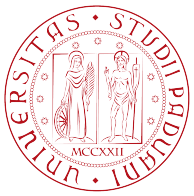

Università degli Studi di Padova

DIPARTIMENTO DI FISICA E ASTRONOMIA "GALILEO GALILEI"

Tesi di Laurea Magistrale in Fisica

Tesi di laurea

Modelling the correlation function of voids in the galaxy distribution

Laureando
Alan Hubert
Matricola 1084177

Relatore
Prof. Sabino Matarrese

Anno Accademico 2018-2019

Contents

Prologue **3**

1 Void description and determination of density thresholds **7**
1.1 Evolution of voids 7
1.2 Uncompensated case with collisionless gas 9
1.3 Calculating density thresholds 13

2 Excursion set approach for haloes **21**
2.1 Gaussian fields topology 22
2.2 Absorbing barrier problem 24
2.3 Excursion set theory related to N-body 29

3 Voids hierarchy **33**
3.1 Void sociology 33
3.2 Void size distribution 38
3.3 Volume conserving model 44
3.4 New treatment of the void in cloud process 55

4 Clustering properties of dark matter haloes **63**
4.1 From joint upcrossing distribution to two-point correlation function 63
4.2 Lagrangian halo correlation function in the Einstein-de Sitter universe . . 69
4.3 A better analytical description of Lagrangian bias 74

5 Description of the two-point correlation function of voids **87**
5.1 A model for the two-point correlation function of voids 88
5.2 Clustering of cosmic voids 91

Conclusions **106**

Prologue

Theoretical and experimental evidences support the idea that little primordial density perturbations born in the early universe, almost homogenous, became bigger in amplitude through a process of gravitational instability giving the complex structure of modern universe.
The distribution of matter evolved in a hierarchical scenario, in a bottom-up way, gathering in cosmic structure formation characterized by filamentary and sheet-like structures. Most of the mass in the universe virialized in haloes, as observed by galaxy redshift surveys and numerical simulations, squeezed in between expanding underdense regions devoid of galaxies, called *voids*. Most of the volume in the universe is occupied by these voids.
The first complete analytical model used to study the dark matter haloes abundance was initially proposed by Press & Shechter (1974) and after modified and improved by Epstein (1983) and Bond [2]. Such a model, called *excursion set approach*, is based on the study of random walks in the presence of a suitable absorbing barrier. As regards clusters, given a random initial perturbation field, this model filters the perturbation with a very large radius, gradually decreasing the filter radius until the highest density spot overcomes the threshold and label the material within that filter radius as an object of the appropriate mass. We are assuming spherical collapse for these regions, to apply this process without problems, but the virializing of cosmic structures is quite different from spherical. Nevertheless the idealization of spherical symmetry is successful for the evolution of low-density regions, because are suited for an excursion set analysis based on a spherical evolution model of voids, based on a theory recently developed by R. Sheth and R. van de Weygaert [3].
Both cases exploit linear approximations to create two threshold values, one for the overdense regions and the other for the underdense ones, telling us that when an high density primordial perturbation overcomes the respective threshold we have a collapse, instead with a low density primordial perturbation we have the formation of a void.
The evolution of clusters and voids is more complicated than this, that is why Bond for clusters and Sheth and van de Weygaert for voids improved the model of the *excursion set approach* proposed by Press & Shechter with the idea of random walks and absorbing barriers to solve the problem of *cloud in cloud* for clusters, namely do not count haloes contained in bigger ones, the corresponding problem of *cloud in cloud* for low density regions called *void in void* and the most crucial aspect of the development of a void

hierarchy that describes the disappearance of small voids as they become embedded in larger scale overdensities, called *void in cloud*.
All we have just said is useful to calculate in the most precise way, through analytical and numerical methods, the mass function number density of clusters and voids in the different regions of space. It is crucial to know how these entities are distributed in order to know how to relate them to calculate correlation functions of haloes in Lagrangian space as reported by Porciani, Matarrese, Lucchin, Catelan [5], who used a stochastic approach to the spatial clustering of dark matter haloes, or by Scannapieco, Thacker [6], who developed an improved analytical construction of correlation functions of clusters with varying masses, formation redshifts and separations.
All this explanation is fundamental to develop what is the main purpose that will be treated in this thesis. We have seen reported in the works [5], [6] methods to study correlation functions of dark matter haloes. Instead, what we are interested in it is to build up an analytical model able to approximately describe the two-point correlation function of voids in the large-scale distribution of galaxies. The basis that will guide the main ideas of our reasoning will be the theories by R. Sheth and R. van de Weygaert. As we said before, they extend the *excursion set approach*, normally adopted to study the dark matter halo abundance, to underdense regions. But to predict the void two-point correlation function we have to modify the study of random walks (in void-size space) in the presence of a suitable absorbing barrier used in [3] in the analysis of multiple space-correlated random walks.
The thesis will begin with a description of the characteristics of the kind of void that will be analyzed and with the calculation of the density thresholds fundamental to understand when overdense regions collapse or turnaround and underdense regions become a void. After we describe the process of the *excursion set approach* first for the haloes and then for the more complex case regarding voids. Finally we will develop the main argument of the thesis, namely the analytical study of the two-point correlation function of voids based on the ideas reported in [5], considering the complications arising in the case of voids, the phenomenon of *void in cloud* and how the different threshold values influence the evolution of a void formation process.

Chapter 1

Void description and determination of density thresholds

Theories about the growth of large scale structure in the universe have been influenced by the discovery of voids in the distribution of galaxies in regions $\geq 10h^{-1}Mpc$, comparable in size with superclusters. All studies about voids express an analytical vision of symmetric voids in a uniform background, with or without compensating ridges, considering them a consequence of a dissipationless clustering scenario, coming from deep underdense regions in the primordial density field.
The most complete study about void evolution is the work done by Bertschinger, first in 1983 and then in [1], using nonrelativistic Newtonian physics, valid for scales of interest much less than the horizon scale requiring perturbations to be linear in the initial conditions, and assuming an Einstein-De Sitter universe, namely $\Omega = 1$, necessary assumption of the similarity solutions which correspond to the translation symmetry of equation of motion of fluid particles relative to different surrounding mass shells developed by voids, as reported in [1].
Void evolution is non linear and their expansion is self similar, and the spherical top-hat model used is justified by the fact that it is successful in describing the void formation and evolution with different density profiles and it is useful to demonstrate that a top-hat configuration is the typical development of a void.

1.1 Evolution of voids

Haloes form from the evolution of overdense regions that increase less rapidly than the background, reach a maximum size and finally turnaround and collapse to small size. The non linearity of the process tends to enhance the initial fluctuations of sphericity present in the primordial perturbations. So they are characterized by filamentary and sheet-like configurations consequence of the anisotropic flattening of structures before the final collapse. The combination of the gravitational forces directed inward and the natural anisotropy of the collapsing process develops to a contraction of initially aspherical density peaks along preferred axis. These conclusions make the ellipsoidal model of

collapse reasonable to be used in the case of overdense regions.
The evolution of voids is completely different. As they increase the spherical model is more and more adapt to describe the expansion of underdense regions. Namely, the anisotropic force field directed outwards will cause the strongest acceleration along the shortest axis and the weakest acceleration along the longest one. So the tendency of underdense regions is to nullify the asphericities present in the primordial Gaussian field, evolving to an even more spherical top-hat geometry.
It is very important in the void characterization the interaction with its surroundings in the analysis of the hierarchical picture evolution, since there are not isolated voids. But for a complete comprehension of the basic ideas about void structure and evolution is useful to analyze isolated negative density perturbations. A void forms after the gradual merging of its internal substructures, present on all scales, in a bottom-up process of assembly, that continue to expand until matter from the interior overtakes the outer shells. The spherical model applied to the evolution of these underdense regions is the most adapt and it has the advantage to make possible an analytical solution basing all the reasoning on a one dimensional treatment.
The evolution of voids is basically characterized by the following properties:

Expansion: differently from overdense regions which collapse, voids undergo simple expansion;

Evacuation: during the expansion the enhancement of the volume provokes a continuous density decreasing through the redistribution of mass over it. Also gravity contributes to density decreasing with the accumulation of mass near the edges;

Spherical shape: voids evolve more and more towards a spherical form during the expansion;

Top-hat density profile: the distribution of matter is characterized by a reverse top-hat profile because the repulsive effect pushing matter out the void is less important approaching the edges;

'Super Hubble' velocity field: the peculiar velocity field in voids has a constant 'Hubble like' interior velocity divergence, so voids transform into 'super-Hubble bubbles';

Suppressed structure growth: all density inhomogeneities and interior structure formation is stopped when underdense regions start to become bigger;

Boundary ridge: around the void grow ridges because matter moves from the interior to the boundaries;

Shell crossing: inner shells pass across outer ones when the void achieves the non linear stage of its evolution.

The last property of the list is fundamental because in the moment when shell crossing happens we have the transformation from an underdense region to a self similar spherical void expanding outwards nonlinearly. Shell crossing is the main characteristic of voids

with a collisionless gas solution, concerning particles with negligible cross section that has not effect on the system. Voids can be also with a collisional gas solution, but the dissipation of kinetic energy caused by the shock waves induced by the interaction between particles of this type makes impossible to have shell crossing. In the next section we will concentrate in giving some solutions of voids with a finite mass deficit at large distance, called uncompensated, considering the case of a collisionless fluid.
It is shell crossing that determines the void formation. The different steps of the evolution are influenced by the density parameter Ω, the initial density deficit Δ_i and the steepness of the density profile. Figure 1.1 shows these features. The panel on the left illustrates the development of an initial uncompensated top-hat depression, with initial density deficit set to $\Delta_{lin,0} = -10$, and comoving initial radius $R_{i,0} = 5h^{-1}Mpc$. The continuous decreasing of the density value together with the mass streaming to the outer shells make clear the tendency of void to expand. The evolution of these underdense regions does not present collapse as final configuration, moving their shells faster than the background. This peculiar behaviour makes possible the creation of dense walls of matter around the voids that were not present in the initial phase of the expansion.
In the panel on the right we have the same initial density deficit and characteristic radius of the left panel, but in the right one the initial profile is more adapt to represent the cosmological circumstances of void evolution because we have a radially averaged density profile in a Gaussian random field of cold dark matter density fluctuations. However we notice that the top-hat spherical model gives an accurate representation not only of isolated voids, but also of reality because the profile presented in the right panel in its evolution is more and more like the top-hat void in the left panel.
These voids present a minimum value of the density in their centre, in fact the density deficit $|\Delta(r)|$ decreases as a function of radius r, so the interior layers of the void move towards the edges more rapidly and the matter coming from the inner shells catches up with that present in the boundaries, leading to an abrupt increasing of the density in the outer realms. A consequence of the expansion of the inner void layers is the fact that the flat part of the density distribution becomes important in the void interior.

1.2 Uncompensated case with collisionless gas

Different ideas were developed in order to explain the creation of voids through analyzing nonlinear evolution of isolated, spherical, negative density perturbations in an Einstein-De Sitter universe. One of them is to consider voids without surrounding shells forming from a spectrum of overlapping random perturbations. In this scenario only voids of size $r \approx 4h^{-1}Mpc$ can be produced, but they are lost in the background because they do not expand rapidly enough to maintain their identity and they cannot be distinguished from those ones appearing in later times. So to justify the birth of the voids that are observed, with larger sizes ($r > 10h^{-1}Mpc$), spherical voids with surrounding shells in an Einstein-De Sitter universe result more adapt. If in the initial perturbation the shell does not present mass deficit we call the void compensated, namely surrounded by a compensating overdense mass shell; otherwise if this deficit exists the void is called

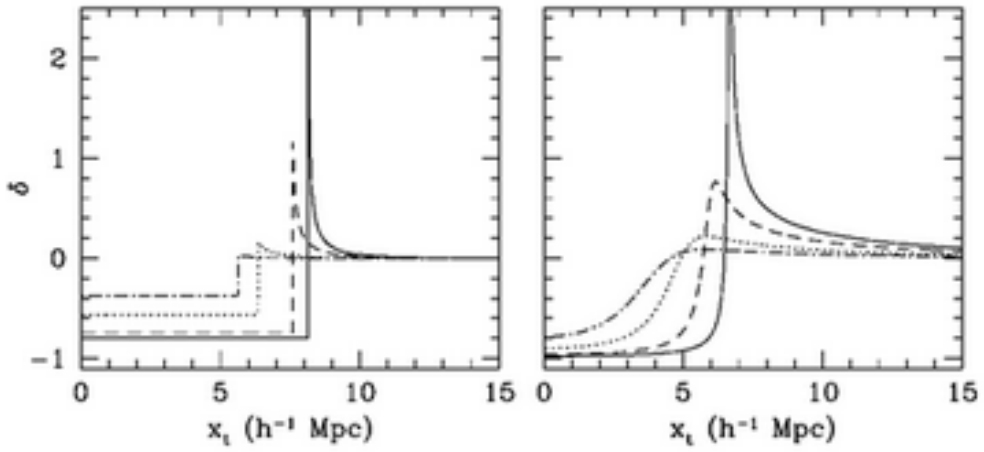

Figure 1.1: Spherical model for the evolution of voids. Left: pure uncompensated top-hat void evolving up to the epoch of shell crossing. Time steps: $a = 0.05, 0.1, 0.2, 0.3$. Right: void with angular averaged SCDM profile. The tendency of this void is to evolve into a top-hat configuration when shell crossing occurs. If we have a sufficiently steep initial profile we have shell crossing and the consequent formation of a ridge.[3]

uncompensated. These two kinds of voids have two different time dependence evolution of their shell mass and of their radius. In the first case shell mass grows as $t^{2/5}$ and radius as $t^{4/5}$; for uncompensated voids, instead, we have an increasing time dependence of shell mass and radius respectively as $t^{2/3}$ and $t^{8/9}$. These data show that uncompensated voids propagate more rapidly; the most important differences are due to the fact that considering mass deficit in the surrounding shells of a void energy increases with radius because binding energy of matter outside the void is lost with matter removed from the center, while energy input of the compensated void is constant at large distances, and the trajectories in front of the dense shell are no longer unperturbed Hubble flow.
The expressions of energy and scaled radius in both cases make evident why we have done these assumptions about voids evolution. The only independent initial parameters that can be used to express the physical quantities calculated studying voids formation and expansion are k, indicating shell width, the initial cosmic time t_i, the initial radius of the void R_i, the critical density at the initial cosmic time $\rho_i \equiv (6\pi G {t_i}^2)^{-1}$ and the initial density deficit $\delta_i \ll 1$. Then outside the void when there is no net mass deficit the value of the energy is

$$E = \frac{8\pi}{45} \frac{\rho_i {R_i}^5}{{t_i}^2} \delta_i g(k, \delta_i) \tag{1.1}$$

where the form factor g depends on void geometry and initial conditions but usually its value is of order unity. Instead, in case we are treating an uncompensated void, the

energy value explicitly increases with radius $r(t) = r_i(t/t_i)^{2/3}$ and its expression is

$$E = \frac{4\pi}{9}\frac{\rho_i {R_i}^3}{{t_i}^2}\delta_i r^2 \tau^{-4/3} \tag{1.2}$$

with $\tau \equiv t/t_i$. These energy values are useful to get the expressions of the self similar propagation of the radius in the two different cases. In particular the scaled radius for a compensated void is

$$\lambda \equiv r\left(\frac{Et^2}{\rho_H}\right)^{-1/5} \propto rt^{-4/5}. \tag{1.3}$$

