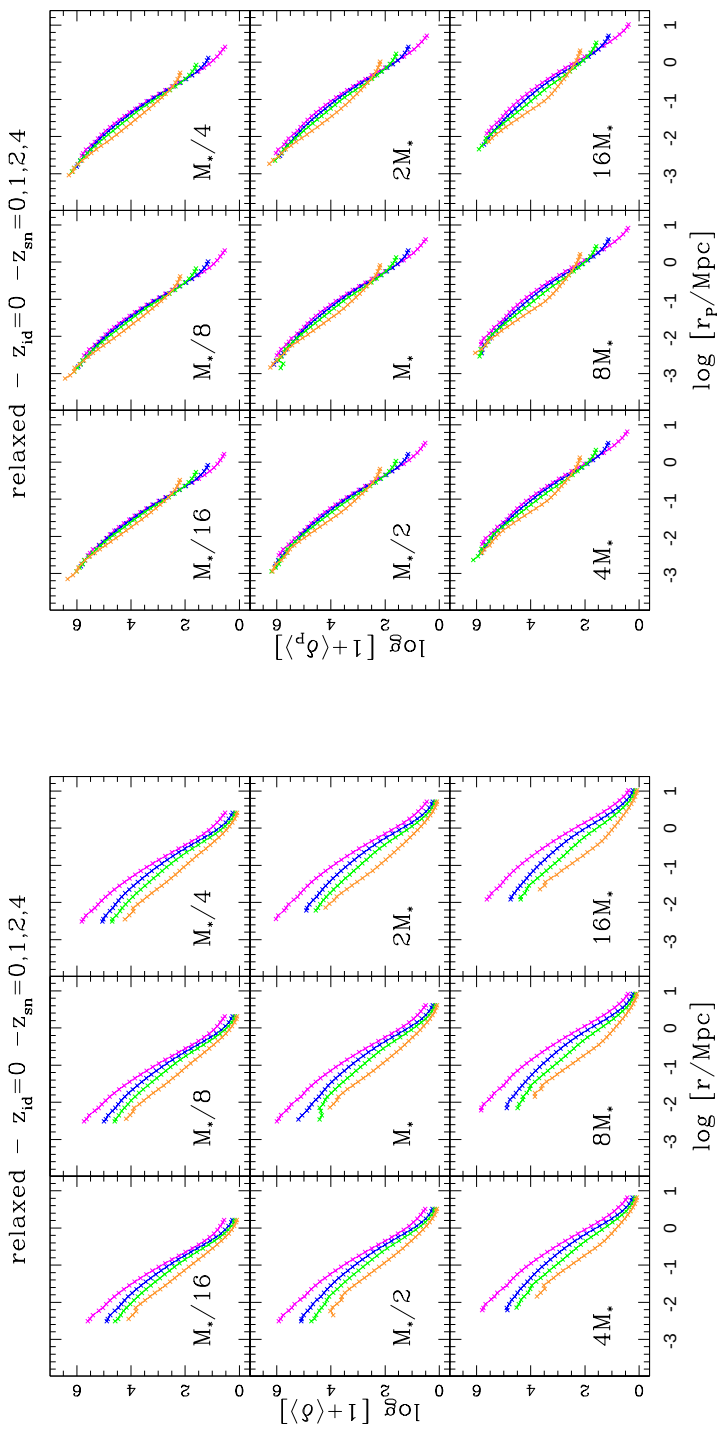

Figure 7.11: Comparison of the averaged profile for halos identified at $z_id = 0$ within nine mass bins and the profiles obtained tracing the halo particles back to $z_{sn} = 1$ (blue), $z_{sn} = 2$ (green) and $z_{sn} = 4$ (orange) and centering on the halo particle center of mass.

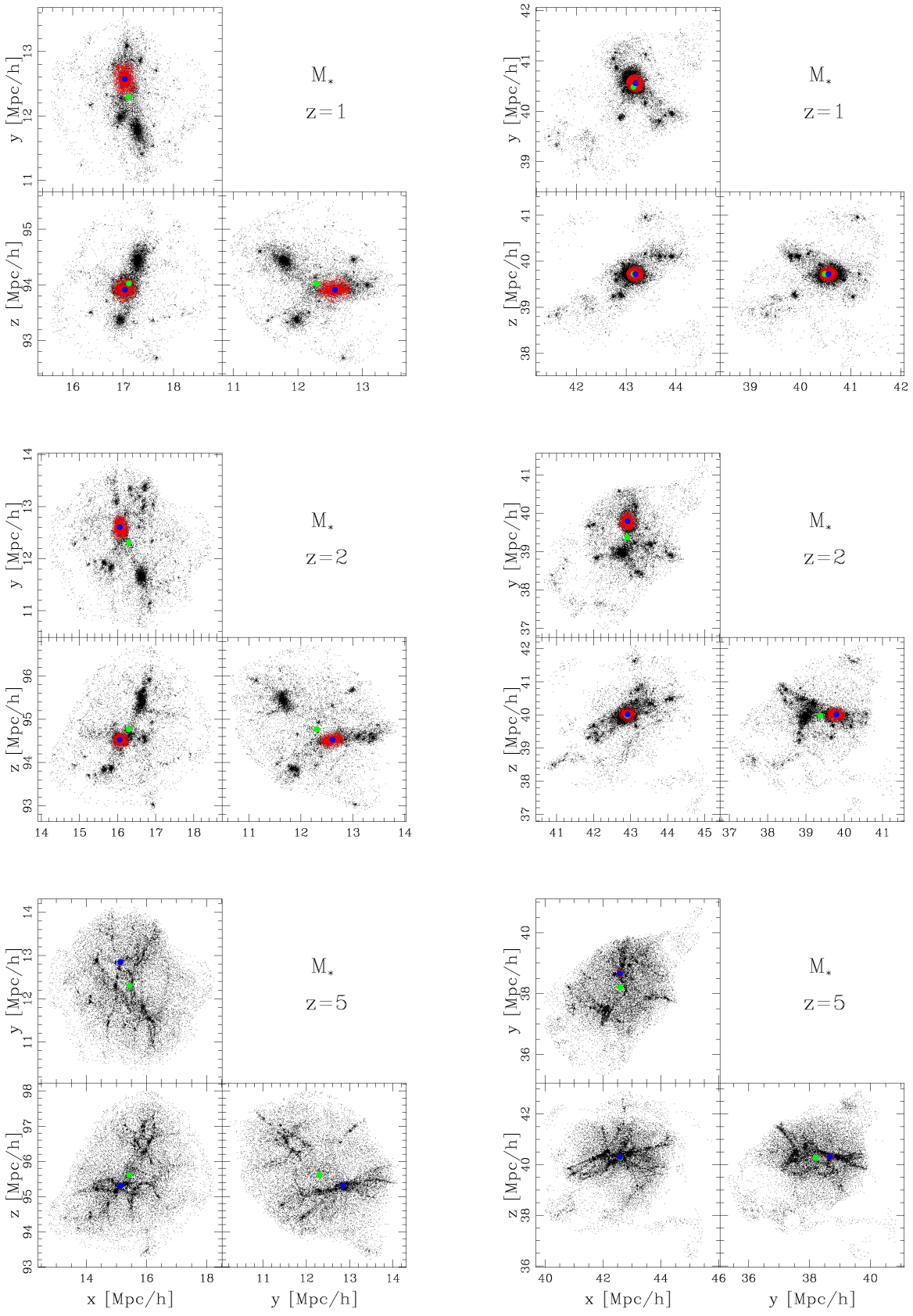

Figure 7.12: Particles of two halos identified at $z_{id} = 0$ and traced back to z_{sn}=1,2 and 5. Black dots: all particles of the halo. Red dots: all particles of the main progenitor. Green circle: mass center of the halo particles. Blue circle: mass center of the main progenitor.

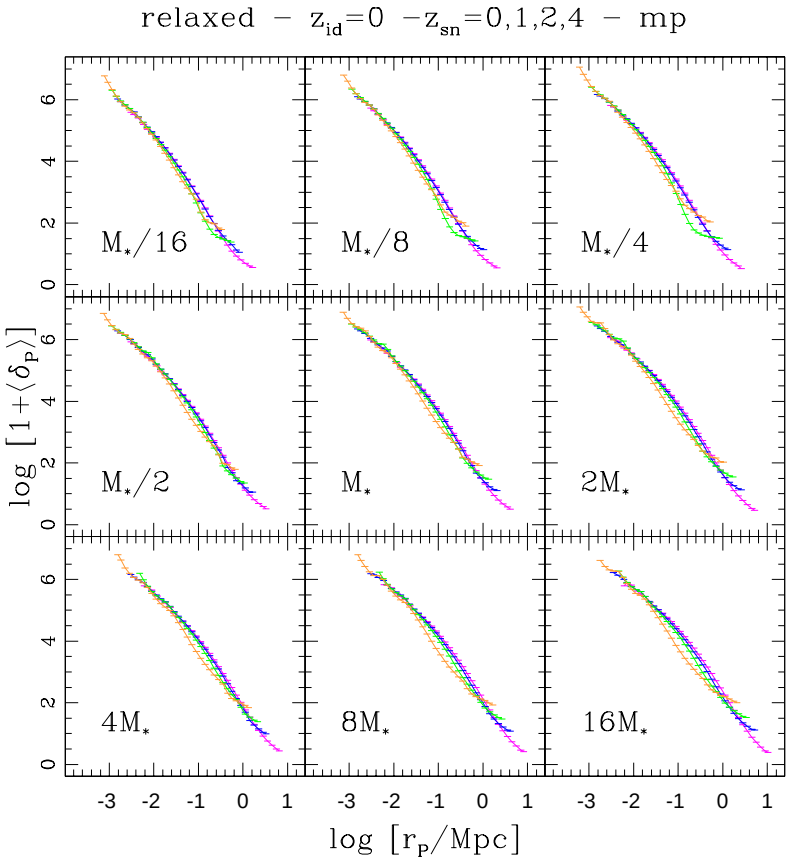

Figure 7.13: Comparison of the averaged profile for halos identified at $z_id = 0$ within nine mass bins and the profiles obtained tracing the halo particles back to $z_{sn} = 1$ (blue), $z_{sn} = 2$ (green) and $z_{sn} = 4$ (orange), and centering on the main progenitor center of mass.

redshift, the correlation is higher and in the outermost regions the dependence on q grows. Finally, we inspected the central region of halo profiles at different z. We showed that the innermost region is very similar when considering profiles in proper coordinates. However, the match is not perfect cause the center of mass of halo particles at high redshifts can be found in less dense regions. When considering halo particle profiles built around the center of mass of main progenitors, we found a better agreement.

Summary and Conclusions

The main aim of this work is the study of the formation of dark matter halos starting from initially small density fluctuations. More specifically, we would like to understand the importance of a few parameters (e.g. shear and formation time) in determining the collapse in the light of both the excursion set theory and the peaks formalism. We performed this analysis by employing Le SBARBINE cosmological set, run by the numerical cosmology group at the Physics and Astronomy Department of Padua University.

In Chapter 4 we presented the methods employed to identify halos at different redshifts, and selected appropriate mass bins and relaxation criteria. Moreover, we calculated the parameters of the ellipsoidal collapse (EC) model at the grid points of the simulation boxes, and tested that their distribution was consistent with theoretical predictions (Doroshkevich (1970), Bardeen et al. (1986), Sheth and Tormen (2002)). In the last Section of the Chapter we focused on the halo formation times. Whereas a fit for the distribution of halo formation times was provided by previous authors (Lacey and Cole (1993), Giocoli et al. (2012b)), the presence of an simpler and predictable approximation specifically tailored to dynamically relaxed halos had never been investigated. We showed that relaxed halos can be well approximated with a Weibull distribution, depending on only two free parameters. Moreover, this approximation is very good when considering, in the definition of the formation time, the mass fraction $f = 0.5$, employed in this work and often adopted in the literature. We plan to check if the same distribution can also be extended to larger identification redshifts. In following chapters we will employ these quantities in order to investigate the properties of the halo seeds.

In Chapter 5 we focussed on the characterization of the halo seeds in the framework of the peaks formalism. Moreover, we presented an alternative to the peaks approach by considering a dipole term emerging from perturbation theory. The fact that collapse generally occurs at the position of density peaks has been largely investigated (Bardeen et al. (1986), Appel and Jones (1990), Ludlow and Porciani (2011), Paranjape and Sheth

(2012), Paranjape et al. (2013)). However, attempts to merge the peaks formalism with other approaches that consider the effect of shear have a much shorter history (Castorina et al. (2016)). A good statistical understanding of the location of the halo seeds is an issue of great importance in cosmology, so it is worth investigating in this direction. We decided to analyze the correlation function between protohalo centers and peaks (valleys) identified everywhere in the simulation box, where peaks (valleys) are defined as points larger (smaller) than the six adjacent grid points. This approach not only gives us information on the well investigated correspondence between peaks and centers, but allows us to study other properties, like the clustering at large scales. We first studied the correlation function between protohalo centers and peaks and valleys of the δ distribution smoothed on the Lagrangian scale and four different scales, two smaller and two larger than the Lagrangian one: we wished to understand how this correlation changes when considering different constraints on the peak height. We found that, when no constraints on the peak height are taken into account, the correlation function at the smallest distances is maximum for smoothing scales equal to and immediately lower than the Lagrangian one, whereas the correlation function is lowest when δ is smoothed on the largest scale. As ν and the threshold increase, the correlation function in the innermost region grows for scales equal to and lower than the Lagrangian scale. In general, when imposing a weak constraint on the peak heights, the correlation function with peaks of δ smoothed on the Lagrangian scale tends to stand above the other ones when ν is large enough. Moreover, when considering a constraint that directly accounts for the value of the traceless shear when smoothed on the same scale of peaks, we selected peaks involved in the halo formation. We then calculated the same correlation function, but considering peaks and valleys on the distribution of q. Here we wanted to understand if the fact that the distribution of q has a null gradient is correlated with protohalo centers. When considering no other constraints, we found that valleys of q smoothed on scales equal or smaller than the Lagrangian one show a high correlation with protohalo centers; on the other hand the correlation with peaks of q shows the opposite behaviour. However, we showed that the value of q is more important than the fact that q is a valley or a peak. In a future work, we will study the correlation function at high distances as a function of ν and identification redshifts i.e. we will investigate more deeply the clustering of peaks at distances larger than the Lagrangian radius. Moreover, we will try to characterize more precisely the combination of δ and shear in the definition of a realistic threshold.