Using this last equation, considering as E value that one of the equation 1.2, we have the definition of the radius of an uncompensated void:

$$r \equiv \lambda R(t), \qquad R(t) \equiv {\delta_i}^{1/3} R_i \tau^{8/9} = \left(\frac{9}{4\pi}\frac{Et^2}{r^2\rho_H}\right)^{1/3}. \tag{1.4}$$

Looking at the way of expansion of uncompensated voids it is possible that mass deficit in the surrounding shell put in danger the similarity solution influencing the mass shell to move outwards more rapidly than in the Einstein-De Sitter case, namely the hypothesis that the mean density relative to the critical density Ω tends to 0 as time goes by. The shell would comove with the matter in front of it, but initial conditions determine the internal density structure. Then similarity conditions regard shell propagation, but if it is rapid enough shocks or caustics may overtake fluid shells before the possible reaction to the mass deficit, so we have the Einstein-De Sitter $\Omega = 1$ condition for the dynamics of particles.
In compensated and uncompensated voids we have similarity solutions regarding different kinds of fluids: collisional gas, for example baryons, collisionless gas, massive neutrino or cold dark matter, or $\Omega \ll 1$ collisional gas in the gravitational potential of a $\Omega = 1$ collisionless solution. We choose to examine in more detail the uncompensated void with a collisionless gas solution because the approximate volume filling domains observed for most of the range of cosmological structure formation scenarios are constituted by primordial underdensities that developed in uncompensated voids through shell crossing, phenomenon that happens only in case of a collisionless solution. The total mass inside a sphere of initial radius $r_i \gg R_i$ is

$$m = \frac{4}{3}\pi\rho_i {r_i}^3\left(1 - \delta_i\frac{{R_i}^3}{{r_i}^3}\right) \equiv \frac{4}{3}\pi\rho_i {r_i}^3\Big(1 - \Delta\Big) \tag{1.5}$$

with $\delta_i(4\pi/3)\rho_i {R_i}^3$ as net mass deficit. Moreover the integrated mass for $\Delta \ll 1$ is

$$m = \frac{4}{3}\pi\rho_H R^3(t) M(\lambda). \tag{1.6}$$

M approaches the unperturbed solution $M = \lambda^3$ as $\lambda \to \infty$ but with an additional term representing the idea of mass deficit

$$M = \lambda^3 - \frac{3}{5} + O(\lambda^{-}1) \qquad \lambda \gg 1. \tag{1.7}$$

The term expressing finite total mass deficit is

$$m_d(t) \equiv \lim_{r\to\infty}\left[\frac{4}{3}\pi\rho_H r^3 - m(r,t)\right] = \frac{4}{5}\pi\rho_H R^3(t). \tag{1.8}$$

The term m_d increases with time as $t^{2/3}$, so it is not the same as the initial mass deficit $\delta_i(4\pi/3)\rho_i {R_i}^3$. But conservation of mass is not violated because m_d is the deficit as $r \to \infty$ and mass conservation is applied to a finite volume.
The equation of motion, derived from Newton's law

$$\frac{d^2r}{dt^2} = -\frac{Gm}{r^2}, \tag{1.9}$$

relative to a collisionless fluid particle around an uncompensated void is

$$\frac{d^2\lambda}{d\xi^2} = (1-2\eta)\frac{d\lambda}{d\xi} + \eta(1-\eta)\lambda - \frac{2}{9}\frac{M}{\lambda^2}, \tag{1.10}$$

where $\eta = \frac{8}{9}$ and with λ given in the 1.4, $\xi = \ln\frac{t}{t_0}$ the nondimensionalized time and M defined by

$$M \equiv m\left[\frac{4\pi}{3}\rho_H R^3(t)\right]^{-1} \propto mt^{-2/3}. \tag{1.11}$$

The initial conditions are

$$\lambda = \lambda_0, \qquad \frac{d\lambda}{d\xi} = V(\lambda_0) - \eta\lambda_0 \qquad at \qquad \xi = ln\left(\frac{t}{t_0}\right) = 0, \tag{1.12}$$

where $V(\lambda_0)$ is given by the expression

$$v \equiv \frac{R(t)}{t}V(\lambda) = \frac{dr}{dt} = \frac{r}{t}\frac{\sinh\theta(\sinh\theta - \theta)}{(\cosh\theta - 1)^2}, \tag{1.13}$$

with θ a dimensionless radial variable.
The particles are fixed but in comoving coordinates the void expands, so the fluid elements start to move to decreasing λ, the void soon catches up with the particles that pass through the shell and they move again to increasing λ because they are no more decelerated. We thus calculate from the equation of motion with the initial conditions given the following oscillating solution

$$\lambda = \lambda_0 e^{[(2/3)-\eta]\xi}. \tag{1.14}$$

Knowing ${\lambda_\pm}^{(i)}$ express caustics in the shell crossing phenomenon we have to consider $\lambda_0 > {\lambda_+}^{(1)}$, when shell crossing is not happened yet, because it occurs only when $\lambda < {\lambda_+}^{(1)}$, with $M(\lambda)$ given by the equation

$$M(\lambda) = M_0\Sigma_i(-1)^{i-1}\exp[-(3\eta - 2)\xi_i]. \tag{1.15}$$

where ξ_i is the value of ξ at the ith point where $\lambda = \lambda(\xi)$. In general there are few intersections for most values of λ, but the series converges rapidly if there are many. The fixed value of convergence as $\xi \to \infty$ is $\lambda_\infty = 1.11803$. The thickness of the collisionless shell and the mean density between ${\lambda_-}^{(1)}$ and ${\lambda_+}^{(1)}$ are

$$\frac{\Delta\lambda_s}{\lambda_s} \equiv \frac{{\lambda_+}^{(1)} - {\lambda_-}^{(1)}}{{\lambda_+}^{(1)}} = 7.20246 \times 10^{-2} \qquad \overline{D} = \frac{M({\lambda_+}^{(1)})}{{{\lambda_+}^{(1)}}^3 - {{\lambda_-}^{(1)}}^3} = 2.66487. \quad (1.16)$$

The shell is thinner and less dense than that around a compensated void because being the void uncompensated $M(\lambda) < \lambda^3$ for $\lambda > {\lambda_+}^{(1)}$. In an uncompensated void passes more mass because it expands more rapidly than a compensated one, despite the amount of matter passing through a caustic is less in an uncompensated void. Moreover in an uncompensated void caustics are weaker and will be smeared out if the collisionless gas has finite temperature before shell crossing and the pressure is less because the velocity dispersion in the shell is less.
The value of the thickness of the shell through a measure in redshift is almost twice that one in true radius, but it is still 37% less than the velocity thickness of the collisionless shell around a compensated void:

$$\frac{\Delta\lambda_V}{\lambda_V} = 0.137248. \quad (1.17)$$

An observer in the shell overestimates the Hubble constant by 33% with a dispersion less than 10%, so local measures of H_0 can be perturbed. The uncompensated void solutions for the different kinds of energy are

$$\frac{t_m}{E} = 1.21519, \qquad \frac{t_{th}}{E} = 3.63354 \times 10^{-3}, \qquad \frac{w}{E} = -0.218824, \quad (1.18)$$

with $E \equiv t_m + t_{th} + w$. The less velocity dispersion makes the thermal energy in an uncompensated void less than in the compensated case. The kinetic energy is dominated by the mean mass outflow.

1.3 Calculating density thresholds

The values of the density thresholds, calculated at the present epoch, whether for haloes or voids are useful in the studying of a spherical perturbation. Knowing that for an Einstein-De Sitter universe the critical density ρ_c in the universe is constant, when a density perturbation is $\rho > \rho_c$ we have an overdense perturbation evolving in time into a virialized cluster, instead if $\rho < \rho_c$ we have an underdense perturbation becoming a void as time goes by. The moment in which density perturbations become cosmic objects is given by the thresholds values. In order to calculate them we have to proceed first using a linear theory to express the value of the velocity of the perturbation, then using a non linear vision to achieve the equation of motion of the shells, that it is useful to determine if the perturbation continues to expand or not and when we have turnaround or collapse

for haloes and shell crossing for voids.
The density perturbation $\rho(\vec{x},t)$ has mean value $\overline{\rho}(t)$. The density contrast is

$$\delta(\vec{x},t) = \frac{\rho(\vec{x},t)}{\overline{\rho}(t)} - 1. \tag{1.19}$$

The conservation of mass at the linear order of the density contrast implies

$$\frac{\partial\delta(\vec{x},t)}{\partial t} + \frac{1}{a(t)}\nabla\cdot\vec{v}(\vec{x},t) = 0, \tag{1.20}$$

where a is the scale factor and $\vec{v}$ is the peculiar velocity field, namely the velocity of the perturbation relative to the velocity of background, typical of the expansion of the universe. The linear density contrast can be expressed as a product of two factors, one spatial and the other temporal, introducing the linear density growth factor $D(t)$:

$$\delta(\vec{x},t) = D(t)\delta(\vec{x}), \tag{1.21}$$

from which

$$\frac{\partial\delta(\vec{x},t)}{\partial t} = \frac{dD(t)}{dt}\delta(\vec{x}) = \frac{1}{D(t)}\frac{dD(t)}{dt}\delta(\vec{x},t) = \frac{d\ln D(t)}{dt}\delta(\vec{x},t) = \frac{d\ln D(t)}{d\ln a(t)}\frac{d\ln a(t)}{dt}\delta(\vec{x},t). \tag{1.22}$$

Defining $g(t) := d\ln D(t)/d\ln a(t)$ the linear growth rate, with $H(t) \equiv \dot{a}(t)/a(t)$ Hubble constant and using equation 1.20 we have

$$\nabla\cdot\vec{v}(\vec{x},t) = -g(t)a(t)H(t)\delta(\vec{x},t). \tag{1.23}$$

We integrate this last equation on a sphere of radius r and through the divergence theorem we obtain

$$\vec{v}(\vec{x},t) = -g(t)a(t)H(t)\frac{1}{3}\frac{3}{4\pi r^2}\int_{Vs}\delta(\vec{x},t)d^3x. \tag{1.24}$$

Moreover expressing the mean density contrast inside the radius r as

$$\Delta(r,t) := \frac{3}{4\pi r^3}\int_{Vs}\delta(\vec{x},t)d^3x \tag{1.25}$$

and substituting this in the equation 1.24 we obtain a direct relation between the radial velocity field and the mean density contrast inside the sphere:

$$v(r,t) = -\frac{1}{3}g(t)a(t)H(t)\Delta(r,t)r = -\frac{1}{3}g(t)H(t)\Delta(r,t)r(t). \tag{1.26}$$

Now we pass to the determination of the equation of motion of the shells, useful to study if a region results overdense or underdense and consequently how it evolves, considering spherical symmetric density fluctuations $\delta(r)$ with mean density contrast

$$\Delta(r) = \frac{3}{r^3}\int_0^r \delta(r')r'^2dr'. \tag{1.27}$$

The mass of the spherical perturbation of radius r is

$$M(r) = \frac{4\pi r^3}{3}\bar{\rho}(1 + \Delta(r)), \tag{1.28}$$

and fundamental it is to understand the distribution of mass taking the idea of Birkhoff's theorem, namely an object spherically symmetric influences the others as if all its mass would be concentrated in a point at the center, thus the perturbation can be considered point form.
Our goal is to obtain the parametric form of the expansion factor $a(t)$ and of the time t of the perturbation in the two cases of the overdense and underdense regions in order to express the values of the density thresholds of turnaround and collapse in the haloes scenario and of shell crossing in the voids scenario.
We start the reasoning with Newton's law

$$\frac{1}{2}\dot{r}^2 - \frac{GM(r)}{r} = cost. \tag{1.29}$$

We substitute the expression in the right hand side of the equation 1.28 to $M(r)$ in the equation relative to Newton's law, obtaining

$$\dot{r}^2 - \frac{8\pi G}{3}\bar{\rho}r^2(1 + \Delta(r)) = 2cost. \tag{1.30}$$

A particular case of this equation is that when we find the critical density $\rho_c = 3H^2/8\pi G$ with $H = \dot{a}/a = \dot{r}/r$ taking $\Delta = 0$ and $cost = 0$.
We define $\Omega_m := \bar{\rho}/\rho_c$ the parameter that compares the mean density of the perturbation $\bar{\rho}$ with the critical density of the universe, meanwhile the background density of the universe ρ is a different parameter. With these expressions we can rewrite the equation 1.30 as

$$\dot{r}^2 - \Omega_m H^2 r^2(1 + \Delta(r)) = 2cost. \tag{1.31}$$

At this point we want to express the initial velocity $\dot{r}$ in the last equation as sum of two contributes: the Hubble flux $\dot{r} = H_i r_i$, due to the expansion of the universe, and the peculiar velocity, namely the velocity of the perturbation respect to that one of the universe. At the initial time the peculiar velocity has a low value and so we can use the linear approximation of the equation 1.26 and omitting the time dependence we obtain

$$\dot{r}_i \cong r_i H_i - \frac{1}{3}g_i H_i r_i \Delta_i(r_i). \tag{1.32}$$

The equation 1.31 is a constant and so we can evaluate it at whatever time instant, in particular the initial one, therefore the expression for the initial velocity just seen can be used to rewrite the equation 1.31:

$$\begin{aligned}(r_i H_i)^2\left(1 - \frac{1}{3}g_i\Delta_i\right)^2 - (r_i H_i)^2(1 + \Delta_i)\Omega_i = 2cost \\ (r_i H_i)^2\left(1 - \frac{2}{3}g_i\Delta_i\right) - (r_i H_i)^2(1 + \Delta_i)\Omega_i = 2cost.\end{aligned} \tag{1.33}$$

Now we define two parameters:

$$\alpha_i := -\frac{2}{3}g(t_i)\Delta_i \qquad 1+\Delta_{ci} := \Omega_i(1+\Delta_i). \tag{1.34}$$

Substituting these in the previous equation we obtain:

$$(r_iH_i)^2(1+\alpha_i) - (r_iH_i)^2(1+\Delta_{ci}) = (r_iH_i)^2(\alpha_i - \Delta_{ci}) = 2cost. \tag{1.35}$$

Substituting this value to the right hand side of the equation 1.31 we have

$$\frac{\dot{r}^2}{r^2} = \frac{(r_iH_i)^2(\alpha_i - \Delta_{ci})}{r^2} + \Omega_m H^2(1+\Delta(r)). \tag{1.36}$$

For the conservation of mass it is valid that $M_i(r) = M(r)$ from which we can write the two following expressions:

$$(1+\Delta(r)) = \frac{\dot{r}^3}{r^3}(1+\Delta_i)\frac{\bar{\rho}_i}{\bar{\rho}} \qquad \frac{\bar{\rho}_i}{\bar{\rho}} = \frac{\rho_{ci}\Omega_i}{\rho_c\Omega_m} = \frac{{H_i}^2}{H^2}\frac{\Omega_i}{\Omega_m}. \tag{1.37}$$

Exploiting these relations the equation 1.36 can be expressed finally representing the equation of motion of the shells, assuming that the perturbation is spherically symmetric with an only one degree of freedom, the radius r which is the distance of the shells from the center of the overdense or underdense region:

$$\frac{\dot{r}^2}{r^2} = {H_i}^2\left(\left(\frac{r_i}{r}\right)^2(\alpha_i - \Delta_{ci}) + \left(\frac{r_i}{r}\right)^3(1+\Delta_{ci})\right). \tag{1.38}$$

The evolution of a region is uniquely determined by the initial over or under density perturbation at the initial radius r_i. The relation between the two parameters α_i and Δ_{ci} makes evident the destiny of the perturbation. If $\Delta_{ci} > \alpha_i$ the shell is closed, this means that the perturbation will turnaround and collapse; instead if $\Delta_{ci} < \alpha_i$ the shell is opened, namely the perturbation will continue to expand. We find a shell critic situation if $\Delta_{ci} = \alpha_i$.
The solution of the equation 1.38 can be expressed in a parametric form, first for the condition $\Delta_{ci} > \alpha_i$ (closed shell), as

$$\frac{r(\Theta)}{r_i} = \frac{1}{2}\frac{1+\Delta_{ci}}{\Delta_{ci}-\alpha_i}(1-\cos\Theta) \qquad t(\Theta) = \frac{1}{2}\frac{1+\Delta_{ci}}{\Delta_{ci}-\alpha_i}^{3/2}(\Theta - \sin\Theta), \tag{1.39}$$

second for $\Delta_{ci} < \alpha_i$ (opened shell) as

$$\frac{r(\Theta)}{r_i} = \frac{1}{2}\frac{1+\Delta_{ci}}{\alpha_i-\Delta_{ci}}(\cosh\Theta - 1) \qquad t(\Theta) = \frac{1}{2}\frac{1+\Delta_{ci}}{\alpha_i-\Delta_{ci}}^{3/2}(\sinh\Theta - \Theta), \tag{1.40}$$

where Θ is a dimensionless parameter defined as $d\Theta = \frac{r_i}{r}\sqrt{\frac{5}{3}\Delta_i}H_i dt$.
We choose to continue analyzing the system in an Einstein-De Sitter universe, so $\Omega = 1$ all time. With the approximation $g(\Omega) \approx \Omega^{0.55}$ the equation 1.38 becomes

$$\frac{\dot{r}^2}{r^2} = {H_i}^2\left(\left(\frac{r_i}{r}\right)^3(1+\Delta_i) - \frac{5}{3}\Delta_i\left(\frac{r_i}{r}\right)^2\right). \tag{1.41}$$

Its solutions, taking density contrast $\Delta_i \ll 1$, are

$$\frac{r}{r_i} = \frac{1}{2}\left(\frac{5}{3}\Delta_i\right)^{-1}(1-\cos\Theta) \qquad H_i t = \frac{1}{2}\left(\frac{5}{3}\Delta_i\right)^{-3/2}(\Theta-\sin\Theta) \tag{1.42}$$

for a $\Delta_i > 0$ closed shell and then

$$\frac{r}{r_i} = \frac{1}{2}\left(\frac{5}{3} \mid \Delta_i \mid\right)^{-1}(\cosh\Theta-1) \qquad H_i t = \frac{1}{2}\left(\frac{5}{3} \mid \Delta_i \mid\right)^{-3/2}(\sinh\Theta-\Theta) \tag{1.43}$$

for a $\Delta_i < 0$ opened shell with Θ imaginary in this second solution.
These solutions are now analyzed in both cases of overdense and underdense regions with the aim to find the threshold values that when overcome by the perturbations indicate the formation of a virialized cluster or of a void. The analysis of voids needs all the threshold values.
First we analyze the motion of an overdense shell, characterized by $\Delta_i > 0$. The perturbation achieves a point called of turnaround and then collapses to a virialized cosmological structure. In order to have the density threshold value of turnaround the velocity of expansion of the perturbation has to be null, therefore $\dot{r} = 0$. From the turnaround condition we use the equation 1.41 and the parametric solution of the space to obtain the following relation:

$$1 + \Delta_i = \frac{5}{3}\Delta_i\frac{r}{r_i} = \frac{1}{2}(1-\cos\Theta_{ta}). \tag{1.44}$$

Since the density contrast is $\Delta_i \ll 1$, the parameter of turnaround is $\Theta_{ta} \approx \pi$.
Using the first equation in 1.37 and Friedmann's third equation integrated, without pression, we obtain

$$1 + \Delta = (1+\Delta_i)\left(\frac{r}{r_i}\right)^{-3}\left(\frac{a}{a_i}\right)^{3}. \tag{1.45}$$

We rewrite this equation substituting the first parametric expression in 1.42 and using the equation $(a/a_i)^3 = (3/2H_i t)^2$ for an Einstein-De Sitter universe:

$$1 + \Delta = (1+\Delta_i)\frac{(\frac{3}{2}H_i t)^2}{\left[\frac{1}{2}(\frac{5}{3}\Delta_i)^{-1}(1-\cos\Theta)\right]^3}. \tag{1.46}$$

Remembering the parametric solution for time in 1.42 the last equation becomes

$$1 + \Delta = (1+\Delta_i)\frac{9(\Theta-\sin\Theta)^2}{2(1-\cos\Theta)^3}. \tag{1.47}$$

Now this equation is written in order to have the possibility to substitute to Θ the parametric turnaround value $\Theta_{ta} = \pi$ and finally to express the value for the contrast density at turnaround:

$$1 + \Delta_{ta} = (1+\Delta_i)\left(\frac{3\pi}{4}\right)^2 \simeq 5.552. \tag{1.48}$$

Therefore through the equation 1.45 we know that the comoving radius of the perturbation has shrunk by a factor $(1+\Delta_{ta})^{1/3} \simeq 1.771$.
If the perturbation evolved according to linear theory, favoured by $\Delta_i \ll 1$, we can expand to first order whether the expression 1.47 or the parametric equation for time in 1.42:

$$1+\Delta \simeq (1+\Delta_i)\left(1+\frac{3}{20}\Theta^2\right) \qquad H_i t \simeq \frac{1}{2}\left(\frac{5}{3}\Delta_i\right)^{-3/2}\frac{\Theta^3}{6}. \tag{1.49}$$

Taking these two equations, substituting the second in the first we obtain

$$1+\Delta \simeq (1+\Delta_i)\left[1+\Delta_i\left(\frac{3}{2}H_i t\right)^{2/3}\right]. \tag{1.50}$$

Finally, in order to have the value of the density contrast at turnaround in a linear theory we use the parametric solution for time in 1.42 evaluating it at $\Theta_{ta} = \pi$:

$$H_i t_{ta} = \frac{\pi}{2}\left(\frac{5}{3}\Delta_i\right)^{-3/2}. \tag{1.51}$$

Substituting this equation in 1.50 we obtain

$$1+\Delta_{ta} \simeq (1+\Delta_i)\left[1+\frac{3}{5}\left(\frac{3\pi}{4}\right)^{2/3}\right] \simeq 1+1.062, \tag{1.52}$$

where $\delta_{ta} = 1.062$ is the linear density threshold value for turnaround.
After turnaround we imagine that shells contract themselves to $r = 0$ with a collapsing parameter $\Theta_c = 2\pi$, so time of collapsing is linked to that one of turnaround by the relation $t_c = 2t_{ta}$. Analyzing the true process for these values of the parameters what really happens it is not a collapse but a virialization, namely a formation of a cosmic structure in equilibrium. The linear theory express the moment of collapse changing the equation 1.52 in

$$1+\Delta_c \simeq (1+\Delta_i)\left[1+\frac{3}{5}\left(\frac{6\pi}{4}\right)^{2/3}\right] \simeq 1+1.686, \tag{1.53}$$

where $\delta_c = 1.686$ is the linear density threshold value for collapse. The meaning is that when a density perturbation in a spherical volume, evolving linearly, overcomes a density value like δ_c the volume has collapsed forming a virialized object. There is only a problem as regards this idea, that is the phenomenon of cloud in cloud, but this will be solved studying the process with the excursion set approach.
Now we pass to calculate the density threshold for the evolution of an underdense region $\Delta_i < 0$. These regions do not turnaround, in fact imposing $\dot{r} = 0$ the unique solution is $\Theta = 0$ that corresponds to the starting moment of the process. Instead, they continue to expand until they are blocked by an overdense region on a bigger scale. When different shells corresponding to different initial radius crossing each other the system becomes non linear and there is the formation of a void. Therefore the distance between the shells

is null at a certain time with the occurrence of shell crossing: $dr = dt = 0$. So we derive the two parametric solutions 1.43:

$$dr = \frac{1}{2}\left(\frac{5}{3} \mid \Delta_i \mid\right)^{-1}\left[(\cosh\Theta - 1)\left(dr_i - r_i\frac{d\Delta_i}{\Delta_i}\right) + r_i \sinh\Theta d\Theta\right] \tag{1.54}$$

and

$$dt = \frac{1}{2H_i}\left(\frac{5}{3} \mid \Delta_i \mid\right)^{-3/2}\left[(\cosh\Theta - 1)d\Theta - \frac{3}{2}(\sinh\Theta - \Theta)\frac{d\Delta_i}{\Delta_i}\right]. \tag{1.55}$$

Imposing $dt = 0$ and $dr = 0$ from the first condition we obtain

$$d\Theta = \frac{3(\sinh\Theta - \Theta)}{2(\cosh\Theta - 1)}\frac{d\Delta_i}{\Delta_i}. \tag{1.56}$$

This $d\Theta$ value is used in the second condition $dr = 0$ in order to have the equation

$$(\cosh\Theta - 1)\left(dr_i - r_i\frac{d\Delta_i}{\Delta_i}\right) + r_i \sinh\Theta\frac{3(\sinh\Theta - \Theta)}{2(\cosh\Theta - 1)} = 0, \tag{1.57}$$

that can be expressed also in the following way:

$$\frac{d\ln\Delta_i}{d\ln r_i}\left[1 - \frac{3\sinh\Theta(\sinh\Theta - \Theta)}{2(\cosh\Theta - 1)^2}\right] = 1. \tag{1.58}$$

Deriving the expression 1.27 for the density contrast in case of a spherical symmetric perturbation we obtain

$$\frac{d\ln\Delta}{d\ln r} = 3\left(\frac{\delta(r)}{\Delta(r)} - 1\right). \tag{1.59}$$

Utilizing as density profile a top-hat depression, namely

$$\delta_i(r_i) = \begin{cases}\delta_0 & \text{for} \quad r_i < r_0 \\ 0 & \text{for} \quad r_i \geq r_0\end{cases}$$

$$\Delta_i(r_i) = \begin{cases}\delta_0 & \text{for} \quad r_i < r_0 \\ \delta_0(r_0/r_i)^3 & \text{for} \quad r_i \geq r_0\end{cases}$$

with density contrast $\delta_0 < 0$, we obtain

$$\frac{d\ln\Delta_i}{d\ln r_i} = \begin{cases}0 & \text{for} \quad r_i < r_0 \\ -3 & \text{for} \quad r_i \geq r_0\end{cases}$$

Inserting these values in the equation 1.58 in the case $r_i < r_0$ we have the banal solution $\Theta = 0$, while for $r_i \geq r_0$ the parameter of shell crossing Θ_{sc} is given by the equation

$$\frac{\sinh\Theta(\sinh\Theta - \Theta)}{(\cosh\Theta - 1)^2} = \frac{8}{9} \tag{1.60}$$

with solution $\Theta_{sc} \simeq 3.488$. Following now the same process made for the overdense regions we find the density value of shell crossing knowing that for underdense regions the next equation holds:

$$1 + \Delta = (1 + \Delta_i)\frac{9(\sinh\Theta - \Theta)^2}{2(\cosh\Theta - 1)^3} \tag{1.61}$$

Substituting to Θ the value $\Theta_{sc} = 3.488$ the density value of shell crossing is $1 + \Delta_{sc} \simeq 0.2047$. Therefore the comoving radius expanded by a factor $(1 + \Delta_{sc})^{-1/3} \simeq 1.697$. With underdense regions too if we want to have the linear density threshold value we have to expand the equations to first order like for overdense regions, obtaining

$$1 + \Delta \simeq (1 + \Delta_i)\left[1 - \mid \Delta_i \mid \left(\frac{3}{2}H_i t\right)^{2/3}\right]. \tag{1.62}$$

Modifying the parametric solution for time for underdense perturbations we find that

$$\left(\frac{3}{2}H_i t\right)^{2/3} = \left(\frac{3}{4}\right)^{2/3}\left(\frac{5}{3} \mid \Delta_i \mid\right)^{-1}(\sinh\Theta - \Theta)^{2/3}. \tag{1.63}$$

Substituting this expression in the equation 1.62 we solve the following relation for $\Theta_{sc} = 3.488$ in order to determine the linear density threshold value δ_v:

$$1 + \Delta_{sc} \simeq (1 + \Delta_i)\left[1 + \frac{3}{5}\left(\frac{3}{4}\right)^{2/3}(\sinh\Theta_{sc} - \Theta_{sc})^{2/3}\right] \simeq 1 - 2.717. \tag{1.64}$$

This negative density value $\delta_v = -2.717$ indicates that when a perturbation, linearly evolved, overcomes the value δ_v we have the formation of a void. The expansion of a void can be interpreted as a shell moving outwards self similarly, but we have to pay attention to the problems of void in void and void in cloud in order to do not miscount the real number of voids.