In the last Section of Chapter 5 we numerically investigated an alternative to the peaks approach. This alternative is justified by the fact that, whereas the SC and EC model describe collapse around the center of mass of protohalos, it is more appropriate to assume that protohalos form from regions in which the initial dipole vanishes. Whereas theoretical attempts have been made in this direction (Musso et al., to appear), this is the first time that this issue has been numerically investigated. We first considered protohalos of a certain mass and calculated the dipole of spheres in proximity of protohalo centers. When these spheres have a radius equal to the Lagrangian one, the dipole heads toward the protohalo center. If the radius is larger, the dipole has the same direction but it is

smaller. So, protohalo centers are related with points where the dipole is null. We also compared the distance of points with a minimum value of dipole and peaks of δ within a spherical region of $0.4R_L$ and we found that the dipole approach seems to better describe the location of the halo seeds. In next future, we will realize these comparisons considering a wider range of masses and following the evolution of the distance between null dipole and centers by considering halo main progenitors at different redshifts.

Whereas in Ch. 5 we searched for a characterization of the halo seeds through peaks and null dipoles, in Ch. 6 we studied the distribution of Lagrangian parameters and the correlation between some of these parameters and the formation time of halos. It is largely known the importance of shear on determining the threshold at which the collapse occurs. Predictions on an ellipsoidal collapse barrier that incorporates the effect of tidal forces have been made both by theoretical considerations (Sheth et al. (2001)) and by fitting results of numerical simulations (Sheth and Tormen (1999), Sheth et al. (2001), Sheth and Tormen (2002)). One important point is to estimate the importance of shear in the halo formation. Here we first investigated the distribution of Lagrangian parameters for halos identified at four identification redshifts and we compared the distribution for $z_{id} = 0$ with previous works. We found a good agreement with results of Robertson et al. (2009) and slightly different results respect to those of Despali et al. (2013). However, this last comparison is probably affected by the fact that they average the density over the protohalo patches, whereas Robertson et al. (2009) and us assume a spherical shape for protohalos. A more accurate measure of shear is given by the traceless shear parameter q, emerging from a perturbative analysis of the ellipsoidal collapse model. Among other things, this parameter has the advantage of not depending on a ratio of the eigenvalues. Sheth et al. (2001) found that this variate is drawn by a χ^2 with five degrees of freedom. We found that the Lagrangian traceless shear q_L is still a $\chi^2_{(5)}$, opportunely scaled. We used q_L to investigate the importance of traceless shear on determining the barrier of collapse by using a simple model with only a free parameter, q_c, growing as the importance of shear decreases. At $z_{id} = 0$ we found a trend with mass similar to that of Sheth et al. (2013), but with a slightly smaller value of q_c. When considering larger identification redshifts, the effect of shear is lower (q_c grows) and the trend with mass shows nonlinearities. Moreover, the presence of substantial scatter around the median suggests that other quantities other than q_L determine the value of δ_L (Sheth et al. (2013)). We therefore tried to understand if the correlation of δ_L with the traceless shear is direct or is due to the shear-formation time and formation time-overdensity correlations. We found that both shear (at given formation time) and formation time (at given shear) are important. More specifically, when ν is low the importance of shear is dominant, whereas at larger ν both correlation decreases and show similar values. Future improvements of these analysis should take into account the nonlinearities on the correlation of q_L.

Whereas in Ch. 5 we took into account Lagrangian quantities, in Ch. 7 we considered profiles. We wished first to understand how does the slope of the initial profile within the

Lagrangian radius correlate with its height on larger scales. This sort of correlation is often called "assembly bias", although in the context of halo formation it is also sometimes called non-local or stochastic bias. We measured the average steepness of the inner profile by measuring the different height of profiles between the innermost point and the Lagrangian radius. We found that it is nearly invariant with mass but depends slightly on the halo identification redshift (larger z_{id} means a lower steepness of the inner profile). On the contrary, the outer profile becomes steeper as ν increases. We studied how the initial shear (the stochastic variable) impacts halo concentrations and large scale bias: we considered four bins in shear and we find that both the height of profile at all scales and the steepness of the outermost profile are larger. We also find similar results when binning in the formation time, but in this case the correlation of *delta* profiles with z_{50} is larger for the innermost profile and lower for the outermost one. However, when considering Eulerian profiles we found that the correlation with formation time is always more important than the correlation with shear.

Given the importance of dark matter haloes for the understanding of the evolution of galaxies, understanding their internal structure is a fundamental issue in cosmology. Hence, at a later stage, we qualitatively analyzed the evolution with time of the the steep of the outer profile and the height of the innermost profile. We then considered halo profiles at the identification redshifts and the profiles of the halo particles once they have been traced back to three larger redshifts: $z = 1, 2$ and 4. We first compare the outermost region of Eulerian profiles with Lagrangian profiles once the have been evolved by using the nonlinear predictions o f Bernardeau (1993). We found that the outermost region of final halo profiles at high ν is consistent with non linear predictions. Moreover, we show how the accordance gets better as the redshift of the halo particle profiles grows. Finally, we considered the evolution with time of the innermost profile. Whereas the profile in comoving coordinates gets higher as the redshift decreases, when considering proper coordinates the innermost region of profile superpose very well. This rescaling is particularly accurate when we built profiles of the halo particles around the center of mass of the main progenitor. The fact that the physical size of the central region of (cumulative) density halo profiles remains the same as the halo evolves can be linked to recent theoretical insights about the formation of halos. Moreover, the profile evolution could be related to the evolution of concentration in different phases e.g. the first phase of rapid growth in the matter-dominated epoch and the second phase of slow growth in the dark energy epoch (Lu et al. (2006), Ludlow et al. (2013) and Correa et al. (2015)).

APPENDIX A

Tables: halo statistics

Number of halos															
z_{id}=0															
	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
ν	0.53	0.57	0.62	0.67	0.73	0.81	0.90	1.00	1.13	1.28	1.46	1.69	1.97	2.32	2.77
Ada	11770	6764	3655	1978	1109	580	334	180	96						
Bice				14550	8244	4409	2489	1236	703	350	172	73			
Cloe							17885	10066	5249	2696	1376	577	272	105	24
Dora										19505	10540	4982	2013	668	174
TOT	11770	6764	3655	16528	9353	4989	20708	11482	6048	22551	12088	5632	2285	773	198
z_{id}=1															
ν	0.88	0.95	1.03	1.11	1.21	1.34	1.49	1.65	1.86	2.11	2.41	2.79	3.25	3.83	4.58
Ada	13707	7489	3875	2083	1082	575	268	127	58						
Bice				14864	8151	4239	2090	990	422	172	69	22			
Cloe							15444	7828	3537	1426	511	180	35	2	0
Dora										11457	4249	1232	247	30	3
TOTAL	13707	7489	3875	16947	9233	4814	17802	8945	4017	13055	4829	1434	282	32	3
z_{id}=2															
ν	1.29	1.39	1.50	1.62	1.76	1.95	2.17	2.41	2.73	3.09	3.52	4.08			
Ada	12925	6742	3494	1603	754	321	127	59	17						
Bice				12515	5996	2761	1138	428	140	37	11	11			
Cloe							9134	3448	1076	333	58	4			
Dora										2347	454	58			
TOTAL	12925	6742	3494	14118	6750	3082	10399	3935	1233	2717	523	73			
z_{id}=4															
ν	2.09	2.28	2.49	2.71	2.95	3.27	3.64	4.04	4.57						
Ada	6952	2985	1128	390	120	47	5	2	0						
Bice				3456	1093	292	71	7	0						
Cloe							515 87	9							
TOTAL	6952	2985	1128	3846	1213	339	591	95	9						

Table A.1: Number of halos identified at z_{id} = 0,1,2 and 4 in the nine mass bins mostly employed in this work and in six additional mass bins.

Number of relaxed halos															
z_{id}=0															
	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
ν	0.53	0.57	0.62	0.67	0.73	0.81	0.90	1.00	1.13	1.28	1.46	1.69	1.97	2.32	2.77
Ada	8566	5010	2730	1493	850	444	237	50	11						
Bice			11091	6336	3344	1906	965	561	238	59	8				
Cloe						14034	7889	4058	2066	1050	433	179	35	1	
Dora										15489	8192	3837	1483	486	119
TOT	8566	5010	2730	12584	7186	3788	16177	8904	4619	17793	9301	4278	1662	521	120
z_{id}=1															
ν	0.88	0.95	1.03	1.11	1.21	1.34	1.49	1.65	1.86	2.11	2.41	2.79	3.25	3.83	4.58
Ada	9682	5376	2805	1487	752	394	190	49	10						
Bice			10687	5836	3099	1467	710	296	107	35	3				
Cloe							11190	5617	2483	1011	351	131	22	1	0
Dora										7742	3000	891	175	22	2
ABCD	9682	5376	2805	12174	6588	3493	12847	6376	2789	8860	3386	1025	197	23	2
z_{id}=2															
ν	1.29	1.39	1.50	1.62	1.76	1.95	2.17	2.41	2.73	3.09	3.52	4.08	4.75	5.60	6.68
Ada	8889	4612	2463	1109	511	208	81	22	5						
Bice				8346	4136	1940	780	298	104	30	4	0			
Cloe							6230	2412	755	234	37	2			
Dora										1606	329	41			
ABCD	8889	4612	2463	9455	4647	2148	7091	2732	859	1870	370	43			
z_{id}=4															
ν	2.09	2.28	2.49	2.71	2.95	3.27	3.64	4.04	4.57						
Ada	4703	2092	789	265	87	25	1	0	0						
Bice				2390	784	201	49	6	0						
Cloe							373	67	7						
ABCD	4703	2092	789	2655	871	226	422	73	7						

Table A.2: Number of relaxed halos identified at $z_{id} = 0, 1, 2$ and 4 in the nine mass bins mostly employed in this work and in six additional mass bins.

APPENDIX B

Tables: statistics of eigenvalues and EC parameters

Variance of the elements of the deformation tensor.																	
	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$		
R [Mpc]	0.30	0.38	0.47	0.60	0.75	0.95	1.2	1.51	1.90	2.40	3.02	3.80	4.79	6.04	7.602	4.790	12.07
xdx																	
Ada	3.43	2.97	2.56	2.18	1.85	1.55	1.28	1.05	0.84	0.67	0.52						
Bice				1.78	1.52	1.29	1.07	0.88	0.72	0.58	0.46	0.36	0.28	0.22			
Cloe							1.04	0.86	0.72	0.57	0.45	0.35	0.27	0.21	0.16	0.11	0.083
Dora										0.54	0.43	0.34	0.26	0.20	0.14	0.10	0.074
ydy																	
Ada	3.33	2.88	2.47	2.09	1.76	1.47	1.21	0.98	0.79	0.63	0.49						
Bice				1.77	1.50	1.27	1.05	0.86	0.70	0.56	0.44	0.35	0.27	0.20			
Cloe						1.04	0.86	0.70	0.56	0.45	0.35	0.27	0.20	0.15	0.11	0.080	
Dora									0.54	0.43	0.34	0.26	0.19	0.14	0.10	0.074	
zdz																	
Ada	3.34	2.90	2.49	2.11	1.78	1.48	1.22	0.99	0.79	0.63	0.49						
Bice				1.77	1.50	1.28	1.06	0.86	0.70	0.60	0.44	0.34	0.26	0.20			
Cloe							1.03	0.85	0.69	0.56	0.44	0.34	0.26	0.20	0.15	0.10	0.072
Dora										0.54	0.43	0.33	0.26	0.19	0.14	0.10	0.073
XY																	
Ada	1.12	0.97	0.83	0.71	0.60	0.50	0.41	0.33	0.27	0.21	0.16						
Bice				0.59	0.50	0.42	0.35	0.28	0.23	0.18	0.14	0.11	0.08	0.061			
Cloe						0.34	0.28	0.23	0.18	0.15	0.11	0.087	0.066	0.49	0.36	0.26	
Dora										0.18	0.14	0.11	0.086	0.065	0.048	0.035	0.024
XZ																	
Ada	1.12	0.97	0.83	0.71	0.60	0.50	0.40	0.33	0.26	0.21	0.16						
Bice				0.59	0.50	0.43	0.36	0.29	0.24	0.19	0.15	0.12	0.090	0.067			
Cloe							0.34	0.28	0.23	0.18	0.15	0.11	0.087	0.065	0.047	0.034	0.023
Dora										0.18	0.14	0.11	0.086	0.065	0.048	0.034	0.024
YZ																	
Ada	1.11	0.96	0.83	0.70	0.59	0.49	0.40	0.33	0.26	0.21	0.16						
Bice				0.59	0.50	0.42	0.35	0.29	0.23	0.18	0.14	0.11	0.083	0.062			
Cloe							0.34	0.28	0.23	0.19	0.15	0.11	0.089	0.066	0.049	0.035	0.025
Dora										0.18	0.14	0.11	0.086	0.065	0.048	0.034	0.024

Table B.1: Variance of the elements of the three diagonal elements of the deformation tensor plus the averaged symmetric elements for all four simulations. We show the values obtained smoothing on nine different scales with a top-hat filter.

MEAN

							λ_1									
R [Mpc]	0.30	0.38	0.47	0.60	0.75	0.95	1.2	1.51	1.90	2.40	3.02	3.80	4.79	6.04	7.60	4.79	12.07
Ada	2.57	2.40	2.23	2.07	1.91	1.75	1.60	1.45	1.31	1.17	1.04	0.92					
Bice			1.59	1.47	1.35	1.23	1.11	1.00	0.90								
Cloe				1.22	1.10	1.00	0.89	0.80	0.70	0.62	0.54	0.46					
Dora				0.98	0.88	0.78	0.69	0.61	0.53	0.45	0.38	0.32					
							λ_2 - units of 10^{-3}										
Ada	2.78	1.59	0.354	-0.891	-2.00	-2.85	-3.47	-4.15									
Bice				-0.29	-0.34	0.67	0.68	-0.47									
Cloe			0.55	0.54	0.71	0.91	1.1	1.2	1.11	0.74	0.014						
Dora			0.097	0.12	0.13	0.16	0.22	0.32	0.41	0.47	0.50						
							λ_3										
Ada	-2.57	-2.40	-2.24	-2.07	-1.91	-1.75	-1.60	-1.45	-1.31	-1.17	-1.04	-0.92					
Bice			-1.59	-1.46	-1.35	-1.23	-1.11	-1.0	-0.90 -0.60	-0.52	-0.36						
Cloe				-1.22	-1.11	-1.00	-0.89	-0.80	-0.70	-0.62	-0.54	-0.46					
Dora				-0.98	-0.88	-0.79	-0.69	-0.52	-0.45	-0.39	-0.32						

VARIANCE

								λ_1									
R [Mpc]	0.30	0.38	0.47	0.60	0.75	0.95	1.2	1.51	1.90	2.40	3.02	3.80	4.79	6.04	7.602	4.790	12.07
Ada	2.79	2.45	2.12	1.83	1.56	1.31	1.10	0.904									
Bice				1.30	1.10	0.94	0.78	0.64	0.52	0.41							
Cloe			0.76	0.63	0.51	0.41	0.33	0.25	0.20	0.15	0.11						
Dora				0.49	0.39	0.32	0.25	0.19	0.14	0.11	0.076	0.054					
								λ_2									
Ada	2.62	2.29	1.98	1.701	1.45	1.22	1.01	0.836									
Bice					1.18	1.00	0.85	0.70	0.58	0.47	0.37	0.31	0.18	0.099			
Cloe			0.69	0.57	0.46	0.37	0.29	0.23	0.18	0.13	0.097						
Dora				0.45	0.36	0.29	0.22	0.17	0.13	0.096	0.069	0.049					
								λ_3									
Ada	2.99	2.61	2.25	1.92	1.63	1.37	1.14	0.942									
Bice				1.30	1.10	0.94	0.77	0.63	0.51	0.41							
Cloe							0.76	0.63	0.51	0.41	0.33	0.26	0.20	0.15			
Dora				0.49	0.40	0.32	0.25	0.19	0.14	0.10	0.076	0.054					

Table B.2: Mean and variance of the eigenvalues of the deformation tensor.

relative frequence of different triplets combinations				
	$\lambda_1, \lambda_2, \lambda_3 < 0$	$\lambda_1, \lambda_2 < 0; \lambda_3 > 0$	$\lambda_1 < 0; \lambda_2, \lambda_3 > 0$	$\lambda_1, \lambda_2, \lambda_3 > 0$
$R(M_*/128)$				
Ada	0.07754	0.4203	0.4174	0.08200
$R(M_*/64)$				
Ada	0.07784	0.4217	0.4176	0.8321
$R(M_*/32)$				
Ada	0.7811	0.4224	0.4179	0.8013
$R(M_*/16)$				
Ada	0.07830	0.4182	0.4203	0.08200
Bice	0.08000	0.4203	0.4200	0.07961
$R(M_*/8)$				
Ada	0.07848	0.4173	0.4184	0.08012
Bice	0.0800	0.4201	0.4193	0.0800
$R(M_*/4)$				
Ada	0.07868	0.4184	0.4201	0.08002
Bice	0.08013	0.4200	0.4202	0.07958
$R(M_*/2)$				
Ada	0.07875	0.4183	0.4201	0.08135
Bice	0.08035	0.4198	0.4202	0.07966
Cloe	0.07994	0.4200	0.4202	0.07994
$R(M_*)$				
Ada	0.08002	0.4201	0.4181	0.07895
Bice	0.08010	0.4191	0.4211	0.07902
Cloe	0.07996	0.4198	0.4201	0.08021
$R(2M_*)$				
Ada	0.08201	0.4197	0.4203	0.0798
Bice	0.08103	0.4173	0.4201	0.07982
Cloe	0.08002	0.4196	0.4198	0.08049
$R(4M_*)$				
Bice	0.07903	0.4213	0.4202	0.07972
Cloe	0.08017	0.4196	0.4195	0.08072
Dora	0.07966	0.4205	0.4202	0.07964
$R(8M_*)$				
Bice	0.07963	0.4213	0.4214	0.07962
Cloe	0.08037	0.4196	0.4190	0.08100
Dora	0.07953	0.4206	0.4202	0.07965
$R(16M_*)$				
Bice	0.08023	0.4173	0.4205	0.08052
Cloe	0.08088	0.4192	0.4186	0.08127
Dora	0.07946	0.4205	0.4206	0.07949
$R(32M_*)$				
Cloe	0.08150	0.4199	0.4169	0.08175
Dora	0.07913	0.4208	0.4210	0.07919
$R(64M_*)$				
Cloe	0.08171	0.4239	0.4202	0.08011
Dora	0.07910	0.4218	0.4220	0.07921
$R(128M_*)$				
Cloe	0.08073	0.4203	0.4209	0.07911
Dora	0.07910	0.4218	0.4220	0.07921
expected values	0.08	0.42	0.42	0.08

Table B.3: Comparison between relative frequence of different combination of eigenvalues and the values presented in literature. There is clearly a good correspondance.

APPENDIX C

Tables: statistics of peaks

C.1 Peaks and valleys of δ

Table C.1: Number of peaks and valleys of δ

Number of peaks: δ

	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Ada	335431	203862	125398	77963	48289	30045	19125	12219	8666						
Bice				754708	476393	310088	195695	125758	82570	55905	37447	26775			
Cloe							375286	241345	156934	103108	68779	47304	32602	23257	17713
Dora										436105	291260	198486	139213	99973	72951

Number of valleys: δ

	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Ada	331129	201489	123495	76393	47224	29964	18975	12196	8070						
Bice				752447	477964	311915	195925	126030	82960	55224	36737	26392			
Cloe							376729	240617	157687	103866	69744	47862	33150	24279	18109
Dora										434407	290805	199073	140486	100593	74397

Number of peaks of $\delta > \delta_c(z_{id} = 0)$

	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Ada	221095	128730	75037	44257	25456	14368	8100	4420	2594						
Bice				399038	230824	134936	76641	42883	22817	11953	6047	2491			
Cloe							199363	112604	62733	33743	16936	7898	3363	1044	274
Dora										167039	84454	39939	16231	5180	1257

Number of peaks of $\delta > 2\delta_c(z_{id} = 0)$

	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Ada	150055	81527	43340	22823	11453	5329	2269	828	335						
Bice				201533	101881	49956	22606	9340	3436	1153	405	33			
Cloe							70724	30476	11566	3711	963	127	8	0	0
Dora										19899	4957	827	79	3	0

Number of peaks of $\delta > 3\delta_c(z_{id} = 0)$

	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Ada	82383	39974	18186	7858	3072	1086	319	92	8						
Bice				71682	28952	10871	3236	867	181	6	0	0			
Cloe							17885	12155	3323	657	76	3	0	0	0
Dora										482	26	0	0	0	0

Table C.2: Number of peaks of q

Number of peaks of q

	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Ada	703525	423911	256636	15521	95741	59531	37368	23800	15222						
Bice				1199195	744869	477472	293745	185539	117525	77701	51884	34229			
Cloe							750238	468738	297463	188165	123031	81967			
Dora										876486	567188	374105	251165	172279	121234

Number of peaks of q with $\delta > \delta_c(z_{id} = 0)$

	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Ada	213780	123747	6	44257	25456	14368	8100	4420	2594						
Bice				754708	476393	310088	195695	125758	82570	55905	37447	26775			
Cloe							375286	241345	156934	103108	68779	47304	32602	23257	17713
Dora										436105	291260	198486	139213	99973	72951

Number of peaks of q with $\delta > 2\delta_c(z_{id} = 0)$

	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Ada	221095	128730	75037	44257	25456	14368	8100	4420	2594						
Bice				754708	476393	310088	195695	125758	82570	55905	37447	26775			
Cloe							375286	241345	156934	103108	68779	47304	32602	23257	17713
Dora										436105	291260	198486	139213	99973	72951

Number of peaks of q with $\delta > 3\delta_c(z_{id} = 0)$

	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Ada	82383	39974	18186	7858	3072	1086	319	92	8						
Bice				754708	476393	310088	195695	125758	82570	55905	37447	26775			
Cloe							375286	241345	156934	103108	68779	47304	32602	23257	17713
Dora										436105	291260	198486	139213	99973	72951

Table C.3: Number of valleys of q

Number of peaks of q														
	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Ada	804174	475962	283004	170267	101638	61504	38859	24043	15080						
Bice				1300516	794037	494454	299557	186983	117467	75268	49228	33469			
Cloe							954448	579659	354865	219975	138983	88462			
Dora										1162798	725161	458027	293662	193188	131152
Number of valleys of q with $\delta > \delta_c(z_{id} = 0)$															
	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Ada	804174	475962	283004	170267	101638	61504	38859	24043	15080						
Bice				754708	476393	310088	195695	125758	82570	55905	37447	26775			
Cloe							375286	241345	156934	103108	68779	47304	32602	23257	17713
Dora										436105	291260	198486	139213	99973	72951
Number of peaks of q with $\delta > 2\delta_c(z_{id} = 0)$															
	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$
Bice				117543	55731	25502	10570	4011	1411	466	123	9			
Cloe							32890	12280	4144	1105	175	34	0	0	0
Dora										5372	1140	129	9	0	0
Number of peaks of q with $\delta > 3\delta_c(z_{id} = 0)$															
	$M_*/128$	$M_*/64$	$M_*/32$	$M_*/16$	$M_*/8$	$M_*/4$	$M_*/2$	M_*	$2M_*$	$4M_*$	$8M_*$	$16M_*$	$32M_*$	$64M_*$	$128M_*$

APPENDIX D

Figures: correlation functions

D.1 Protohalo centers and peaks of δ

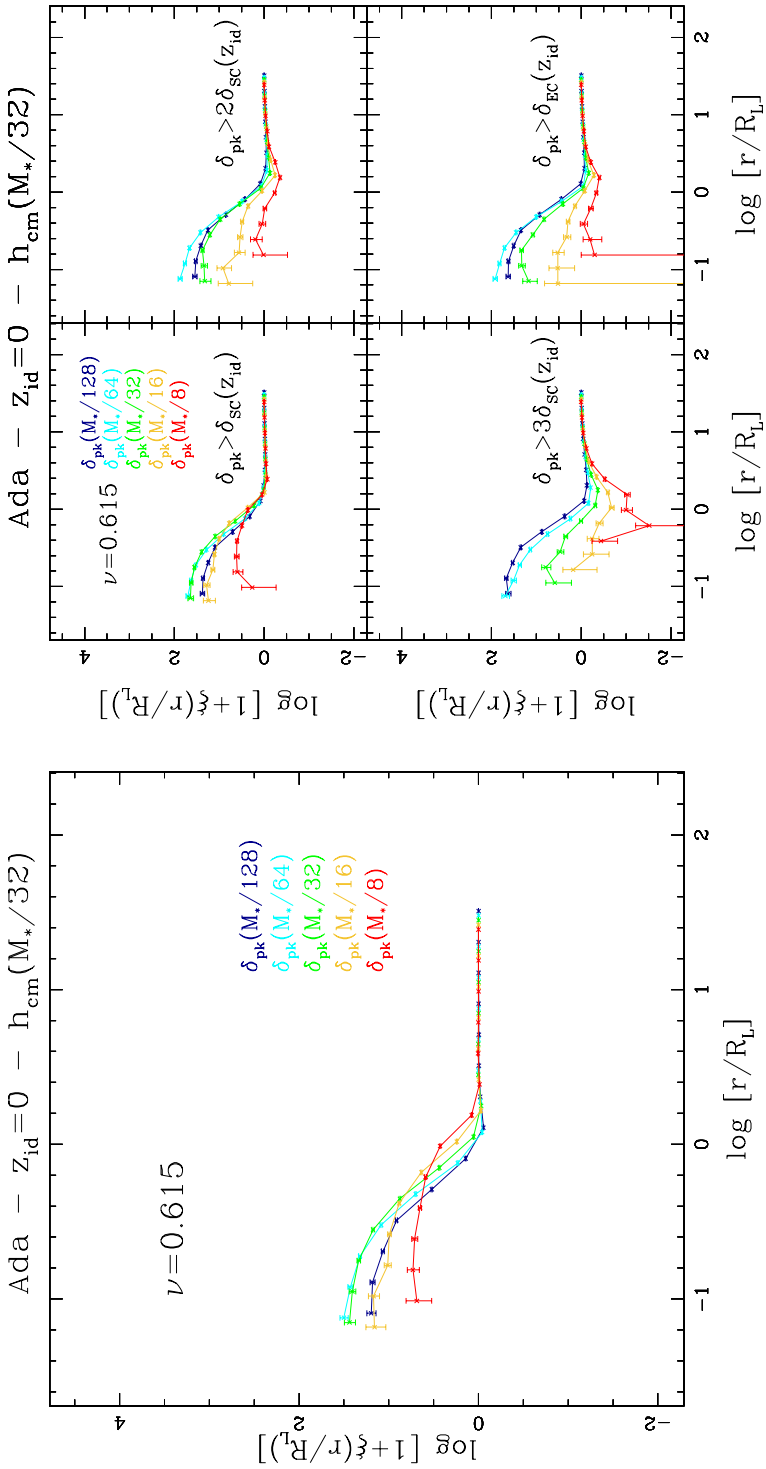

Figure D.1: Correlation functions between centers of mass of protohalos with $\nu = 0.615$ and δ peaks on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 0$. Halos have been identified with Ada simulation. Left figure: all peaks; right figure: peaks larger than four different thresholds.