Chapter 2

Excursion set approach for haloes

In models concerning hierarchical gravitational instability in order to obtain from the statistics of the initial density field the mass function $n(M)dM$ of virialized objects have been used different formulations. The first derivation made by Press and Schechter (1974) agreed well with N-body experiments about mass functions. They calculated an equation for the mass function at any given time observing that a part of the mass in collapsed objects more massive than a mass M is related to a fraction of volume samples when the smoothed initial density perturbations are above some density threshold. However this derivation is not so convincing because they account for only half of the mean mass density of the universe in the expression of the mass function and they solve the problem by multiplying by a factor 2 their final result. This little rigorous way of proceeding is closely due to the cloud in cloud problem.
To solve the cloud in cloud we propose the so called theory of the excursion sets of $\delta(\boldsymbol{r}, R_f)$, the four dimensional initial density perturbation field smoothed with a continuous hierarchy of filters of radii R_f. The mass fraction of matter in virialized objects with mass greater than M is related to the fraction of space in which we smooth on some filter of radius greater than or equal to $R_f(M)$ the initial density contrast lying above a critical overdensity. Decreasing R_f at constant position r we obtain the differential mass function given by the rate of first upcrossings of the critical overdensity. We have the possibility to choose between different filter functions that influence the shape of the mass function: we can calculate the first upcrossing rate analytically with a sharp k-space filter, in which the field makes a Brownian random walk changing the resolution; we derive analytical upper and lower bounds for Gaussian or top-hat filters, though the mass function is calculated numerically generating an ensemble of field trajectories. Over a limited mass range all these filters fit well the N-body simulations, a part some imprecisions in identifying the particles which form low mass groups, although the sharp k-space is preferred.

2.1 Gaussian fields topology

We assume that the linear density fluctuations form a homogeneous and isotropic Gaussian random field $\delta(\boldsymbol{r}, z)$, with $\boldsymbol{r}$ the Lagrangian comoving position and z the redshift. Since hierarchical fields have structure on all scales we consider the $(N+1)$-dimensional field $\delta(\boldsymbol{r}, R_f)$ smoothed by convolving with some filter function $W(\mid \boldsymbol{r'} - \boldsymbol{r} \mid, R_f)$ on resolution scale R_f:

$$\delta(\boldsymbol{r}, R_f) = \int d\boldsymbol{r'} W(\mid \boldsymbol{r'} - \boldsymbol{r} \mid, R_f)\delta(\boldsymbol{r'}, 0) = \frac{1}{(2\pi)^3}\int d\boldsymbol{k}\tilde{W}(kR_f)\tilde{\delta}(\boldsymbol{k}, 0)e^{-i\boldsymbol{kx}} \qquad (2.1)$$

where $\tilde{W}(kR_f)$ is the Fourier transform of the filter and at each point $\boldsymbol{r}$ the smoothed field represents the weighted average of $\delta(\boldsymbol{r}, 0)$ over a spherical region of characteristic dimension R_f centred in $\boldsymbol{r}$. The statistical properties of the Gaussian field $\delta(\boldsymbol{r})$ are uniquely specified by its power spectrum $P(k)$ through the two-point function in Fourier space $\langle\delta(\tilde{\boldsymbol{k}}_1)\delta(\tilde{\boldsymbol{k}}_2)\rangle = (2\pi)^3\delta_D(\boldsymbol{k}_1+\boldsymbol{k}_2)P(k_1)$. Our Fourier transform convention is $\tilde{\delta}(\boldsymbol{k}) = \int d\boldsymbol{x}\delta(\boldsymbol{x})e^{i\boldsymbol{kx}}$ and δ_D and the brackets $\langle\rangle$ represent respectively the Dirac delta function and ensemble averaging. As regards $\delta(\boldsymbol{r}, R_f)$ its detailed properties depend by the specific choice of the filter function. The most used are the following three kernels: the top-hat filter $W_{TH}(\mid \boldsymbol{r} \mid, R_f) = 3\Theta(R_f - \mid \boldsymbol{r} \mid)/4\pi{R_f}^3$, with $\Theta(r)$ Heaviside step function, the Gaussian filter $W_G(r, R_f) = (2\pi{R_f}^2)^{-3/2}\exp(-x^2/2{R_f}^2)$ and the sharp k-space filter, convenient for the analysis, $W_{SKS}(k, R_f) = \Theta(k_f - k)$, where $k_f = 1/R_f$ and $k_f = \mid \boldsymbol{k}_f \mid$, that is the top-hat filter applied in momentum space. Since our goal is to obtain $n(M)dM$, namely the number density of objects in the mass range M to $M + dM$, it is important to associate a defined mass with R_f for each filter. While the association for real space top-hat filtering is easy, $M_{TH}(R_f) = 4\pi\rho_b{R_f}^3/3$, for the other filters is more difficult to assign a mass. One possibility is to multiply the average density by the volume enclosed by the filter, so we have $M_G(R_f) = (2\pi)^{3/2}\rho_b{R_f}^3$ for the Gaussian filter and $M_{SKS}(R_f) = 6\pi^2\rho_b{k_f}^{-3}$ for the sharp k-space one. An other possibility is to take $M_{SKS}(R_f) = 4\pi\rho_b{R_f}^3/3$, and similarly for the Gaussian case, where R_{TH} is chosen by requiring the equality between the variances of the density fields smoothed with a sharp k-space filter and a top-hat filter: $\sigma^2_{SKS}(R_f) = \sigma^2_{TH}(R_{TH})$. Through this last procedure we obtain correspondence with numerical simulations of clustering growth, where the non linear collapse up to some scale R_f of mass fluctuations forms first the objects of lower mass in a hierarchical bottom-up model of structure formation. Merging and accretion of subunits create haloes of larger mass. The mass distribution and the accretion of matter can be modelled by hierarchical structures formation processes, considering a decreasing resolution scale R_f, obtained by differentiating the equation 2.1:

$$\frac{\partial\delta(\boldsymbol{r}, R_f)}{\partial R_f} = \frac{1}{(2\pi)^3}\int d\boldsymbol{k}\tilde{\delta}(\boldsymbol{k}, 0)\frac{\partial\tilde{W}(kR_f)}{\partial R_f}e^{-i\boldsymbol{kx}} \equiv \eta(\boldsymbol{r}, R_f). \qquad (2.2)$$

This equation, similar to the form of a Langevin equation, shows that the term of the stochastic force $\eta(\boldsymbol{r}, R_f)$ equals the infinitesimal change of the resolution scale R_f influencing the value of the density fluctuation field $\delta(\boldsymbol{r}, R_f)$ in the position $\boldsymbol{r}$. As initial

condition for our first order stochastic differential equation we choose the limit $R_f \to \infty$ that corresponds to $\delta(\boldsymbol{r}, R_f) \to 0$. Therefore varying the resolution scale R_f we associate a trajectory $\delta(\boldsymbol{r}, R_f)$ to each point $\boldsymbol{r}$. Since equation 2.2 is linear the stochastic force is a zero mean Gaussian random field specified by its autocorrelation function

$$\langle \eta(\boldsymbol{r}_1, R_{f1}) \eta(\boldsymbol{r}_2, R_{f2}) \rangle = \frac{1}{2\pi^2} \int_0^\infty dk k^2 P(k) \frac{\partial \tilde{W}(kR_{f1})}{\partial R_{f1}} \frac{\partial \tilde{W}(kR_{f2})}{\partial R_{f2}} j_0(kr), \qquad (2.3)$$

with $r = \mid \boldsymbol{r}_1 - \boldsymbol{r}_2 \mid$ and $j_0(r)$ zeroth order spherical Bessel function. The correlation properties of the stochastic force $\eta(\boldsymbol{r}, R_f)$, namely of the Gaussian field $\delta(\boldsymbol{r}, R_f)$, influence statistically how are associated trajectories corresponding to different neighbouring points. Instead the coherence of each trajectory along the smoothing direction depends on the analytic form of the filter. In particular, using a sharp k-space filter the coherence vanishes, in fact we decrease R_f and add up new sets of Fourier modes of the unsmoothed distribution $\delta(\boldsymbol{r}, R_f)$, and since this is a Gaussian field each increment is independent of the previous ones and each trajectory $\delta(\boldsymbol{r}, R_f)$ becomes a Brownian random walk. The sharp k-space filter has an advantage in the simplification of the notation using as time variable the variance of the filtered density field, $\Lambda \equiv \sigma^2(k_f) = \langle \delta(k_f)^2 \rangle = (2\pi^2)^{-1} \int_0^{k_f} dk k^2 P(k)$, thus we reduce the stochastic process to a Wiener one $\partial \delta(\boldsymbol{r}, \Lambda)/\partial \Lambda = \zeta(\boldsymbol{r}, \Lambda)$ with $\langle \zeta(\boldsymbol{r}, \Lambda) \rangle = 0$ and $\langle \zeta(\boldsymbol{r}, \Lambda_1) \zeta(\boldsymbol{r}, \Lambda_2) \rangle = \delta_D(\Lambda_1 - \Lambda_2)$. The solution of this last Langevin equation in an arbitrary point of space with initial condition $\delta(\Lambda = 0) = 0$ is $\delta(\Lambda) = \int_0^\Lambda d\Lambda' \zeta(\Lambda')$ and averaging it we obtain $\langle \delta(\Lambda) \rangle = 0$ and $\langle \delta(\Lambda_1) \delta(\Lambda_2) \rangle = \min(\Lambda_1, \Lambda_2)$ which uniquely specify the Gaussian distribution $\delta(\Lambda)$.
The homogeneity and isotropy are not preserved in the resolution direction. We know that in linear theory the field $\delta(\boldsymbol{r}, R_f)$ grows in amplitude proportionally to the linear growth factor $D(t)$, that in universes dominated by nonrelativistic matter with $\Omega = 1$ can be expressed as $D(t) = (1+z)^{-1}$ in terms of the redshift z. But in our excursion set approach we have $\delta(\boldsymbol{r}, R_f)$ to be of fixed amplitude and a threshold decreasing with time instead of considering $\delta(\boldsymbol{r}, R_f)$ to be growing in amplitude relative to a fixed threshold δ_c. Therefore, using linear theory, we take $\delta(\boldsymbol{r}, R_f)$ as the initial linear density field extrapolated to the present epoch t_0 and a threshold that decreases with time as $\delta_c/D(t)$ with $D(t_0) = 1$. The gravitationally bound structures forming at any time t are to be identified with regions above some critical initial overdensity $t_f = \delta_c/D(t)$. For large filter radii the cumulative mass fraction, denoted by $\Omega_v(M)/\Omega$ where Ω is the density of the universe in units of the closure density and the mass M is a function of R_f, is easy to associate to regions which lay above the threshold. The problem is that at small filter radii we can have regions break up into equal areas above and below the threshold, even if they were above the threshold at larger filter radii. Therefore it is difficult to link the change in total volume, as a function of the resolution, of excursion regions with the mass in clumps at that scale. In the next section we will explain the study of random walks in the presence of a suitable absorbing barrier in order to obtain a plausible value of the mass function.