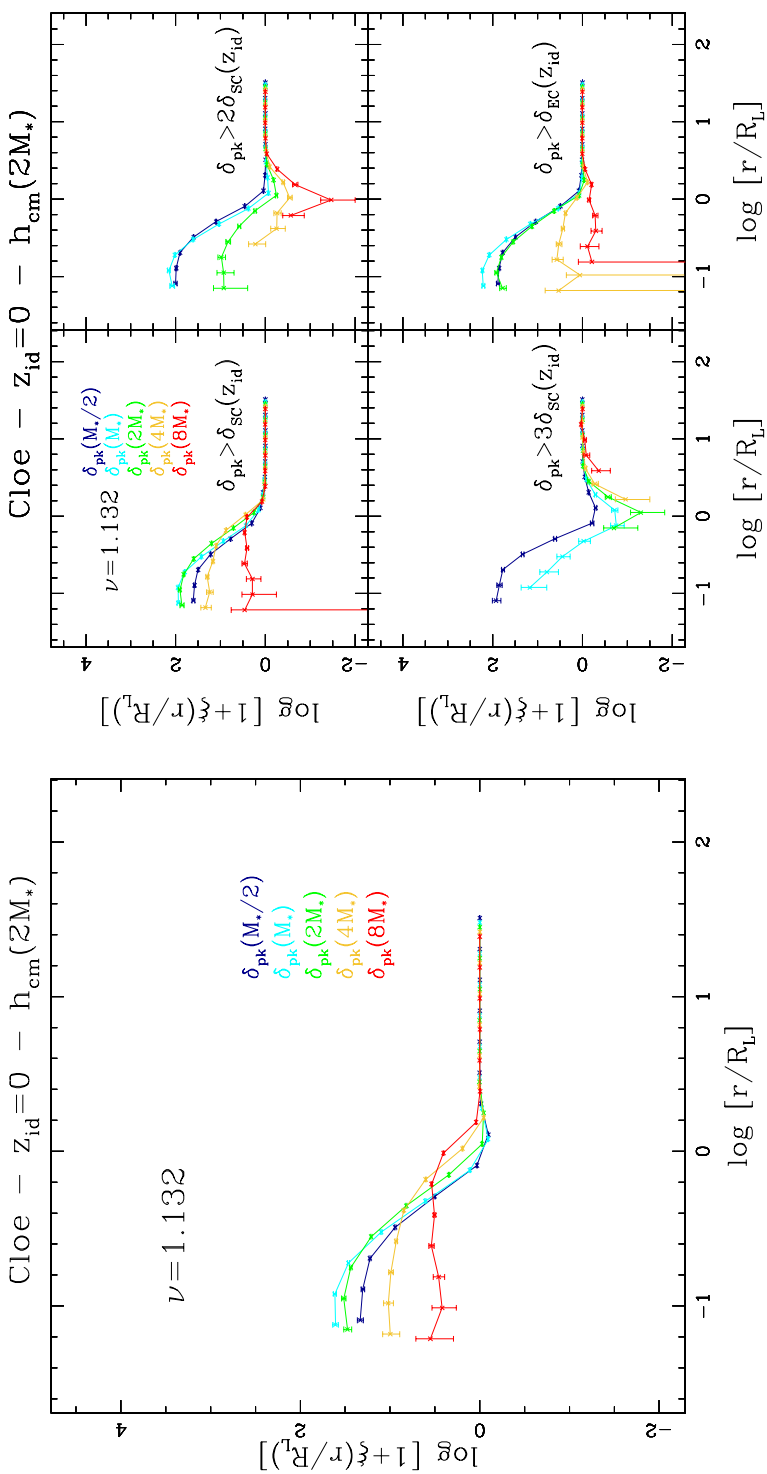

Figure D.2: Correlation functions between centers of mass of protohalos with $\nu = 1.132$ and δ peaks on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 0$. Halos have been identified with Bice simulation. Left figure: all peaks; right figure: peaks larger than four different thresholds.

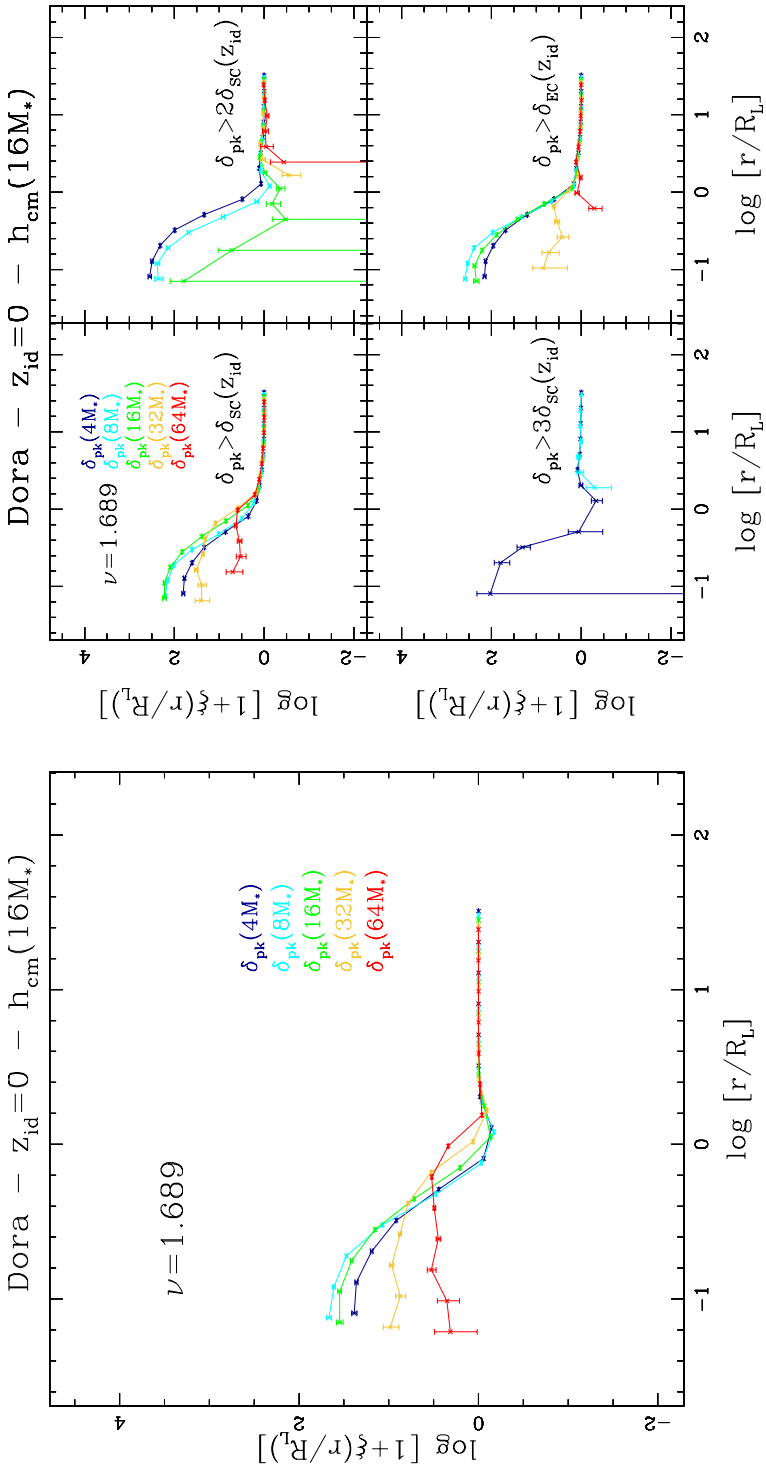

Figure D.3: Correlation functions between centers of mass of protohalos with $\nu = 1.689$ and δ peaks on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 0$. Halos have been identified with Dora simulation. Left figure: all peaks; right figure: peaks larger than four different thresholds.

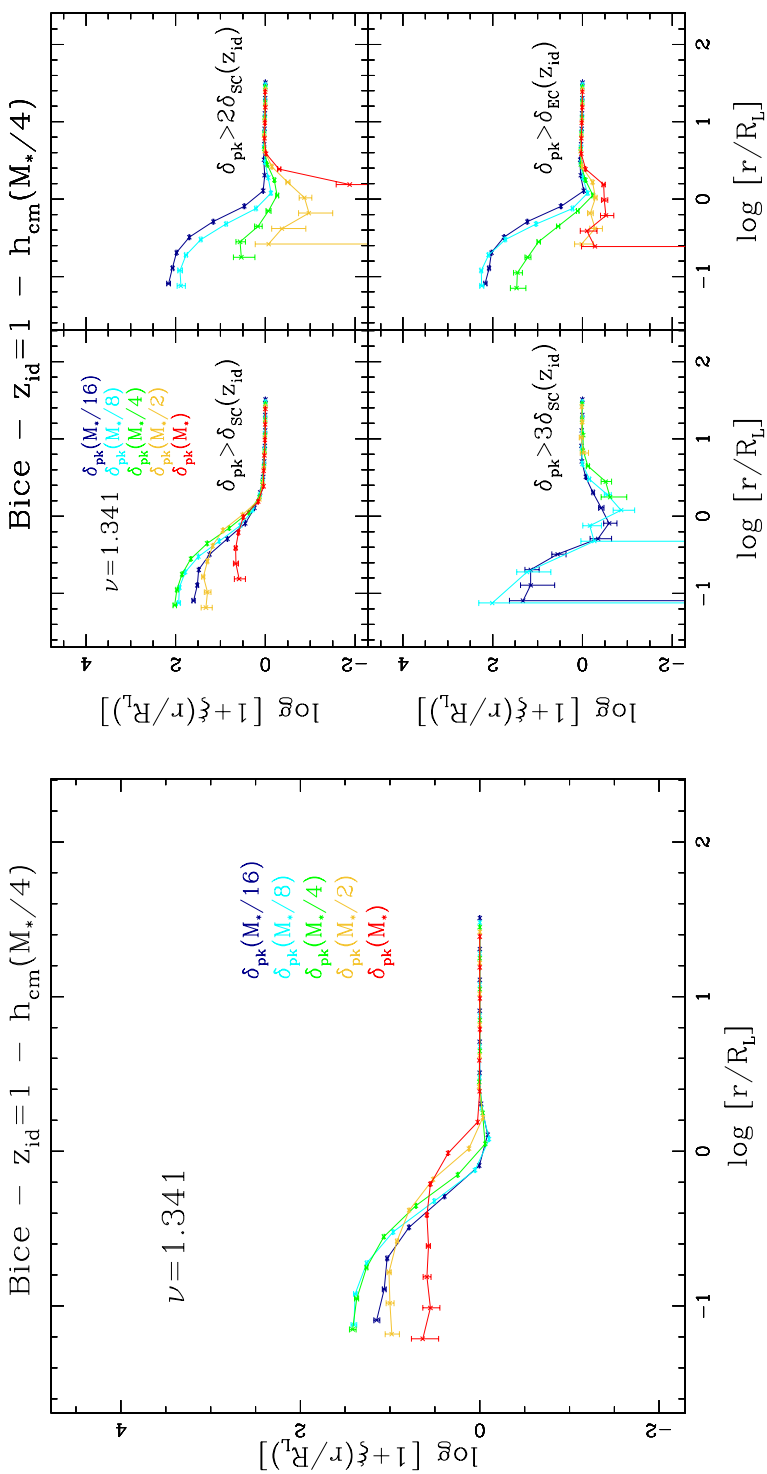

Figure D.4: Correlation functions between centers of mass of protohalos with $\nu = 1.341$ and δ peaks on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 1$. Halos have been identified with Dora simulation. Left figure: all peaks; right figure: peaks larger than four different thresholds.

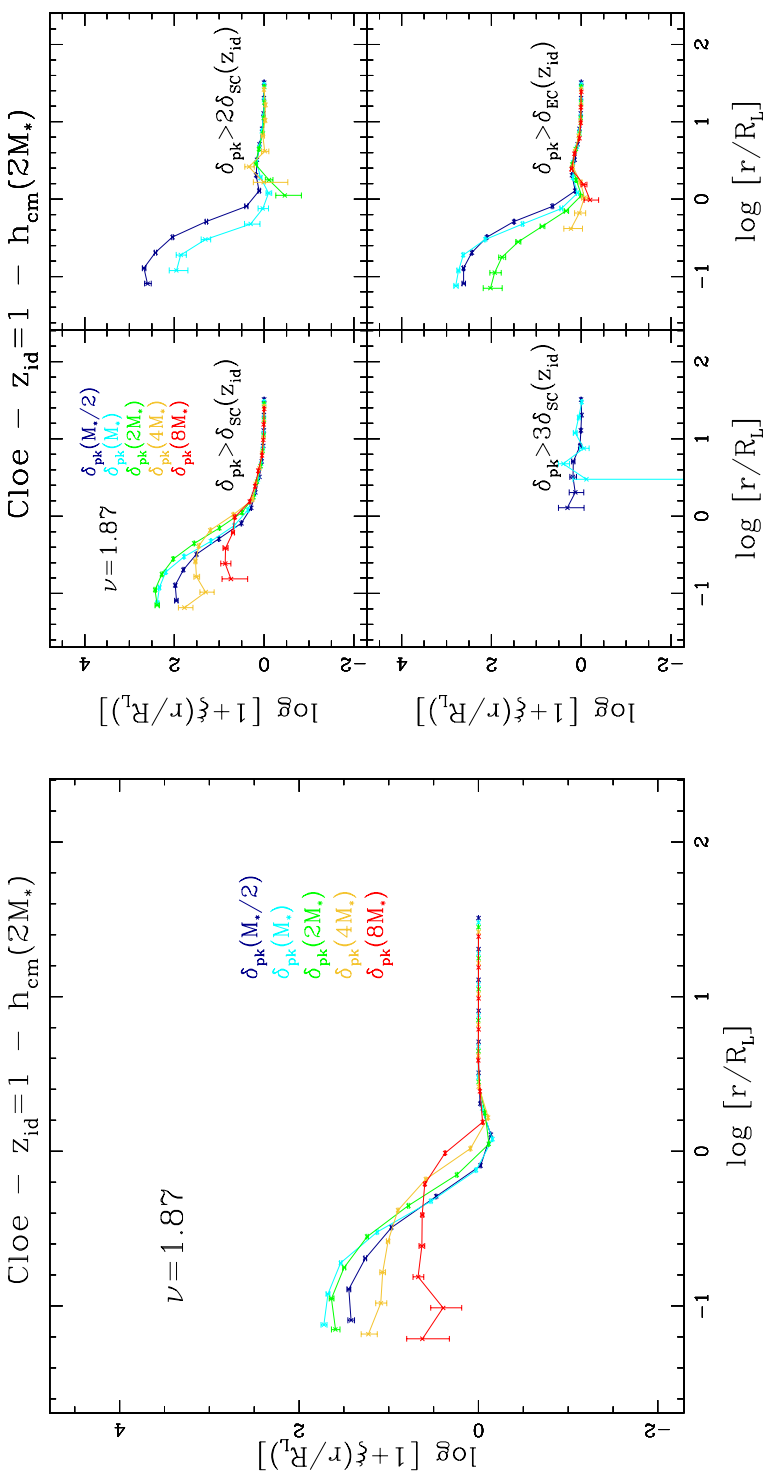

Figure D.5: Correlation functions between centers of mass of protohalos with $\nu = 1.87$ and δ peaks on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 1$. Halos have been identified with Cloe simulation. Left figure: all peaks; right figure: peaks larger than four different thresholds.