2.2 Absorbing barrier problem

The statistical properties of the linear Gaussian density field are the starting point in order to compute the comoving number density of collapsed haloes. The first developed theory was made by Press & Schechter: a fluctuation is part of a collapsed region of scale larger than $M(R_f)$ if at the same scale the value of the smoothed linear density contrast overcomes a threshold t_f. If the density perturbations evolve following a spherical top-hat model it is possible to have an exact value for the threshold t_f. Therefore with a fluctuation of amplitude δ collapsing at a redshift z_f we have $\delta(\boldsymbol{r}) = t_f = \delta_c/D(z_f)$. For general cosmologies δ_c depends on the density parameter, the cosmological constant, the Hubble constant, the redshift. Instead during the matter dominated area in the Einstein-De Sitter universe its value, given by $\delta_c \simeq 1.686$, does not depend on any cosmological parameter. Between the Press-Schechter formula and the N-body simulations was found good agreement, but as we said before the model by Press-Schechter has different problems: the cloud in cloud problem, namely to consider substructures of the same fluctuation belonging to haloes of different mass; the fact that, since in a Gaussian field only half volume is overdense, as we can see in the figure 2.2, with the Press-Schechter model we account only for half of the mass collapsed in objects. Multiplying the result by a factor 2 was the solution adopted by Press-Schechter. Instead we will consider excursion set mass functions in order to solve the cloud in cloud problem and the missing mass in the process of collapse.
Initially the density fluctuations are small and grow in amplitude as time goes by. Therefore it is easier the formation of big cosmological structures in recent times. Assuming linear theory, the field $\delta(\boldsymbol{r}, R_f)$ grows in amplitude proportionally with the linear growth factor $D(t)$. But, instead of considering the field $\delta(\boldsymbol{r}, R_f)$ growing in amplitude and the threshold δ_c to be fixed, we take $\delta(\boldsymbol{r}, R_f)$ fixed in amplitude and we choose as time dependent the value of the threshold. So we evaluate $\delta(\boldsymbol{r}, R_f)$ at present time t_0 and we decrease with time the threshold value thanks to the linear growth factor: $t_f(z_f) = \delta_c/D(z_f)$ with $D(0) = 1$ and z_f formation redshift. The value $\delta_c = 1.686$ is true only in the present time, while it is higher in more ancient times, because the density fluctuations are smaller and it is more difficult the formation of big cosmological objects, in fact we have a hierarchical structure scenario. The choice of the filter influences the form of the mass function. The solution of the problem is simplified for Brownian random walks, that is for sharp k-space filtered density fields. The jagged line in the figure 2.1 represents the overdensity centred on a randomly chosen position in the initial Gaussian random field, as a function of the scale on which the overdensity was computed. The height of the walk $\delta_0(\Lambda)$ is the linear density contrast relative to the density of the background universe. The spatial scale is parametrized by its variance Λ, that in hierarchical models decreases as the scale increases, so the largest spatial scales are on the left, and $\delta(\Lambda) \to 0$ as $\Lambda \to 0$.

In the spherical collapse model, all regions with linear theory densities greater than δ_c can have formed bound virialized objects, and the value of this critical overdensity does not depend of mass scale. The dotted line of the same height at all Λ_m in the figure 2.1 represents this constant value. No mass can escape from a collapsing region

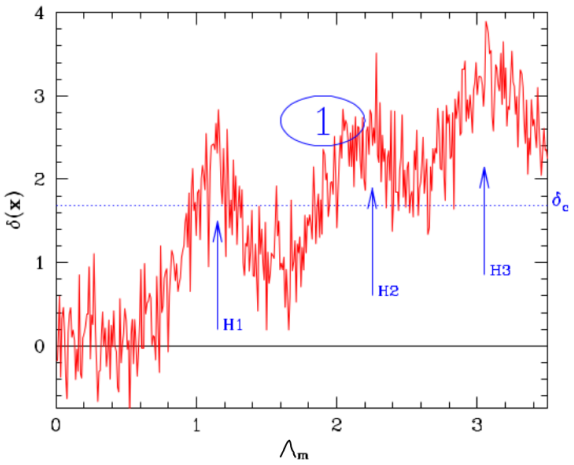

Figure 2.1: Excursion set formalism for the formation of a halo in the present time. Random walk exhibited by the average overdensity δ centred on a randomly chosen position in a Gaussian random field, as a function of smoothing scale, parametrized by the variance Λ_m. The large volumes are on the left and the small ones on the right. Dashed horizontal line indicates the collapse barrier δ_c. The largest scale, corresponding to the smallest value of Λ, on which $\delta(\Lambda)$ overcomes δ_c is an estimate of the mass of the halo that will form around that region.[3]

for the excursion set formalism. If $\delta_0 = \delta_c$ on scale R all the mass contained within R is included in the collapsed object, even if $\delta_0 < \delta_c$ for all $r < R$. Therefore if we have a collapsed object of mass $m \propto R^3$ means that the random walk height δ_0 exceeded the value δ_c after having travelled a distance $\Lambda(R)$. The threshold δ_c may be crossed at many different values of the variance and each crossing corresponds to a different smoothing scale and, since $m \propto R^3$, contains a different amount of mass. But the most meaningful crossing is the first one at the smallest value of $\Lambda(R)$ for which $\delta_0 \geq \delta_c$, since it is this scale which is associated with the most mass. The crossings at smaller scales correspond to condensations of a smaller mass, which have been incorporated in the larger encompassing mass concentration. The further a given walk travels before crossing the barrier, the smaller the mass of the object with which it is associated. So the computation of the mass function is equivalent to calculating the fraction of the trajectories that first upcross the threshold as the scale decreases.
Starting from the Gaussian distribution of particle positions

$$P(\delta)d\delta = \frac{d\delta}{\sqrt{2\pi\Lambda}} \exp\left(-\frac{\delta^2}{2\Lambda}\right), \tag{2.4}$$

one only has to solve the Fokker-Planck equation for the probability density $F(\delta, \Lambda)d\delta$ that the stochastic process at Λ assumes a value in the interval $\delta, \delta + d\delta$,

$$\frac{\partial F(\delta, \Lambda)}{\partial \Lambda} = \frac{1}{2} \frac{\partial^2 F(\delta, \Lambda)}{\partial \delta^2}, \tag{2.5}$$

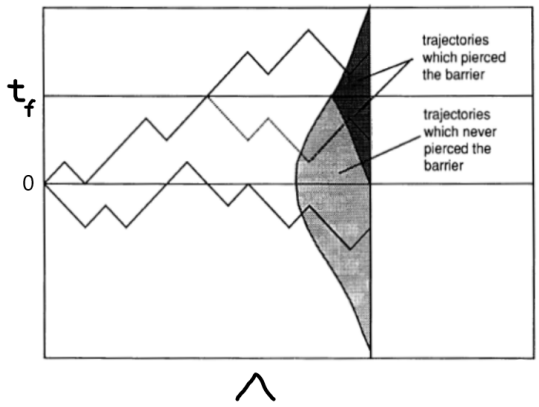

Figure 2.2: In this figure there are examples of random walks of $\delta(\boldsymbol{r}, \Lambda)$ with Λ. In the limit of infinitesimally fine steps in Λ the distribution of trajectories at fixed Λ is Gaussian as indicated by the curve on the right side. The fraction of the trajectories above the barrier $\delta = t_f$ at Λ are represented by the heavily shaded part of the Gaussian above the same barrier. If one reflects in the barrier the portion of the trajectory that lies to the right of where it first pierce the barrier we have the dotted trajectory, so the heavily shaded region below the barrier, reflection of the region above, represents the probability of a trajectory being below the barrier at Λ but having overcome the barrier at some lower value of Λ. The lightly shaded fraction of the area under the Gaussian represents those trajectories which have not overcome the barrier at any value of Λ.[2]

with the absorbing boundary condition $F(t_f, \Lambda) = 0$ and the initial condition $F(\delta, 0) = \delta_D(\delta)$. The solution is influenced by the fact that for each trajectory which reaches a field value above the threshold, there is an equally likely trajectory obtained by reflecting about the threshold the portion of the trajectory beyond the point where it first overcomes the threshold t_f, as we see in the figure 2.2:

$$F(\delta, \Lambda, t_f)d\delta = \frac{1}{\sqrt{2\pi\Lambda}}\left\{\exp\left(-\frac{\delta^2}{2\Lambda}\right) - \exp\left[-\frac{(\delta - 2t_f)^2}{2\Lambda}\right]\right\}d\delta. \tag{2.6}$$

This equation represents the paths which have not met the threshold, and the cumulative mass fraction is just twice the area in the tail of the Gaussian. Knowing that we can define the survival probability of the trajectories as $S(\Lambda, t_f) = \int_{-\infty}^{t_f} d\delta F(\delta, \Lambda, t_f)$ we obtain the density probability distribution of first crossing variances by differentiation:

$$\begin{aligned} f(\Lambda, t_f) &= -\frac{\partial S(\Lambda, t_f)}{\partial \Lambda} = -\frac{\partial}{\partial \Lambda}\int_{-\infty}^{t_f} d\delta F(\delta, \Lambda, t_f) = \\ &= \left[-\frac{1}{2}\frac{\partial F(\delta, \Lambda, t_f)}{\partial \delta}\right]_{-\infty}^{t_f} = \frac{t_f}{\sqrt{2\pi\Lambda^3}}\exp\left(-\frac{t_f^2}{2\Lambda}\right). \end{aligned} \tag{2.7}$$

The function $f(\Lambda, t_f)d\Lambda$ yields the probability that a random walk is absorbed by the barrier in the interval $(\Lambda, \Lambda + d\Lambda)$ or that a fluid element belongs to a structure with mass in the range $[M(\Lambda+d\Lambda), M(\Lambda)]$. At the end the comoving number density of haloes with mass in the interval $[M, M + dM]$ collapsed at redshift z_f is

$$n(M, z_f)dM = \frac{\rho_b}{M} f(\Lambda, t_f)\left|\frac{d\Lambda}{dM}\right| dM, \tag{2.8}$$

and substituting the expression found before for $f(\Lambda, t_f)$ we obtain the equation for the mass function identical to the Press-Schechter formula including the ad hoc multiplicative factor of 2 introduced by Press-Schechter in order to solve the problem of the missing of half of the mass collapsed:

$$n(M, z_f)dM = \frac{\rho_b t_f(z_f)}{\sqrt{2\pi}} \frac{1}{M^2\sqrt{\Lambda(M)}}\left|\frac{d\ln\Lambda}{d\ln M}\right| \exp\left(-\frac{t_f(z_f)^2}{2\Lambda(M)}\right) dM. \tag{2.9}$$

Where the factor 2 comes from corresponds to the mass density in collapsed structures contributed by the second term on the right hand side of equation 2.6. This is related to mass elements that lie below the threshold t_f at resolution $\Lambda(M)$ but exceed the threshold at a lower resolution and are therefore assigned to collapsed objects more massive than M. Press-Schechter, ignoring these trajectories, assert with a wrong derivation that the fraction of mass in objects of mass $\geq M$ is given by the fraction of mass above t_f at resolution $\Lambda(M)$.

Only for sharp k-space filtering it is possible to write an analytic formula for the mass function. Numerical solutions of the cloud in cloud problem using physically more plausible smoothing kernels, as the Gaussian and the top-hat, develop mass functions that are a factor of 2 lower in the high mass tail and have different small mass slopes compared with the analytical result using the sharp k-space filter. The consequence of this result is interpreted as the excursion set approach is not suitable for $M \ll M_*$, where M_*, defined by $\Lambda(M_*) = t_f^2$, is the typical mass collapsing at z_f. In particular for a nonsharp k-space filter we can no longer express the distribution at one time as a simple convolution of the distribution at a slightly earlier time with a fixed transition probability, so the process is non Markov. The determination of the first upcrossing rate is difficult since the non locality condition that the history of the trajectory should lie below the threshold, and the required distribution can be calculated only using a numerical Monte Carlo approach. However we can obtain analytical useful bounds on the fraction of trajectories which have yet to upcross the barrier using the rate of up and downcrossings found without considering the previous history. The integer quantities we can introduce in order to construct the analytical bounds are the number of times the trajectory upcrosses through t_f by resolution Λ denoted by $N_{up}(\Lambda, t_f)$, the number of times the trajectory downcrosses through t_f denoted by $N_{down}(\Lambda, t_f)$. Given these quantities we have $N_{ud}(\Lambda, t_f) \equiv N_{up} - N_{down}$, that is zero or one. The length of time between a downcrossing and an upcrossing is a measure of the coherence of the filter. The statistical averages of these three quantities are analytically tractable. The main aim for us is to find the operator $N_{1up}(\Lambda, t_f)$, which is zero until the first upcrossing occurs, at which it becomes

unity. In the general case we are obliged to calculate this operator only through numerical results. But with the aid of bounds for each trajectory as $N_{up}(\Lambda) \geq N_{1up}(\Lambda) \geq N_{ud}(\Lambda)$ and $N_{up}(\Lambda) - N_{up}(\Lambda - \delta\Lambda) \geq N_{1up}(\Lambda) - N_{1up}(\Lambda - \delta\Lambda)$ we can compare the numerical results for $d\langle N_{1up}\rangle/d\Lambda$ with analytical calculations for $d\langle N_{up}\rangle/d\Lambda$ and $d\langle N_{ud}\rangle/d\Lambda$. More useful are the following statistical rate operators, namely the derivatives of these integer operators:

$$R_{up} \equiv \frac{dN_{up}}{d\Lambda} = \sum_{\Lambda_{up}} \boldsymbol{\delta}(\Lambda - \Lambda_{up}) =\mid \dot{\delta} \mid \Theta(\dot{\delta})\boldsymbol{\delta}(\delta - t_f) \tag{2.10}$$

$$R_{down} \equiv \frac{dN_{down}}{d\Lambda} = \sum_{\Lambda_{down}} \boldsymbol{\delta}(\Lambda - \Lambda_{down}) =\mid \dot{\delta} \mid \Theta(-\dot{\delta})\boldsymbol{\delta}(\delta - t_f) \tag{2.11}$$

$$R_{ud} \equiv R_{up} - R_{down} = \dot{\delta}\boldsymbol{\delta}(\delta - t_f), \tag{2.12}$$

where Θ is the unit Heaviside step function, Λ_{up} denotes all those resolutions at which $\delta = t_f$ with $d\delta/d\Lambda > 0$ and Λ_{down} denotes the downcrossing resolutions $d\delta/d\Lambda < 0$.
We need all these expressions in order to define what we are interested in, that is the operator form of the statistical rate corresponding to the first upcrossing:

$$R_{1up}(\Lambda, t_f) \equiv \frac{dN_{1up}}{d\Lambda} = \boldsymbol{\delta}(\Lambda - \Lambda_{1up}) = R_{up}(\Lambda, t_f)\{1 - \Theta[N_{down}(\boldsymbol{r}, \Lambda)]\}. \tag{2.13}$$