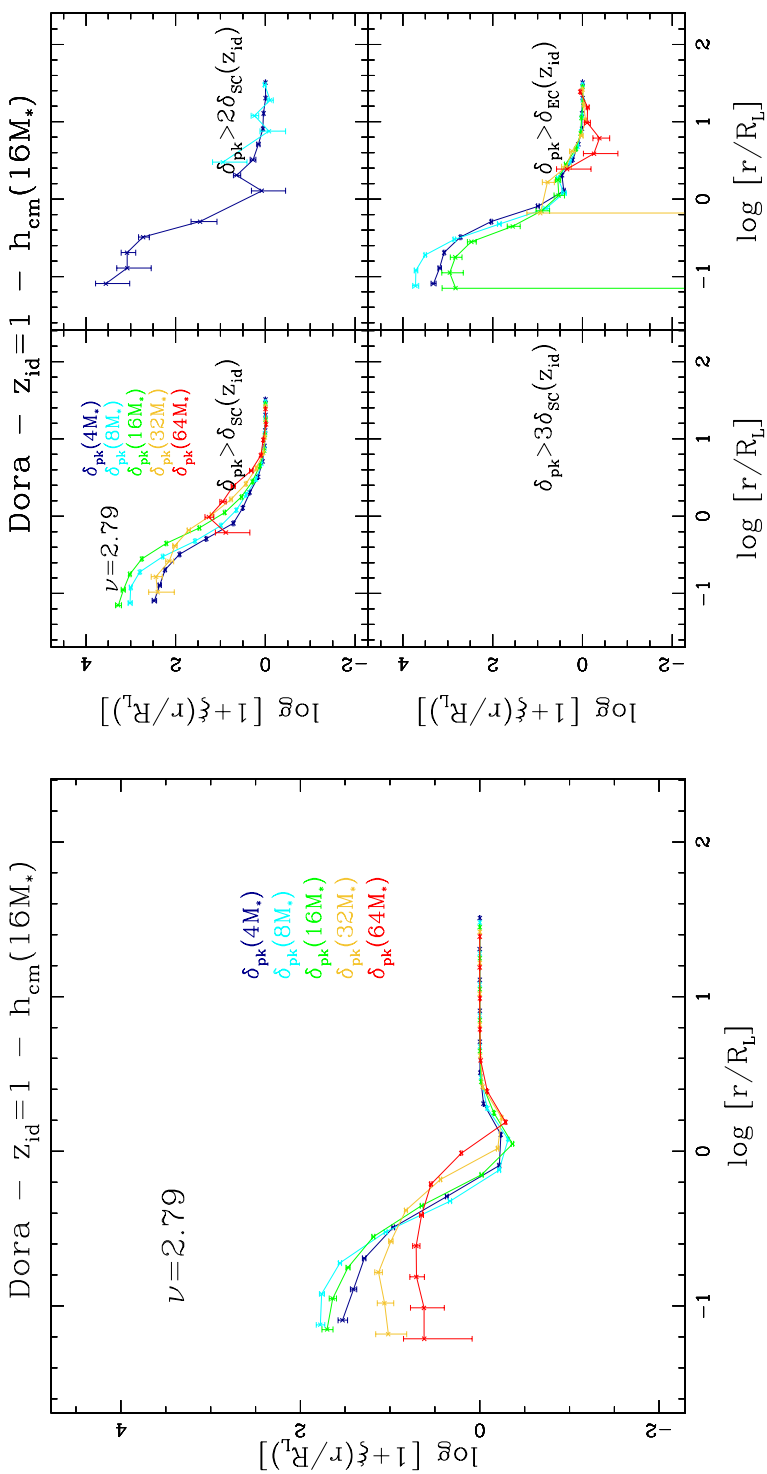

Figure D.6: Correlation functions between centers of mass of protohalos with $\nu = 2.79$ and δ peaks on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 1$. Halos have been identified with Dora simulation. Left figure: all peaks; right figure: peaks larger than four different thresholds.

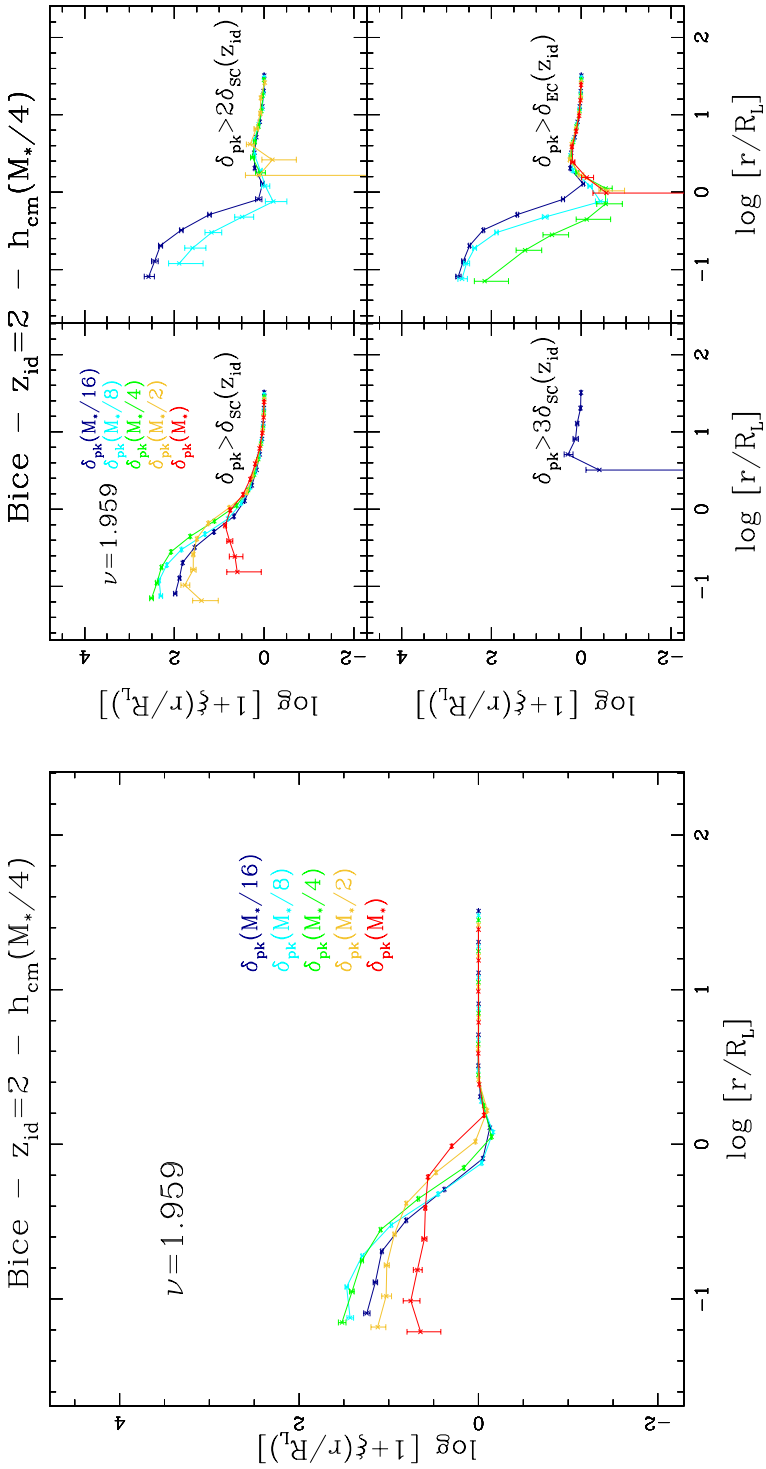

Figure D.7: Correlation functions between centers of mass of protohalos with $\nu = 1.959$ and δ peaks on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 2$. Halos have been identified with Dora simulation. Left figure: all peaks; right figure: peaks larger than four different thresholds.

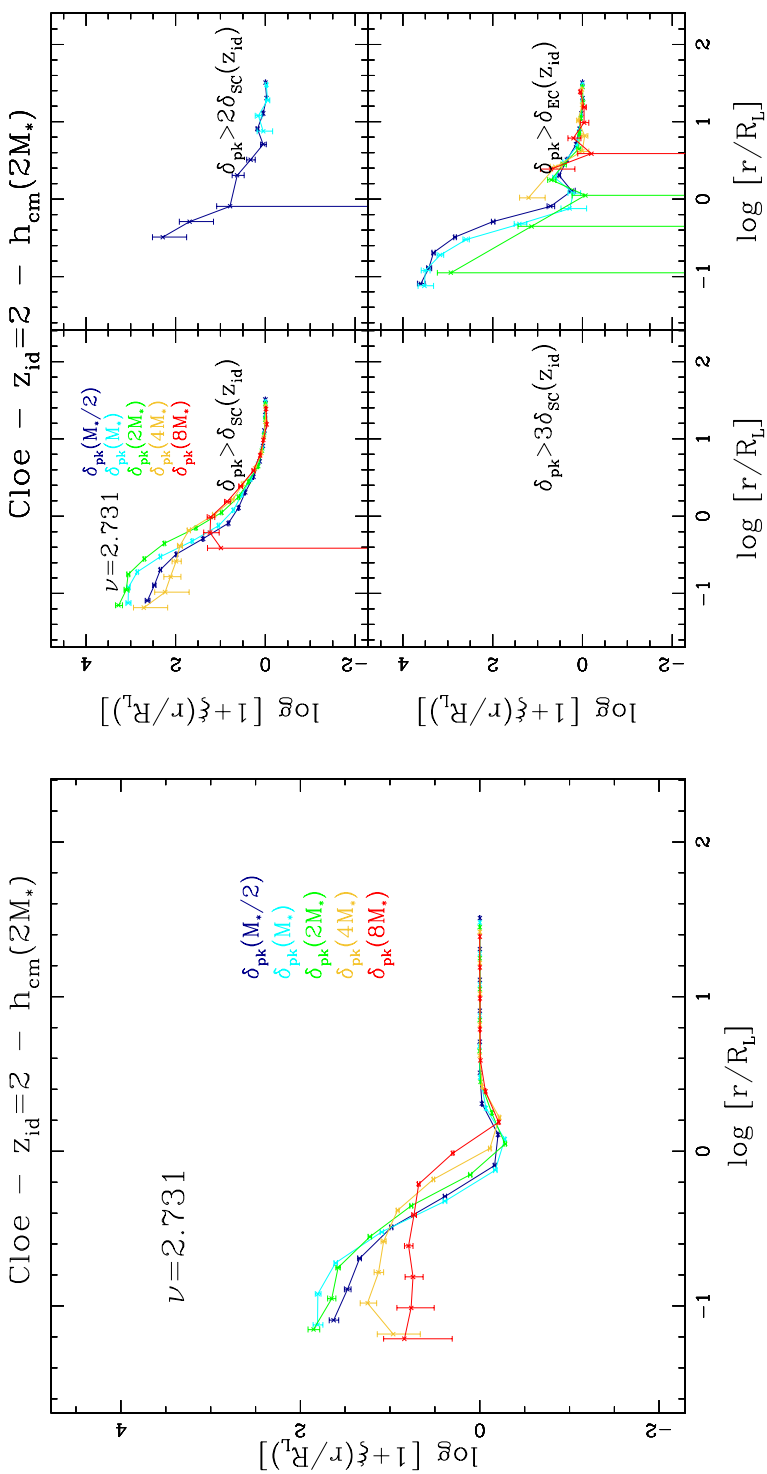

Figure D.8: Correlation functions between centers of mass of protohalos with $\nu = 2.731$ and δ peaks on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 2$. Halos have been identified with Cloe simulation. Left figure: all peaks; right figure: peaks larger than four different thresholds.