A closed form of this expression cannot be given for general filters. We want to evaluate the average $\langle R_{1up}\rangle$, but an analytical solution is not possible. The operator in 2.13 can be expressed also in the following form:

$$R_{1up}(\boldsymbol{r}, \Lambda) = R_{up}(\boldsymbol{r}, \Lambda)\{1 - \Theta[N_{up}(\boldsymbol{r}, \Lambda-)]\}, \tag{2.14}$$

where by $\Lambda-$ we mean $\Lambda-\delta\Lambda$, with $\delta\Lambda$ positive and infinitesimal. Averaged and multiplied by $\delta\Lambda$ this equation expresses the product of the probability that the interval from 0 to $\Lambda-\delta\Lambda$ is devoid of upcrossing points and the number of upcrossing points in the interval from $\Lambda - \delta\Lambda$ to Λ. The latter is the probability that the infinitesimal interval has at least one upcrossing point taking small $\delta\Lambda$. The average that we want to calculate can be given in terms of the derivative with respect to Λ of the probability there is a void of upcrossing points in the interval from 0 to Λ:

$$\langle R_{1up}(\Lambda)\rangle = -\frac{\partial}{\partial\Lambda}P_{noup}(\Lambda), \qquad P_{noup} = 1 - \langle N_{1up}(\boldsymbol{r}, \Lambda, t_f)\rangle. \tag{2.15}$$

Taking the limit we have the void probability:

$$P_{noup} = \exp\left[\sum_{M=1}^{\infty} \frac{(-1)^M}{M!}\langle N_{up}^M\rangle_{cc}\right], \tag{2.16}$$

$$\langle N_{up}^M\rangle_{cc} \equiv \int_0^\Lambda d\Lambda_1\langle R_{up}(\Lambda_1)\rangle \cdots \int_0^\Lambda d\Lambda_M\langle R_{up}(\Lambda_M)\rangle\zeta^{(M)}(\Lambda_1, \ldots, \Lambda_M), \tag{2.17}$$

where $\zeta^{(M)}$ is the reduced continuous M-point correlation function of R_{up}. Therefore the mean of the 1-up density is

$$\langle R_{1up}(\Lambda)\rangle\delta\Lambda = \delta\Lambda \sum_{M=0}^{\infty} \frac{(-1)^M}{M!}\langle R_{up}(\Lambda)N_{up}^M(\Lambda)\rangle_{cc} \exp\left[\sum_{M=1}^{\infty} \frac{(-1)^M}{M!}\langle N_{up}^M(\Lambda)\rangle_{cc}\right] \quad (2.18)$$

$$= \delta\Lambda\langle R_{up}e^{N_{up}}\rangle_{cc} \exp[-\langle 1-\exp(-N_{up})\rangle]. \quad (2.19)$$

That it is impossible to find an accurate analytical expression for the statistical rate operator relative to the first upcrossing it is due to the fact that we made little progress to calculate this beyond the level of the two point function. Peacock and Heavens gave a well motivated analytical description of the numerical results about the first upcrossing through the following starting expression:

$$\langle N_{1up}(\Lambda_j)\rangle \approx [1-\langle N_{ud}(\Lambda_j)\rangle] \prod_{i=1}^{j-1}[1-\langle N_{ud}(\Lambda_i)\rangle], \quad (2.20)$$

supposing the resolution intervals $\delta\Lambda_i = \Lambda_i - \Lambda_{i-1}$ to be large enough in order to make the trajectories in each interval uncorrelated with the adjacent ones.
Despite these difficulties in analyzing the systems analytically with general filters and the differences in the low mass part of the spectrum between the filters approximation and the N-body simulation, that shows to prefer the sharp k-space filter formula, our formalism is able to calculate analytically for sharp k-space filters the probabilities of merging of objects at two redshifts z_1 and z_2 considering two thresholds $t_{f1} = \delta_c/D(z_1)$ and $t_{f2} = \delta_c/D(z_2)$ as a two barrier absorption problem and through Monte Carlo simulations, using a large number of paths, for general filters in terms of a correlation function of the first upcrossing rate at contour level t_{f2},$R_{1up}(\Lambda_2, t_{f2})$, and the first upcrossing rate at contour level t_{f1},$R_{1up}(\Lambda_1, t_{f1})$. For sharp k-space filter we consider the fraction of trajectories which make their first upcrossing of the lower threshold t_{f2} at resolution $\Lambda(M_2)$ and then upcross through the higher threshold t_{f1} for the first time at $\Lambda(M_1)$. We interpret this as the fraction of mass that was in objects of mass M_1 at z_1 that is later in objects of mass M_2 at z_2. We need to find the solution $F(\delta, \Lambda \mid \Lambda_2, t_{f2})$ for the constrained probability of δ at Λ given t_{f2} at Λ_2 that can be expressed in terms of the solution 2.6 of equation 2.5, but in the variables $\delta - t_{f2}$ and $\Lambda - \Lambda_2$: $F(\delta - t_{f2}, \Lambda - \Lambda_2, t_{f1} - t_{f2})$. Knowledge of this quantity makes it possible to calculate many statistics concerning the merging of objects.

2.3 Excursion set theory related to N-body

We know that the Gaussian and the top-hat filter functions give fewer high mass objects than the Press-Schechter formula, that corresponds to the sharp k-space prediction, but the biggest differences between the curves relative to the mass function are at the low mass end. Our aim in this section is to understand if these filter functions provide a reasonable fit to the N-body mass function simulations. These simulations begin with

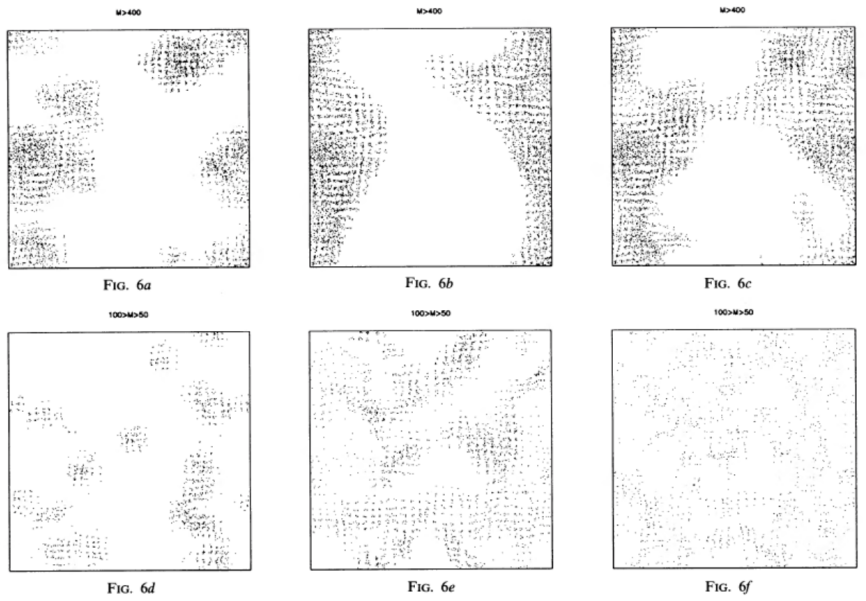

FIG. 6a FIG. 6b FIG. 6c

FIG. 6d FIG. 6e FIG. 6f

Figure 2.3: Initial positions of various subsets of particles in two different ranges of mass in the N-body simulation in (a) and (d), using a sharp k-space filter in (b) and (e) and using a Gaussian filter in (c) and (f).[2]

Gaussian initial fluctuations with power law spectra $P(k) \propto k^{3+n}$ superposed on a flat $\Omega = 1$ background. Also in the N-body simulation convolving the initial density field we can smooth the initial conditions with a set of filters, starting with the filter with the largest R_f, continuing until the smoothed density exceeds the threshold value δ_c/a, with a factor of universe expansion since the start of the simulation and of growth of the linear theory density field, when we assign the filter mass $M(R_f)$ to this particle. To assign a mass to each particle we must adopt a mass filter radius relation, that for the case of a top-hat filter is the mass contained within the filter $M = 4\pi\Omega\rho_{cr}R_{TH}^3/3$, and a value for the linear density threshold t_f, that we choose in the present time $t_f = \delta_c = 1.686$. For other filtering choices we fixed the relation between R_f and R_{TH} by requiring that the relation $M(\Lambda)$ has to be independent of the choice of filter function. Starting from these mass assignments we construct group catalogs as results of the N-body simulations. Most of the groups contain one dominant density center, but sometimes percolation selects a group with two or more distinct density centers joined by bridges. Groups containing only few particles which are often found in the outer haloes of very massive groups can be ignored because due to statistical fluctuations in the poorly sampled density field.
Now we compare the group catalogs that come out assigning mass to each particle using a filter function or the N-body simulation. The figures $6a, 6b, 6c$, respectively for the N-body simulation, the sharp k-space filter and the Gaussian filter, show regions of space occupied by the initial density perturbations which collapse to form massive groups in a range of high mass. For both the filter functions the region selected does not corre-

spond so much with the figure $6a$. There is a void running up the center with the largest structures forming from material initially at the edges of the box. Moreover the filter schemes tend to miss the outer regions of some of the density perturbations and also form bridges between them. But after all analyzing the figures $7a, 7b, 7c$ corresponding to the final positions of these sets of particles we note that the massive groups in the N-body simulation of figure $7a$ are well represented in figures $7b, 7c$ using sharp k-space and Gaussian filters indicating a good correspondence between the two different methods. Instead for massive groups of particles which end up in lower mass groups, like those plotted in their initial positions in the figures $6d, 6e, 6f$ and in their final positions in the figures $7d, 7e, 7f$, the degree of correspondence between the regions of space occupied by the density perturbations in the N-body simulation and regions of space selected by the two filtering choices is not good. Therefore filtering on a mass scale $M \geq M_*$ we obtain similar final conditions between the regions our filtering schemes select and the regions which actually form groups of the corresponding mass in the N-body simulation. Instead for $M \ll M_*$, where objects are constructed out of the mass left over by the larger ones, the filtering schemes show bad correspondence with the N-body simulation. This is an inevitable problem in any local filtering scheme, and cannot be avoided by judicious choice of filtering function. Thus it seems reasonable to abandon the excursion set approach adopting non local criteria to study the formation of structures locating mass around peaks in the initial density field. But these non local criteria are too difficult to develop analytically and it is too elaborate to describe the formation of low mass objects accurately. Therefore the excursion set approach is adapt to reproduce the N-body simulations, fitting well the evolution of groups catalogs inside a restricted range of masses, especially at the high mass end.

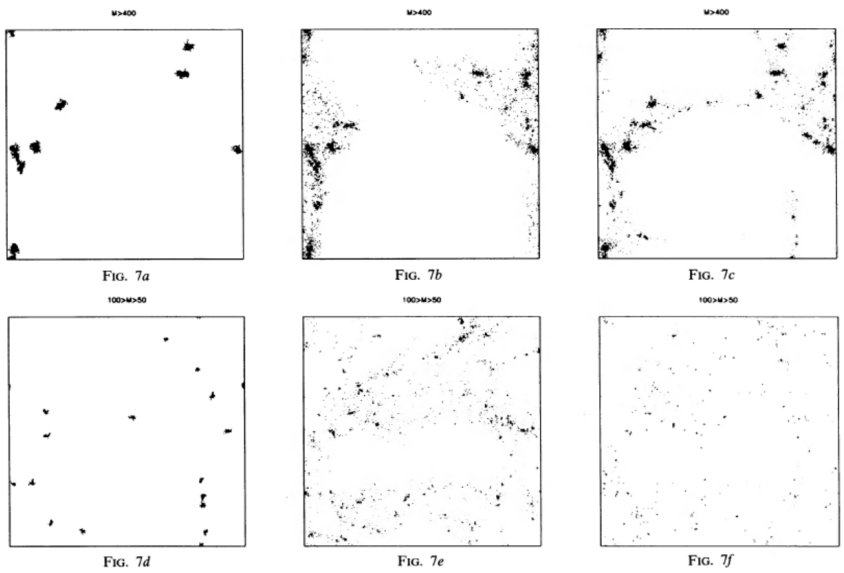

Figure 2.4: Final positions of various subsets of particles in two different ranges of mass in the N-body simulation in (a) and (d), using a sharp k-space filter in (b) and (e) and using a Gaussian filter in (c) and (f).[2]

Chapter 3

Voids hierarchy

The size distribution of voids, contrarily to the distribution of virialized halo masses which does not have a small scale cut off, is well peaked about a characteristic void size that evolves self similarly in time. In the excursion set approach of cluster abundance and evolution the determination of the mass function requires the study of random walks crossing one barrier with the only complication of the cloud in cloud problem, namely counting as clusters those objects which are not embedded in larger clusters. In this chapter we will use this same formalism to calculate the mass function and the evolution of voids, but we need the study of a two barrier problem, because two processes rule the destiny of voids: one barrier accounts for the void in void problem, that is the process that affects voids which are embedded in larger underdense regions, a larger void made up by merging smaller voids, like when in the cloud in cloud we have clusters forming from the mergers of less massive progenitors; the second barrier is unique to voids and accounts for the void in cloud problem, that happens when voids are embedded within a larger scale overdensity, that squeezes voids out of existence as it collapses around them. This last process produces the cut off at small scales and so it is very important to understand how to treat the relation between the voids inside the collapsing clouds around them and the clouds themselves to reproduce in the best possible way the distribution of void masses, especially at the low mass end. Therefore we analyze a model of considering the void in cloud problem that improves the vision by Sheth and van de Weygaert in [3] studying voids in terms of the late time field, relating Eulerian volumes in the late time field and Lagrangian ones in the initial field.