D.2 Protohalo centers and peaks/valleys of q

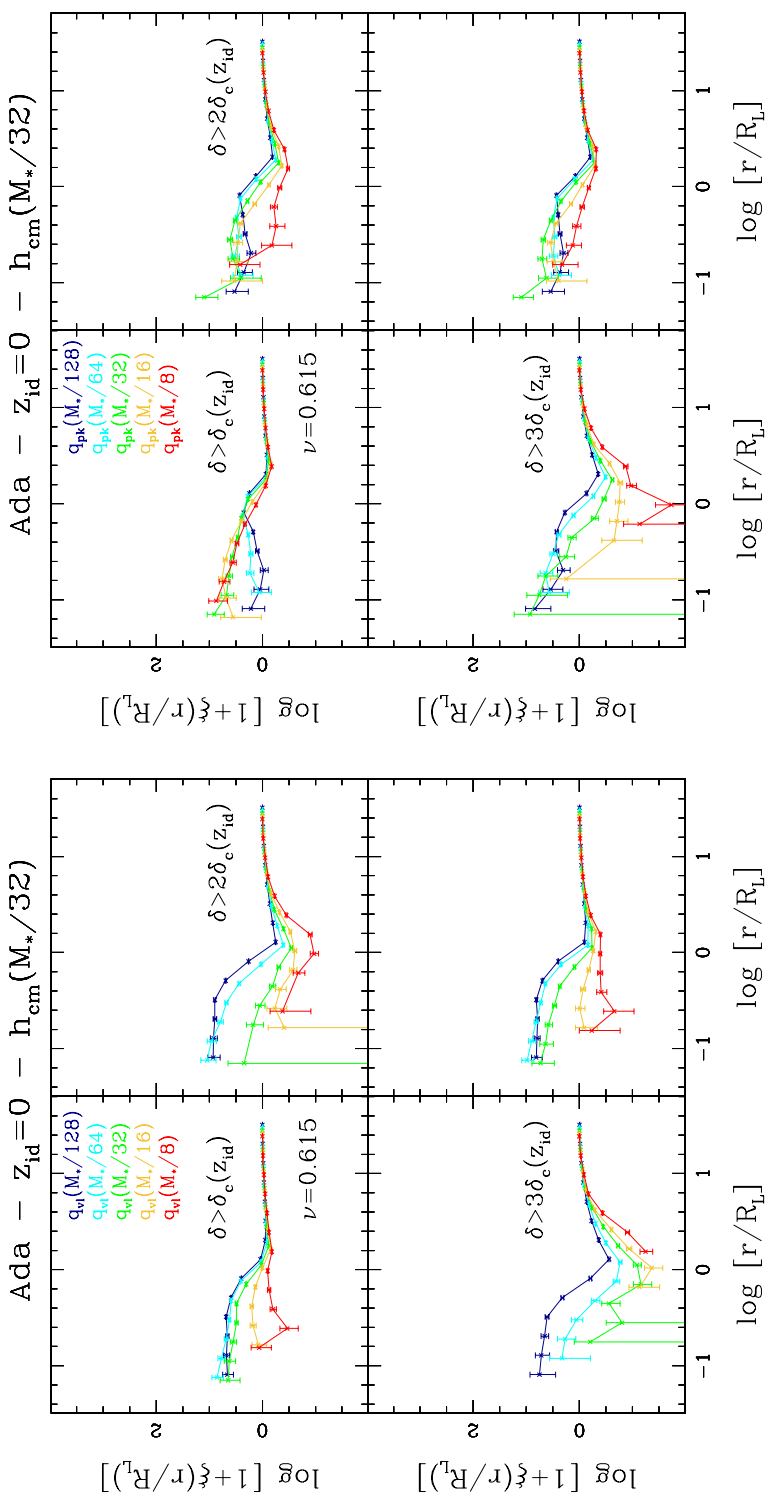

Figure D.9: Correlation functions between centers of mass of protohalos with $\nu = 0.615$ and q peaks (right) and valleys (left) on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 0$. Halos have been identified with Ada simulation. We consider four different threshold for the value of δ.

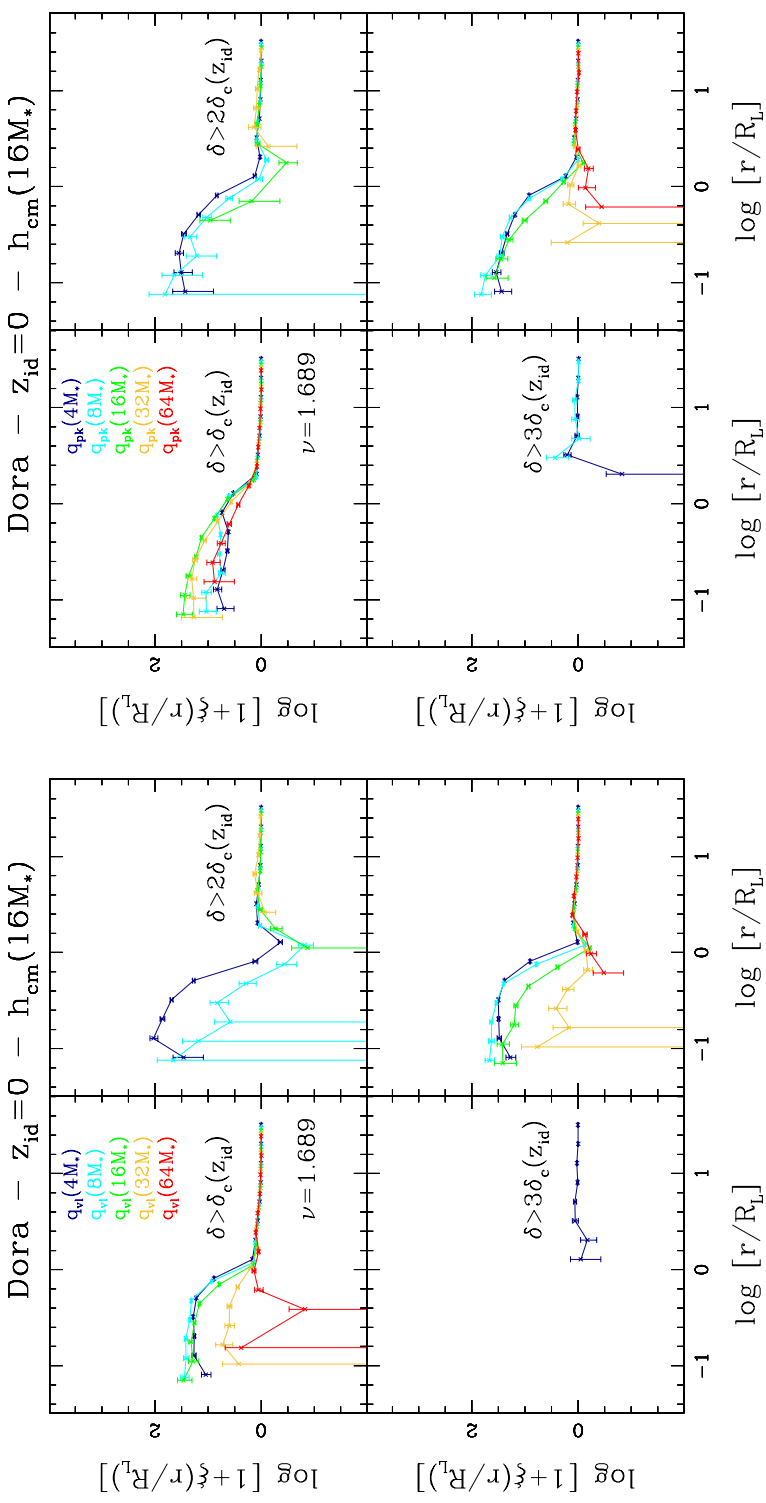

Figure D.10: Correlation functions between centers of mass of protohalos with $\nu = 1.689$ and q peaks (right) and valleys (left) on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 0$. Halos have been identified with Dora simulation. We consider four different threshold for the value of δ.

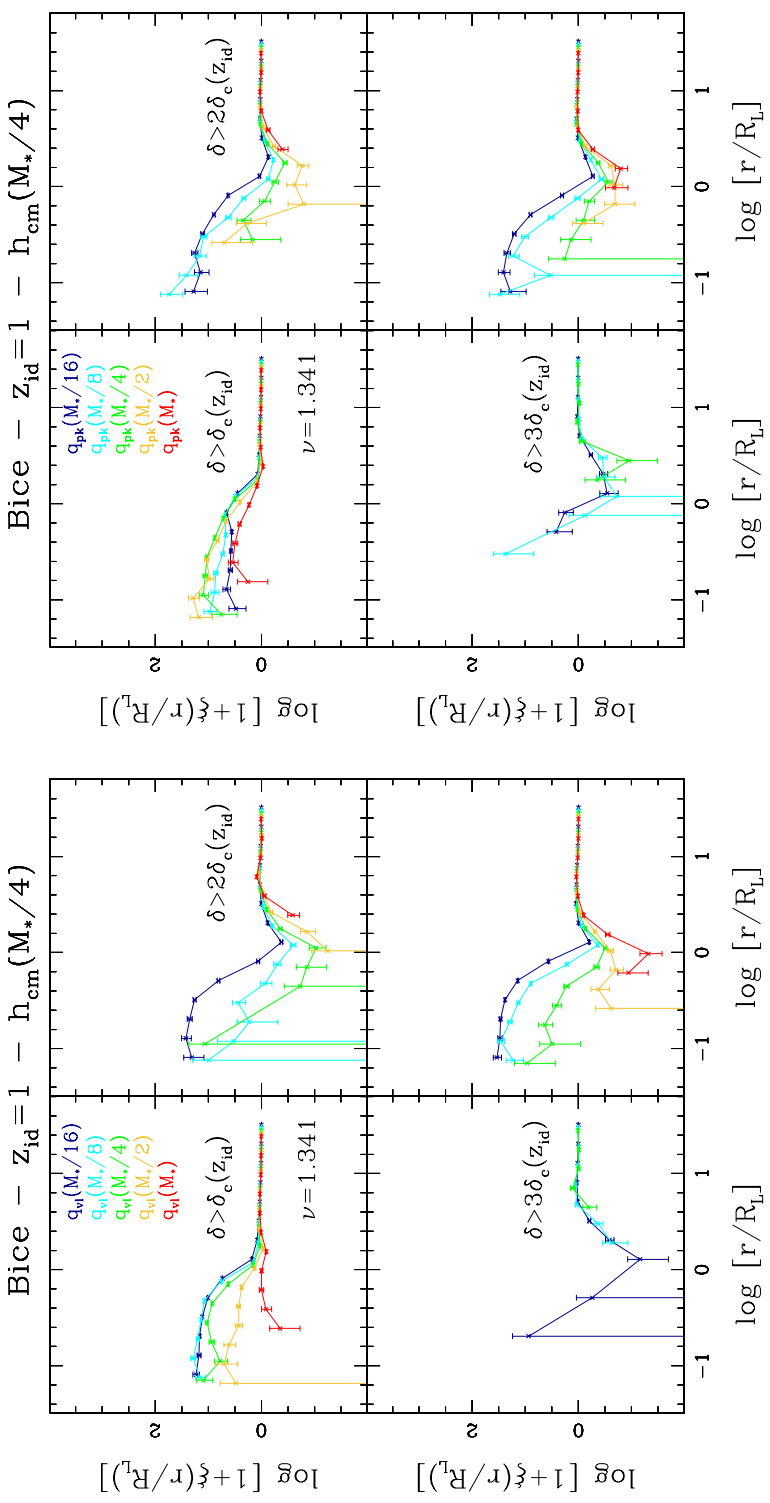

Figure D.11: Correlation functions between centers of mass of protohalos with $\nu = 1.341$ and q peaks (right) and valleys (left) on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 1$. Halos have been identified with Bice simulation. We consider four different threshold for the value of δ.

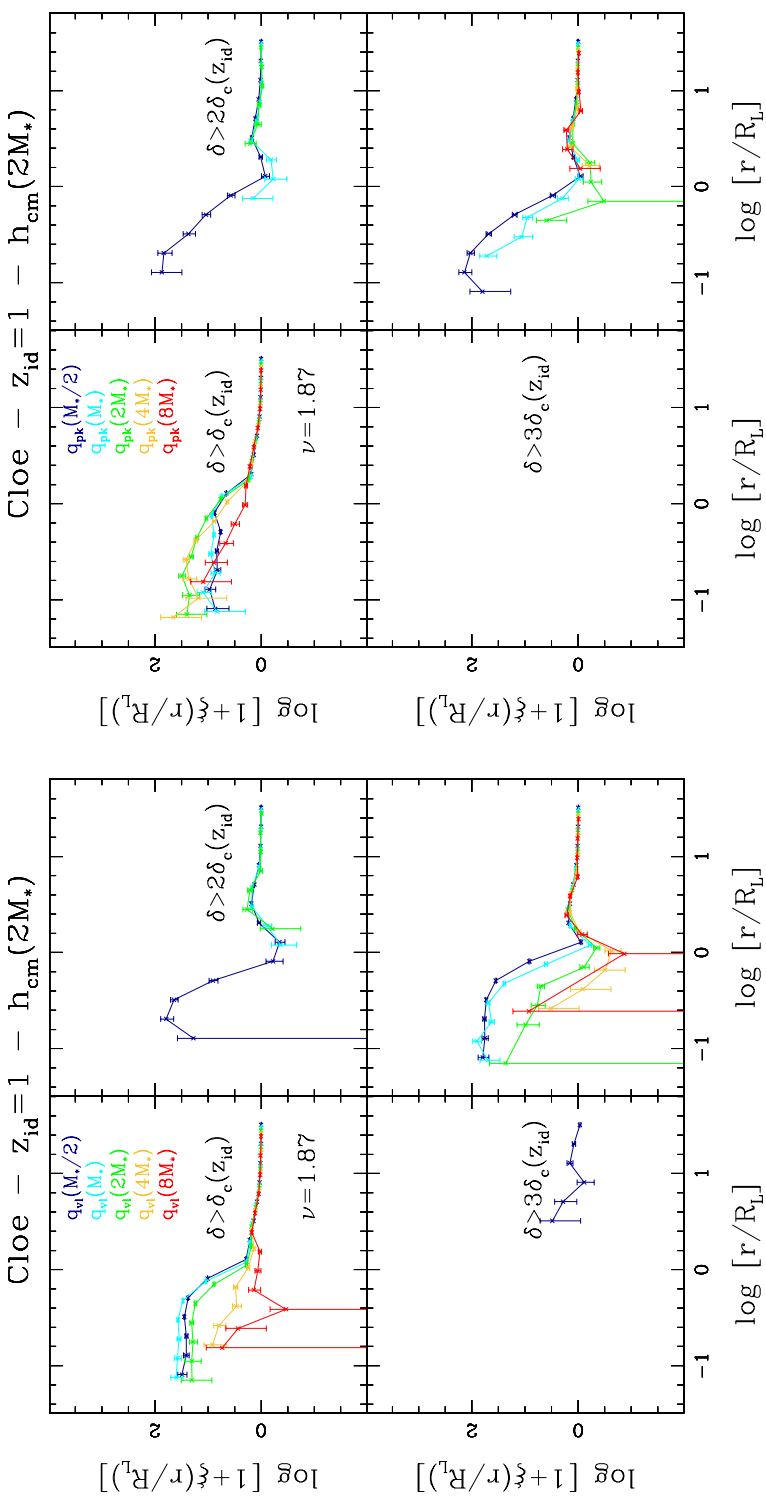

Figure D.12: Correlation functions between centers of mass of protohalos with $\nu = 1.87$ and q peaks (right) and valleys (left) on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 1$. Halos have been identified with Bice simulation. We consider four different threshold for the value of δ.

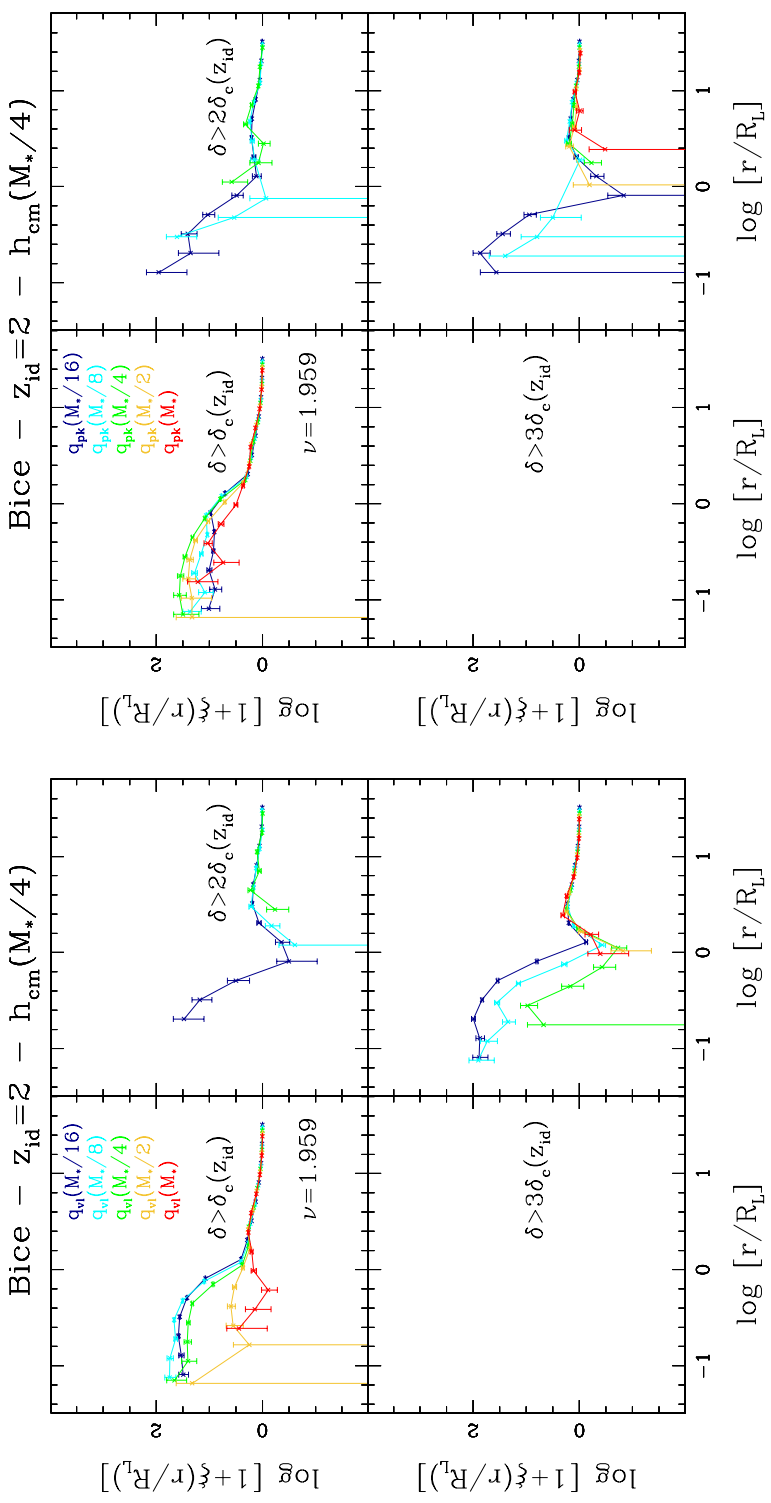

Figure D.13: Correlation functions between centers of mass of protohalos with $\nu = 1.959$ and q peaks (right) and valleys (left) on five different scales, encompassing five different masses at the initial conditions. Protohalo particles are traced back from the halo identification redshift at $z_{id} = 2$. Halos have been identified with Bice simulation. We consider four different threshold for the value of δ.

List of Tables

List of Figures

Bibliography

PAR Ade, N Aghanim, MIR Alves, C Armitage-Caplan, M Arnaud, M Ashdown, F Atrio-Barandela, J Aumont, H Aussel, C Baccigalupi, et al. Planck 2013 results. i. overview of products and scientific results. *Astronomy & Astrophysics*, 571:A1, 2014.

Brandon Allgood, Ricardo A Flores, Joel R Primack, Andrey V Kravtsov, Risa H Wechsler, Andreas Faltenbacher, and James S Bullock. The shape of dark matter haloes: dependence on mass, redshift, radius and formation. *Monthly Notices of the Royal Astronomical Society*, 367(4):1781–1796, 2006.

Lone Appel and BJT Jones. The mass function in biased galaxy formation scenarios. *Monthly Notices of the Royal Astronomical Society*, 245:522, 1990.

Edouard Audit and J-M Alimi. Gravitational lagrangian dynamics of cold matter using the deformation tensor. *Astronomy and Astrophysics*, 315:11–20, 1996.

Edouard Audit, Romain Teyssier, and Jean-Michel Alimi. Non-linear dynamics and mass function of cosmic structures: I analytical results. *arXiv preprint astro-ph/9704023*, 1997.

Vladimir Avila-Reese, Pedro Colín, Stefan Gottlöber, Claudio Firmani, and Christian Maulbetsch. The dependence on environment of cold dark matter halo properties. *The Astrophysical Journal*, 634(1):51, 2005.

James M Bardeen, JR Bond, Nick Kaiser, and AS Szalay. The statistics of peaks of gaussian random fields. *The Astrophysical Journal*, 304:15–61, 1986.

AJ Benson, S Cole, CS Frenk, CM Baugh, and CG Lacey. The nature of galaxy bias and clustering. *Monthly Notices of the Royal Astronomical Society*, 311(4):793–808, 2000.

Francis Bernardeau. The nonlinear evolution of rare events. *arXiv preprint astro-ph/9311066*, 1993.

E Bertschinger. Self-similar secondary infall and accretion in an einstein-de sitter universe. *The Astrophysical Journal Supplement Series*, 58:39–65, 1985.

Edmund Bertschinger and Bhuvnesh Jain. Gravitational instability of cold matter. *arXiv preprint astro-ph/9307033*, 1993.

J.R. Bond. in large-scale motions in the universe. page 419, 1989.

JR Bond and ST Myers. The peak-patch picture of cosmic catalogs. i. algorithms. *The Astrophysical Journal Supplement Series*, 103:1, 1996.

JR Bond, S Cole, G Efstathiou, and Nick Kaiser. Excursion set mass functions for hierarchical gaussian fluctuations. *The Astrophysical Journal*, 379:440–460, 1991.

Richard G Bower. The evolution of groups of galaxies in the press–schechter formalism. *Monthly Notices of the Royal Astronomical Society*, 248(2):332–352, 1991.

RG Carlberg and HMP Couchman. Mergers and bias in a cold dark matter cosmology. *The Astrophysical Journal*, 340:47–68, 1989.

E. Castorina, E. Sefusatti, R. K. Sheth, F. Villaescusa-Navarro, and M. Viel. Cosmology with massive neutrinos II: on the universality of the halo mass function and bias. *JCAP*, 2:049, February 2014. doi: 10.1088/1475-7516/2014/02/049.

Emanuele Castorina and Ravi K Sheth. Stochastic bias in multidimensional excursion set approaches. *Monthly Notices of the Royal Astronomical Society*, 433(2):1529–1536, 2013.

Emanuele Castorina, Aseem Paranjape, Oliver Hahn, and Ravi K Sheth. Excursion set peaks: the role of shear. *arXiv preprint arXiv:1611.03619*, 2016.

S Chandrasekhar. Solutions of two problems in the theory of gravitational radiation. *Physical Review Letters*, 24(11):611, 1970.

Camila A Correa, J Stuart B Wyithe, Joop Schaye, and Alan R Duffy. The accretion history of dark matter haloes–i. the physical origin of the universal function. *Monthly Notices of the Royal Astronomical Society*, 450(2):1514–1520, 2015.

Neal Dalal, Olivier Dore, Dragan Huterer, and Alexander Shirokov. Imprints of primordial non-gaussianities on large-scale structure: Scale-dependent bias and abundance of virialized objects. *Physical Review D*, 77(12):123514, 2008.

Marc Davis, George Efstathiou, Carlos S Frenk, and Simon DM White. The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, 292:371–394, 1985.

Pea de Bernardis, Peter AR Ade, JJ Bock, JR Bond, J Borrill, A Boscaleri, K Coble, BP Crill, G De Gasperis, PC Farese, et al. A flat universe from high-resolution maps of the cosmic microwave background radiation. *Nature*, 404(6781):955–959, 2000.

Vincent Desjacques. Environmental dependence in the ellipsoidal collapse model. *Monthly Notices of the Royal Astronomical Society*, 388(2):638–658, 2008.

Vincent Desjacques and Ravi K Sheth. Redshift space correlations and scale-dependent stochastic biasing of density peaks. *Physical Review D*, 81(2):023526, 2010.

Giulia Despali. Triaxial collapse of dark matter haloes. https://giuliadespali.wordpress.com/research-topics/triaxial-collapse-of-dark-matter-haloes/. Available on line.

Giulia Despali, Giuseppe Tormen, and Ravi K Sheth. Ellipsoidal halo finders and implications for models of triaxial halo formation. *Monthly Notices of the Royal Astronomical Society*, page stt235, 2013.

Giulia Despali, Carlo Giocoli, and Giuseppe Tormen. Some like it triaxial: the universality of dark matter halo shapes and their evolution along the cosmic time. *Monthly Notices of the Royal Astronomical Society*, 443(4):3208–3217, 2014.

Giulia Despali, Carlo Giocoli, Raul E Angulo, Giuseppe Tormen, Ravi K Sheth, Giacomo Baso, and Lauro Moscardini. The universality of the virial halo mass function and models for non-universality of other halo definitions. *Monthly Notices of the Royal Astronomical Society*, 456(3):2486–2504, 2016.

Benedikt Diemer and Andrey V Kravtsov. Dependence of the outer density profiles of halos on their mass accretion rate. *The Astrophysical Journal*, 789(1):1, 2014.

AG Doroshkevich. Spatial structure of perturbations and origin of galactic rotation in fluctuation theory. *Astrophysics*, 6(4):320–330, 1970.

John Dubinski. Cosmological tidal shear. *The Astrophysical Journal*, 401:441–454, 1992.

G Efstathiou and MJ Rees. High-redshift quasars in the cold dark matter cosmogony. *Monthly Notices of the Royal Astronomical Society*, 230(1):5P–11P, 1988.

George Efstathiou, S Michael Fall, and Craig Hogan. Self-similar gravitational clustering. *Monthly Notices of the Royal Astronomical Society*, 189(2):203–220, 1979.

J Einasto. On the construction of a composite model for the galaxy and on the determination of the system of galactic parameters. *Trudy Astrofizicheskogo Instituta Alma-Ata*, 5:87–100, 1965.

Jaan Einasto, Enn Saar, Ants Kaasik, and ARTHUR D Chernin. Missing mass around galaxies-morphological evidence. *Nature*, 252:111–113, 1974.

Richard I Epstein. Proto-galactic perturbations. *Monthly Notices of the Royal Astronomical Society*, 205(1):207–229, 1983.

James A Fillmore and Peter Goldreich. Self-similar gravitational collapse in an expanding universe. *The Astrophysical Journal*, 281:1–8, 1984.

Carlos S Frenk and Simon DM White. Dark matter and cosmic structure. *Annalen der Physik*, 524(9-10):507–534, 2012.

George Gamow and Edward Teller. On the origin of great nebulae. *Physical Review*, 55 (7):654, 1939.

Liang Gao, Volker Springel, and Simon DM White. The age dependence of halo clustering. *Monthly Notices of the Royal Astronomical Society: Letters*, 363(1):L66–L70, 2005.

Liang Gao, N Yoshida, T Abel, CS Frenk, A Jenkins, and V Springel. The first generation of stars in the λ cold dark matter cosmology. *Monthly Notices of the Royal Astronomical Society*, 378(2):449–468, 2007.

Carlo Giocoli, Giuseppe Tormen, and Frank C Van Den Bosch. The population of dark matter subhaloes: mass functions and average mass-loss rates. *Monthly Notices of the Royal Astronomical Society*, 386(4):2135–2144, 2008.

Carlo Giocoli, Giuseppe Tormen, and Ravi K Sheth. Formation times, mass growth histories and concentrations of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 422(1):185–198, 2012a.