3.1 Void sociology

Voids are the dominant component of the megaparsec scale galaxy distribution and they fill most of the universe, while much of the mass in the universe is squeezed in between expanding adjacent colliding voids, confined to tangential motions. We have seen that the void is a bottom up assembly in which it forms through the fusion and erasure of its internal substructure evolving towards a spherical top-hat geometry. The progressive fading of these structures, while matter evacuates towards the enclosing boundary of the

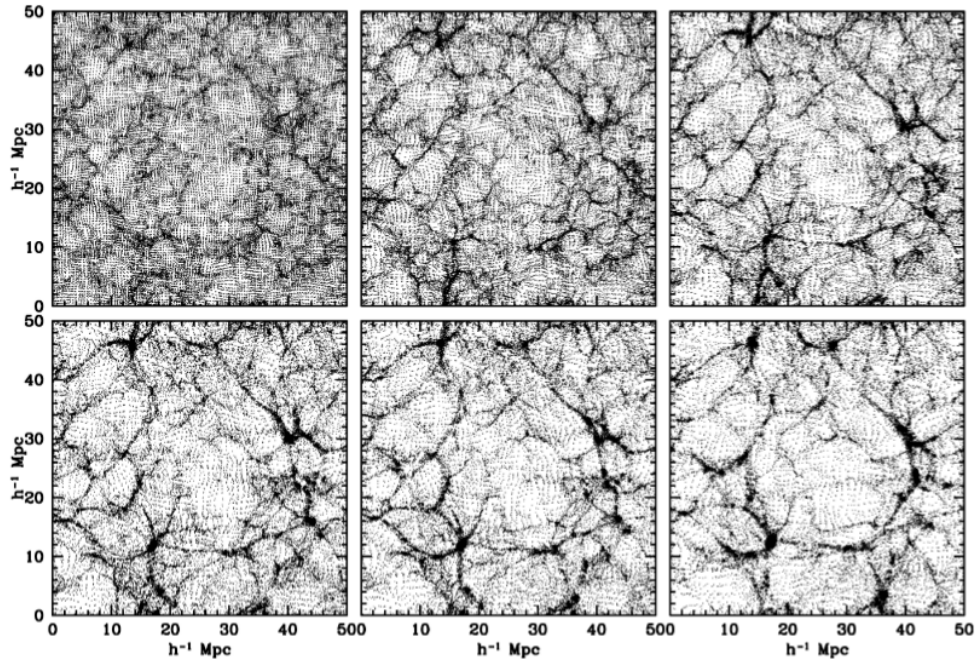

Figure 3.1: Six time steps in the evolution of a void region in a 128^3 particle N-body simulation of structure formation in an SCDM model: top left to bottom right shows expansion factors $a_{exp} = 0.1, 0.2, 0.3, 0.35, 0.4, 0.5$ (the present time has $a_{exp} = 1$). It is illustrated the gradual development of a large void of diameter $\approx 25h^{-1}Mpc$ as the smaller voids within it merge with one another at an earlier time: this is the void in void process.[3]

void, marks the subsequent merging of voids. The void itself approach non linearity in the same time scale his internal substructure is erased, in fact at non linearity small voids merge with one another into one larger void dissolving their separate entities. This natural void hierarchy in which small scale voids merge into a larger one is one of the two effects that characterize the interaction of voids with their surroundings. The figure 3.1 illustrates this scenario. We call this the void in void problem, namely, analyzing the system in the present time, if we have many small regions less dense than the critical threshold of formation of a void δ_v within one larger region less dense δ_v too we must not consider the smaller voids as distinct objects in order to do not overestimate the number of small voids and the total volume fraction in voids. This is analogous to the cloud in cloud problem for overdense regions. But to understand the complex spatial patterns in the universe it is extremely important to study also the second effect that seriously affects the distribution void masses, especially the number of small voids, and that represents a radical dissimilarity between voids and haloes. We call this the void in cloud problem. If we have an overdense perturbation surrounding an underdense one the collapse of the first can squeeze the second so much that it should no longer be counted

as a void. The figure 3.2 shows three examples of this process. Thus we understand that the excursion set formulation of voids needs consideration of two barriers: one associated with the collapse of clouds and the other with the formation of voids. The asymmetry between voids and haloes evolution is evident also from the fact that the cloud in void phenomenon is not important for dark halo formation because virialized haloes within voids are not to be torn apart as the void expands around them.
The values of the critical densities $\delta_c = 1.686$ and $\delta_v = -2.717$ establish the two thresholds that define when overdense and underdense regions can be considered respectively haloes and voids in the present time. So if a random walk δ_0 first overcomes the value δ_c after having travelled a distance $\Lambda(R)$ it represents a cloud of mass $m \propto R^3$ and physical size $\boldsymbol{R} \approx R/1.771$, while if it first drops below the value δ_v after having travelled a distance $\Lambda(R)$ it represents a void of mass $m \propto R^3$ and physical size $\boldsymbol{R} \approx 1.7R$. But we know that the processes of embedding make things more complicated, therefore the excursion set formalism shows how to analyze the evolution of overdense and underdense regions in the correct way.
The figure 3.3 illustrates with four sets of panels the random walk associated with the initial particle distribution through the leftmost column and the same distribution at two later times in the other two columns. The first row regards the cloud in cloud process. The mass of the final object, represented in the third plot, formed when the linear theory variance has value $\Lambda = 0.55$, is given by the merging of the smaller clumps, represented in the second plot, formed at earlier times, when the threshold, intercepted by the random walk at bigger variance values ($\Lambda > 0.55$), was $\delta_c/D(z) > \delta_c$.
The second row illustrates the cloud in void process. We have a low mass clump virialized at early times, $\Lambda > 0.85$, embedded in a region destined to become a void at $\Lambda = 0.4$ containing significantly more mass than the low mass clump. The cloud inside the void is not destroyed by the expansion of the void itself, in fact the mass of the cloud slightly increases from $\Lambda > 0.85$ to $\Lambda = 0.85$. So in this case the random walk represents an halo at $\Lambda = 0.85$ and we have the certainty that to determine halo abundances we do not need to know anything about the barrier δ_v.
The third row shows the void in void process, namely the formation of a large void by the merger of smaller ones. In this case the random walk seems to be the inverse of that for the cloud in cloud process. The void at the present time $\Lambda \sim 0.4$ contains more mass than it did in the past ($\Lambda > 0.4$). Instead a random walk centred on one of the mass elements that constitute the filaments within the large void can be described like a cloud in void walk.
Finally, the fourth row regards the void in cloud process. This last one describes a ring of objects that collapses around a large void formed at earlier times squeezing it to a much smaller size. Comparing this process to the cloud in void would be wrong because the void would be considered a relatively large object containing mass at $\Lambda = 1$ inside a cloud containing much more mass at $\Lambda \sim 0.3$, and it is absurd that a so massive cloud virializes around a large void without squeezing it.
Considered all these situations the excursion set formulation associates voids only to those walks that cross δ_v without first crossing at larger scales δ_c, because if a walk first

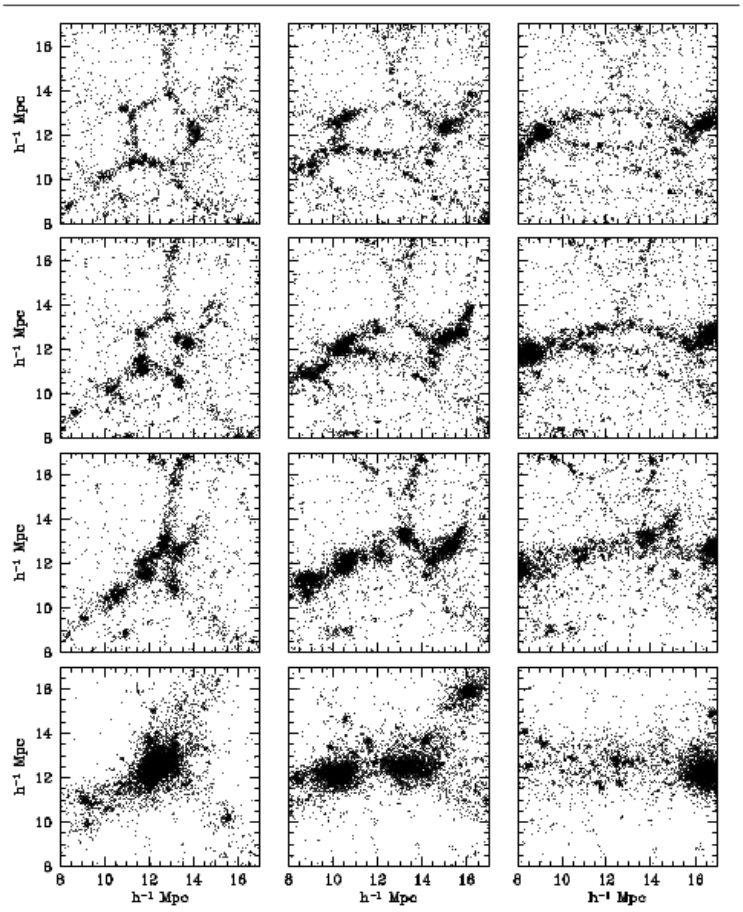

Figure 3.2: Top to bottom panels in three different examples from left to right of the void in cloud process show the evolution of the particle distribution in comoving coordinates from early to late times (respectively $a = 0.3, 0.4, 0.5, 1.0$, the current epoch) in numerical simulations of structure formation in an SCDM scenario ($\Omega_0 = 1.0, h_0 = 0.5$). The initially underdense regions are crushed by the collapse of the overdense regions that surround them.[3]

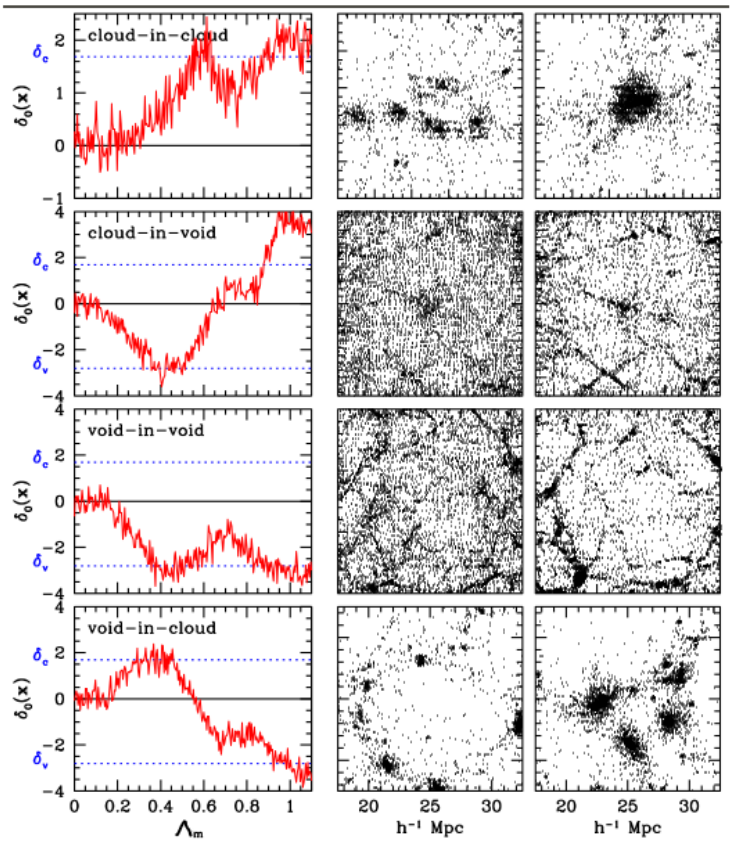

Figure 3.3: From top to bottom each row illustrates one of the four basic modes of hierarchical clustering: the cloud in cloud, the cloud in void, the void in void, the void in cloud. Each mode is described using three panels. The leftmost shows the random walk associated to the local density perturbation $\delta_0(\boldsymbol{r})$ as a function of mass resolution scale Λ_m, with the dashed horizontal lines indicating the collapse barrier δ_c and the shell crossing void barrier δ_v. The other two panels on the right illustrate how the particle distribution evolves.[3]

overcomes δ_c before to cross δ_v on a smaller scale the larger overdense region becomes a cloud that collapses around the smaller void within it squeezing it. The fraction of random walks that first cross δ_v at Λ and did not cross δ_c at any $\Lambda' < \Lambda$ determines the fraction of mass in voids. Now is evident the reason why the void hierarchy is a two barrier problem.
However the analysis of the void in cloud problem is more complicated. Choosing $\delta_c = 1.686$ as threshold value associated with complete collapse we allow the void to have the maximum possible size, unless it is within a fully collapsed halo, in which case it has zero size. Instead, if the void is embedded in a region that has not yet collapsed completely its size is intermediate between the maximum size that one would have estimated in case of an isolated spherical evolution and zero. Therefore we overestimate the typical void size if we do not count only voids within regions collapsed completely. If we consider an other threshold value $\delta_{ta} = 1.062$, the one that defines the moment of turnaround, namely considering regions that achieve the collapsing phase but that have not yet started collapsing, we ignore all voids that are within these kind of regions, even though they may still have not negligible sizes, and so we underestimate the abundance of large voids. So to determine the distribution and size of voids is not so obvious, especially at the low mass end where the void in cloud problem is more significant. While haloes within voids may be observable, voids within collapsed haloes are not.