Carlo Giocoli, Giuseppe Tormen, and Ravi K Sheth. Formation times, mass growth histories and concentrations of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 422(1):185–198, 2012b.

JR Gott. On the formation of elliptical galaxies. In *Symposium-International Astronomical Union*, volume 69, pages 271–285. Cambridge Univ Press, 1975.

James E Gunn and J Richard Gott III. On the infall of matter into clusters of galaxies and some effects on their evolution. *The Astrophysical Journal*, 176:1, 1972.

JE Gunn. Massive galactic halos. i-formation and evolution. *The Astrophysical Journal*, 218:592–598, 1977.

Alan H Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, 23(2):347, 1981.

Oliver Hahn and Aseem Paranjape. The locations of halo formation and the peaks formalism. *Monthly Notices of the Royal Astronomical Society*, page stt2256, 2013.

Geraint Harker, Shaun Cole, John Helly, Carlos Frenk, and Adrian Jenkins. A marked correlation function analysis of halo formation times in the millennium simulation. *Monthly Notices of the Royal Astronomical Society*, 367(3):1039–1049, 2006.

Edward R Harrison. Fluctuations at the threshold of classical cosmology. *Physical review D*, 1(10):2726, 1970.

Eric Hayashi and Simon DM White. Understanding the halo-mass and galaxy-mass cross-correlation functions. *Monthly Notices of the Royal Astronomical Society*, 388(1):2–14, 2008.

Y Hoffman. The dynamics of superclusters-the effect of shear. *The Astrophysical Journal*, 308:493–498, 1986.

James H Jeans. The stability of a spherical nebula. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 199:1–53, 1902.

Adrian Jenkins, CS Frenk, Simon DM White, JM Colberg, Shaun Cole, AE Evrard, HMP Couchman, and N Yoshida. The mass function of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 321(2):372–384, 2001.

YP Jing and Yasushi Suto. Triaxial modeling of halo density profiles with high-resolution n-body simulations. *The Astrophysical Journal*, 574(2):538, 2002.

YP Jing, HJ Mo, and Gerhard Börner. Spatial correlation function and pairwise velocity dispersion of galaxies: Cold dark matter models versus the las campanas survey. *The Astrophysical Journal*, 494(1):1, 1998.

Neal Katz, Thomas Quinn, and James M Gelb. Galaxy formation and the peaks formalism. *Monthly Notices of the Royal Astronomical Society*, 265(3):689–705, 1993.

G Kauffmann, Simon DM White, and B Guiderdoni. The formation and evolution of galaxies within merging dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 264(1):201–218, 1993.

Guinevere Kauffmann, Adi Nusser, and Matthias Steinmetz. Galaxy formation and large-scale bias. *Monthly Notices of the Royal Astronomical Society*, 286(4):795–811, 1997.

Tetsu Kitayama and Yasushi Suto. Semi-analytic predictions for statistical properties of x-ray clusters of galaxies in cold dark matter universes. *arXiv preprint astro-ph/9604141*, 1996.

Anatoly Klypin, Stefano Borgani, Jon Holtzman, and Joel Primack. Damped lyman alpha systems vs. cold+ hot dark matter. *arXiv preprint astro-ph/9410022*, 1994.

Lawrence M Krauss and Michael S Turner. The cosmological constant is back. *General Relativity and Gravitation*, 27(11):1137–1144, 1995.

Cedric Lacey and Shaun Cole. Merger rates in hierarchical models of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 262(3):627–649, 1993.

Tsz Yan Lam and Ravi K Sheth. Ellipsoidal collapse and the redshift-space probability distribution function of dark matter. *Monthly Notices of the Royal Astronomical Society*, 389(3):1249–1258, 2008.

Jounghun Lee and Sergei F Shandarin. The cosmological mass distribution function in the zeldovich approximation. *The Astrophysical Journal*, 500(1):14, 1998.

Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of cosmic microwave background anisotropies in closed friedmann-robertson-walker models. *The Astrophysical Journal*, 538(2):473, 2000.

Evgenii Mikhailovich Lifshitz. On the gravitational stability of the expanding universe. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki*, 16:587–602, 1946.

Yu Lu, HJ Mo, Neal Katz, and Martin D Weinberg. On the origin of cold dark matter halo density profiles. *Monthly Notices of the Royal Astronomical Society*, 368(4):1931–1940, 2006.

VA Lubimov, EG Novikov, VZ Nozik, EF Tretyakov, and VS Kosik. An estimate of the ν e mass from the β-spectrum of tritium in the valine molecule. *Physics Letters B*, 94 (2):266–268, 1980.

Aaron D Ludlow and Cristiano Porciani. The peaks formalism and the formation of cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 413(3):1961–1972, 2011.

Aaron D Ludlow, Julio F Navarro, Michael Boylan-Kolchin, Philip E Bett, Raúl E Angulo, Ming Li, Simon DM White, Carlos Frenk, and Volker Springel. The mass profile and accretion history of cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 432(2):1103–1113, 2013.

Donald Lynden-Bell. Statistical mechanics of violent relaxation in stellar systems. *Monthly Notices of the Royal Astronomical Society*, 136:101, 1967.

Alberto Manrique and Eduard Salvador-Sole. The confluent system formalism: I. the mass function of objects in the peak model. *arXiv preprint astro-ph/9505083*, 1995.

P Meszaros. The behaviour of point masses in an expanding cosmological substratum. *Astronomy and Astrophysics*, 37:225–228, 1974.

Jordi Miralda-Escudé. The dark age of the universe. *Science*, 300(5627):1904–1909, 2003.

HJ Mo and Simon DM White. An analytic model for the spatial clustering of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 282(2):347–361, 1996.

Houjun Mo, Frank Van den Bosch, and Simon White. *Galaxy formation and evolution*. Cambridge University Press, 2010.

Marcello Musso and Ravi K Sheth. One step beyond: The excursion set approach with correlated steps. *Monthly Notices of the Royal Astronomical Society: Letters*, 423(1): L102–L106, 2012.

J Navarro and SD White. The structure of cold dark matter halos. In *SYMPOSIUM-INTERNATIONAL ASTRONOMICAL UNION*, volume 171, pages 255–258. KLUWER ACADEMIC PUBLISHERS GROUP, 1996.

Julio F Navarro, Eric Hayashi, Chris Power, AR Jenkins, Carlos S Frenk, Simon DM White, Volker Springel, Joachim Stadel, and Thomas R Quinn. The inner structure of λcdm haloes–iii. universality and asymptotic slopes. *Monthly Notices of the Royal Astronomical Society*, 349(3):1039–1051, 2004.

Julio F Navarro, Aaron Ludlow, Volker Springel, Jie Wang, Mark Vogelsberger, Simon DM White, Adrian Jenkins, Carlos S Frenk, and Amina Helmi. The diversity and similarity of simulated cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 402(1):21–34, 2010.

Angelo F Neto, Liang Gao, Philip Bett, Shaun Cole, Julio F Navarro, Carlos S Frenk, Simon DM White, Volker Springel, and Adrian Jenkins. The statistics of λ cdm halo concentrations. *Monthly Notices of the Royal Astronomical Society*, 381(4):1450–1462, 2007.

Adi Nusser and Ravi K Sheth. Mass growth and density profiles of dark matter haloes in hierarchical clustering. *Monthly Notices of the Royal Astronomical Society*, 303(4): 685–695, 1999.

Jeremiah P Ostriker and Paul J Steinhardt. Cosmic concordance. *arXiv preprint astro-ph/9505066*, 1995.

JP Ostriker, PJE Peebles, and A Yahil. The size and mass of galaxies, and the mass of the universe. *Particle Physics and Cosmology: Dark Matter*, 6:24, 2012.

Aseem Paranjape and Ravi K Sheth. Peaks theory and the excursion set approach. *Monthly Notices of the Royal Astronomical Society*, 426(4):2789–2796, 2012.

Aseem Paranjape, Ravi K Sheth, and Vincent Desjacques. Excursion set peaks: a self-consistent model of dark halo abundances and clustering. *Monthly Notices of the Royal Astronomical Society*, page stt267, 2013.

JA Peacock and AF Heavens. Alternatives to the press–schechter cosmological mass function. *Monthly Notices of the Royal Astronomical Society*, 243(1):133–143, 1990.

JA Peacock and RE Smith. Halo occupation numbers and galaxy bias. *Monthly Notices of the Royal Astronomical Society*, 318(4):1144–1156, 2000.

Phillip James Edwin Peebles. *The large-scale structure of the universe*. Princeton university press, 1980.

Collaboration Planck, PAR Ade, N Aghanim, M Arnaud, M Ashdown, J Aumont, C Baccigalupi, M Baker, A Balbi, AJ Banday, et al. Planck early results. i. the planck mission. 2011.

Cristiano Porciani, Avishai Dekel, and Yehuda Hoffman. Testing tidal-torque theory–ii. alignment of inertia and shear and the characteristics of protohaloes. *Monthly Notices of the Royal Astronomical Society*, 332(2):339–351, 2002.

William H Press and Paul Schechter. Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. *The Astrophysical Journal*, 187:425–438, 1974.

Brant E Robertson, Andrey V Kravtsov, Jeremy Tinker, and Andrew R Zentner. Collapse barriers and halo abundance: testing the excursion set ansatz. *The Astrophysical Journal*, 696(1):636, 2009.

Graziano Rossi, Ravi K Sheth, and Giuseppe Tormen. Modelling the shapes of the largest gravitationally bound objects. *Monthly Notices of the Royal Astronomical Society*, 416 (1):248–261, 2011.

Eduard Salvador-Solé, Jordi Viñas, Alberto Manrique, and Sinue Serra. Theoretical dark matter halo density profile. *Monthly Notices of the Royal Astronomical Society*, 423(3): 2190–2202, 2012.

Jiajian Shen, Tom Abel, HJ Mo, and Ravi K Sheth. An excursion set model of the cosmic web: the abundance of sheets, filaments, and halos. *The Astrophysical Journal*, 645(2): 783, 2006.

Ravi K Sheth and Giuseppe Tormen. Large-scale bias and the peak background split. *Monthly Notices of the Royal Astronomical Society*, 308(1):119–126, 1999.

Ravi K Sheth and Giuseppe Tormen. An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier. *Monthly Notices of the Royal Astronomical Society*, 329(1):61–75, 2002.

Ravi K Sheth and Giuseppe Tormen. On the environmental dependence of halo formation. *Monthly Notices of the Royal Astronomical Society*, 350(4):1385–1390, 2004.

Ravi K Sheth and Rien Van De Weygaert. A hierarchy of voids: much ado about nothing. *Monthly Notices of the Royal Astronomical Society*, 350(2):517–538, 2004.

Ravi K Sheth, HJ Mo, and Giuseppe Tormen. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 323(1):1–12, 2001.

Ravi K Sheth, Kwan Chuen Chan, and Román Scoccimarro. Nonlocal lagrangian bias. *Physical Review D*, 87(8):083002, 2013.

George F Smoot, CL Bennett, A Kogut, EL Wright, J Aymon, NW Boggess, ES Cheng, G De Amici, S Gulkis, MG Hauser, et al. Structure in the cobe differential microwave radiometer first-year maps. *The Astrophysical Journal*, 396:L1–L5, 1992.

Volker Springel. The cosmological simulation code gadget-2. *Monthly Notices of the Royal Astronomical Society*, 364(4):1105–1134, 2005.

Volker Springel, Simon DM White, Giuseppe Tormen, and Guinevere Kauffmann. Populating a cluster of galaxies–i. results at z= 0. *Monthly Notices of the Royal Astronomical Society*, 328(3):726–750, 2001a.

Volker Springel, Naoki Yoshida, and Simon DM White. Gadget: a code for collisionless and gasdynamical cosmological simulations. *New Astronomy*, 6(2):79–117, 2001b.

Jeremy Tinker, Andrey V Kravtsov, Anatoly Klypin, Kevork Abazajian, Michael Warren, Gustavo Yepes, Stefan Gottlöber, and Daniel E Holz. Toward a halo mass function for precision cosmology: The limits of universality. *The Astrophysical Journal*, 688(2):709, 2008.

Giuseppe Tormen. *Formazione delle strutture cosmiche*. 2010.

Giuseppe Tormen, Lauro Moscardini, and Naoki Yoshida. Properties of cluster satellites in hydrodynamical simulations. *Monthly Notices of the Royal Astronomical Society*, 350 (4):1397–1408, 2004.

HY Wang, HJ Mo, and YP Jing. Environmental dependence of cold dark matter halo formation. *Monthly Notices of the Royal Astronomical Society*, 375(2):633–639, 2007.

Michael S Warren, Kevork Abazajian, Daniel E Holz, and Luis Teodoro. Precision determination of the mass function of dark matter halos. *The Astrophysical Journal*, 646(2): 881, 2006.

Risa H Wechsler, Andrew R Zentner, James S Bullock, Andrey V Kravtsov, and Brandon Allgood. The dependence of halo clustering on halo formation history, concentration, and occupation. *The Astrophysical Journal*, 652(1):71, 2006.

Simon DM White and MJ Rees. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. *Monthly Notices of the Royal Astronomical Society*, 183(3):341–358, 1978.

SIMON DM White and JOSEPH Silk. The growth of aspherical structure in the universe-is the local supercluster an unusual system. *The Astrophysical Journal*, 231:1–9, 1979.

YA B Zel'Dovich. Gravitational instability: An approximate theory for large density perturbations. *Astronomy and astrophysics*, 5:84–89, 1970.

Ya B Zeldovich. A hypothesis, unifying the structure and the entropy of the universe. *Monthly Notices of the Royal Astronomical Society*, 160(1):1P–3P, 1972.

Andrew R Zentner, Andreas A Berlind, James S Bullock, Andrey V Kravtsov, and Risa H Wechsler. The physics of galaxy clustering. i. a model for subhalo populations. *The Astrophysical Journal*, 624(2):505, 2005.