3.2 Void size distribution

In order to study the abundance of voids we start following the spherical evolution model described by Sheth and van de Weygaert with critical density threshold defined to be when the expanding shells of the voids cross, namely when for an Einstein-de Sitter universe the non linear average density within the void reaches $\rho_v = 0.2\rho_m$, with ρ_m background matter density, or when the linear density threshold reaches $\delta_v = -2.717$.
In calculating the first crossing distribution for voids $f(\Lambda, \delta_v, \delta_c)$ Sheth and van de Weygaert proposed that the excursion set approach applied to voids needs a second barrier, the threshold δ_c for collapse of overdense regions, which should lie in between $\delta_c = 1.06$, the value at turnaround in the spherical collapse model, and $\delta_c = 1.686$, the value at the point of collapse. The voids counted through this method are those associated to trajectories that cross the barrier δ_v given that they have not overcome δ_c on any larger scale, so the distribution $f(\Lambda, \delta_v, \delta_c)$ represents those random walks that satisfy both these conditions. We have to subtract to the distribution of voids $f(\Lambda, \delta_v)$ without considering the void in cloud process those voids subjected to the void in cloud process:

$$f(\Lambda, \delta_v, \delta_c) = f(\Lambda, \delta_v) - \int_0^{\Lambda} f(\Lambda', \delta_c) f(\Lambda, \delta_v \mid \Lambda', \delta_c) d\Lambda'. \tag{3.1}$$

The last term in the integral is the product of the fraction of all the trajectories that overcome δ_c at the scale $\Lambda' < \Lambda$ with the fraction of the trajectories that cross δ_v at scale Λ if they overcame δ_c at Λ'.
The Sheth and van de Weygaert formula for the abundance of voids in linear theory is

given by:

$$\frac{dn_L}{d\ln M} = \frac{\rho_m}{M} f(\Lambda, \delta_v, \delta_c) \frac{d\Lambda}{d\ln M}, \tag{3.2}$$

where the first crossing distribution is the expression

$$\Lambda f(\Lambda, \delta_v, \delta_c) = \sum_{j=1}^{\infty} \frac{j^2\pi^2 D^2}{\delta_v^2/\Lambda} \frac{\sin(j\pi D)}{j\pi} \exp\left(-\frac{j^2\pi^2 D^2}{2\delta_v^2/\Lambda}\right). \tag{3.3}$$

The quantity D is the void and cloud parameter and it is defined as

$$D := \frac{\mid \delta_v \mid}{\delta_c + \mid \delta_v \mid}. \tag{3.4}$$

It parametrizes the impact of halo evolution on the evolving populations of voids, in fact the probability of smaller voids being crushed through the void in cloud process decreases as the relative value of the collapse barrier δ_c becomes larger respect to the void barrier δ_v. This value is useful to calculate the mass fraction in voids given by

$$\int d\Lambda f(\Lambda, \delta_v, \delta_c) = 1 - D = \frac{\delta_c}{\delta_c + \mid \delta_v \mid}. \tag{3.5}$$

If $\delta_c \gg \mid \delta_v \mid D$ is small and voids account for nearly all the mass in the universe, because it is more difficult to overcome δ_c. Instead, if the void in cloud process becomes more and more important the mass fraction in voids is less than unity and voids are squeezed to vanishingly small size, in fact for $\delta_c \ll \mid \delta_v \mid D$ is big and almost all the mass in the universe is contained in the haloes.
In an Einstein-de Sitter universe the values $\delta_c, \delta_v, \sigma(m)$ have the same time dependence, so the equation 3.3 evolves self similarly, that is a quite accurate approximation because in more general world models the time dependences are only slightly different.
The equation 3.3 converges rapidly but, since it is difficult to appreciate the associated shape, for values of $\delta_c / \mid \delta_v \mid \geq 1/4$ of so it is well approximated by

$$\nu f(\nu) \approx \sqrt{\frac{\nu}{2\pi}} \exp\left(-\frac{\nu}{2}\right) \exp\left(-\frac{\mid \delta_v \mid}{\delta_c} \frac{D^2}{4\nu} - 2\frac{D^4}{\nu^2}\right), \tag{3.6}$$

where we have set $\nu \equiv \delta_v^2/\Lambda \equiv \delta_v^2/\sigma^2(m)$ and $\nu f(\nu) d\nu/\nu = \Lambda f(\Lambda) d\Lambda/\Lambda$. Figure 3.4 shows the resulting void size distribution. The mass fraction in small voids depends strongly on δ_c, whereas the mass fraction enclosed by the largest voids depends only on δ_v. In the figure 3.4 are plotted three curves relative to three different values of δ_c with fixed δ_v. The dotted curve, obtained with $\delta_c \to \infty$, represents the process of ignoring the void in cloud. This distribution overestimates enormously the number of small voids. In fact for $\delta_c \gg \mid \delta_v \mid$ the second exponential in the equation 3.6 tends towards unity, then the abundance of voids is given by accounting only for the void in void process. The divergence at low ν is removed decreasing δ_c and for the other two curves, corresponding to the other two values of δ_c chosen, $f(\nu)$ cuts off sharply at both small and large values

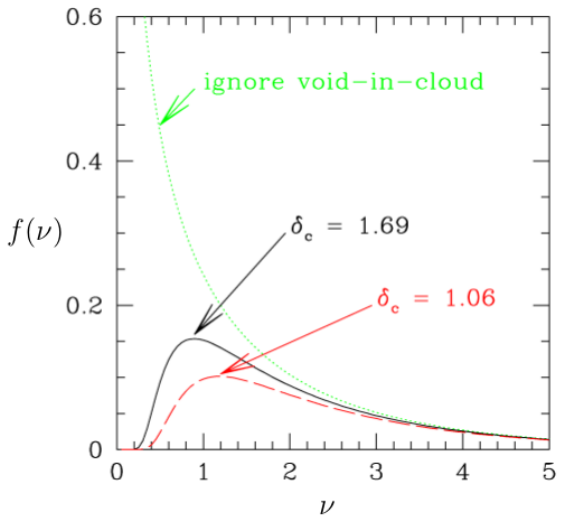

Figure 3.4: Scaled distribution of void masses/sizes: voids that enclose large masses have large values of ν. Curves show equation 3.3 with $\delta_v = -2.717$ and two choices of δ_c as labelled.[3]

of ν and the distribution is well peaked about $\nu \approx 1$. It is evident that the distribution of large voids does not depend on δ_c, in fact to ignore the process of void in cloud does not overestimate the number of large voids and consequently all the curves are similar at large values of ν. This is a consequence of the fact that large underdensities embedded in a larger region of average density are rare, so such regions embedded in large overdensities are rarer still.
We can make an estimate of the scale of the peak in the distribution of void sizes, namely of the typical dimension of voids, considering a model where the initial power spectrum is $P(k) \propto k^n$. We know that for the values of interest of the void and cloud parameter, in the range $0.6 \leq D \leq 0.75$, the distribution is peaked about $\nu_{max} \approx 1$, so from the definition of ν we have $\sigma \sim \mid \delta_v \mid$ and since from the definition of the power spectrum we have $\sigma^2 \propto R^{-n-3}$, with R scale of interest, we have that the initial comoving size R_i of a region that is identified as a void is

$$\frac{\mid \delta_v \mid}{\sigma_8} = \left(\frac{8}{R_i}\right)^{(n+3)/2} \Rightarrow R_i = 8\left(\frac{\sigma_8}{\mid \delta_v \mid}\right)^{2/(n+3)}, \tag{3.7}$$

with σ_8, evaluated ≈ 0.9 in ΛCDM models, denoting the rms fluctuation on scales of $8h^{-1}Mpc$. We know that at the moment of shell crossing the initial comoving radius has expanded by a factor 1.7, so the final size r_v of the void is

$$\frac{r_v}{h^{-1}Mpc} \sim 1.7 \times R_i = 1.7 \times \frac{8}{3^{2/(n+3)}}. \tag{3.8}$$

We obtain a reasonable approximation to CDM spectra on megaparsec scale by setting $n = -1.5$, consequently the typical void radius is $\sim 3h^{-1}Mpc$. Since the correlation length is of the order of $8h^{-1}Mpc$, this makes the typical void diameter of the order of the correlation length.
Typically a cluster forms from a region that had comoving radius $R_i \sim 8(\sigma_8/\delta_c)^{2/(n+3)} \sim 3.5h^{-1}Mpc$. Since few collapsing regions are larger than this, voids that are initially larger than this are unlikely to have been squeezed out of existence.
The number density can be expressed also in terms of the linear theory radius of the void r_L. Defining the volume of a spherical region of an arbitrary radius R as $V(R) := 4\pi R^3/3$ and making the substitution $\rho_m/M = 1/V(r_L)$ we obtain

$$\frac{dn_L}{d\ln r_L} = \frac{1}{V(r_L)} f(\Lambda, \delta_v, \delta_c) \frac{d\Lambda}{d\ln r_L}. \tag{3.9}$$

In the spherical evolution model the actual void expands from its linear radius. At the epoch of shell crossing $\rho_v = 0.2\rho_m$, so, given that

$$\frac{r}{r_L} = \left(\frac{\rho_m}{\rho_v}\right)^{1/3}, \tag{3.10}$$

the spherical expansion predicts that $r \approx 1.7 r_L$. Therefore in the Sheth and van de Weygaert model the void abundance becomes

$$\frac{dn}{d\ln r} = \left.\frac{dn_L}{d\ln r_L}\right|_{r_L = r/1.7} \tag{3.11}$$

because it is assumed that the expansion of isolated voids preserves their total number density. Figure 3.5 shows that in this model $dn/d\ln r$ does not change in amplitude, shifting left to right in scale through the non linear growth. In the Sheth and van de Weygaert model the threshold δ_v is fixed by the shell crossing whereas the threshold δ_c modifies its value within $1.06 \leq \delta_c \leq 1.686$, but in figure 3.5 it is evident that changing δ_c between these values has little effect on the void abundance for large ones, about more than $1h^{-1}Mpc$. Unfortunately for spherical expansion the assumption that the comoving number density of objects is conserved during the evolution $n = n_L$ whereas their sizes change is invalid for large voids. The cumulative volume fraction in voids larger than R defined as

$$F(R) = \int_R^\infty \frac{dr}{r} V(r) \frac{dn}{d\ln r} \tag{3.12}$$

exceeds unity at $R \approx 2h^{-1}Mpc$ and in figure 3.6 we see that changing δ_c within the given range affects only small voids without resolving the problem. In fact, taking $R \to 0$, for the exact $\Lambda f(\Lambda, \delta_v, \delta_c)$ we have

$$F(0) = \left(\frac{r}{r_L}\right)^3 \int_0^\infty \frac{d\sigma}{\sigma} 2\Lambda f(\Lambda, \delta_v, \delta_c) = \left(\frac{r}{r_L}\right)^3 (1 - D). \tag{3.13}$$

Knowing that the expansion factor is $r \approx 1.7 r_L$, given by the spherical evolution, for $\delta_v = -2.7$ and $\delta_c = 1.686$ the mass fraction in voids is $(1 - D) = 0.383$ and the volume

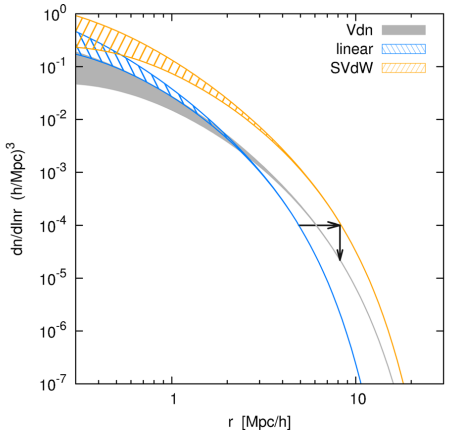

Figure 3.5: Void abundance model predictions. In the Sheth and van de Weygaert model (SVdW) the number density of linear underdensities (blue curve) is unchanged in void formation and the right arrow to the orange curve means that only their sizes change. In the volume conserving model (Vdn) the number density changes in order to conserve the volume fraction in voids, lowering the amplitude at fixed shape, as demonstrated by the low arrow to the grey curve. Varying $1.06 \leq \delta_c \leq 1.686$ we form the shaded and hatched regions and we show that the abundance of voids changes significantly only for small ones, with radius $r \leq 1h^{-1}Mpc$. We take $\delta_v = -2.7$ throughout.[7]

fraction in voids is more than unity, also if we set $\delta_c = 1.06$, indicating that we have a model in which about one third of the mass of the universe is associated with voids that occupy most of the volume. The remaining two thirds of the mass is in between the voids and occupies negligible volume.
We can obtain a more physical value of order unity of the cumulative volume fraction if we reduce $\delta_v \to 0$ simultaneously taking $r \to r_L$ and $D \to 0$. δ_v is given by the shell crossing criterion, but it is interesting to analyze if modifications of this vision can bring the Sheth and van de Weygaert model to a better agreement with simulations. If we change the non linear density at which voids are defined ρ_v/ρ_m, the linear density threshold δ_v and the expansion factor r/r_L change self consistently. Indeed in figure 3.7 we show that the abundance of voids is modified by changing δ_v, in particular decreasing $\mid \delta_v \mid$ decreases the steepness of the abundance function. Therefore lowering δ_v makes the total volume fraction physical and increases the abundance of the largest voids. In the excursion set method the agreement between simulations and the abundance of voids is better if we do not use the idea of isolated spherical expansion of voids, because it is this last assumption that, through the conservation of the number density of voids, causes the problem of miscounting the abundance of voids in the Sheth and van de Weygaert model. One solution to the problems coming from the vision presented in the work by Sheth and van de Weygaert is to change some considerations about the evolution from the linear regime of underdensities to the non linear regime of voids. In [7] voids are defined as spherical regions of ρ_v/ρ_m and using a linear underdensity of $\delta_v = -2.7$ through a

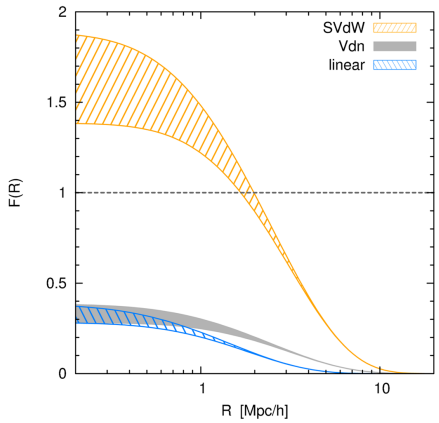

Figure 3.6: The cumulative volume fraction in voids with radii larger than R for the different models: the blue striped region, $R = r_L$, representing linear theory and the orange and the grey shaded regions, $R = r$, representing respectively the SVdW and the Vdn models. Regions correspond to expected range of $1.06 \leq \delta_c \leq 1.686$ and we take $\delta_v = -2.7$ throughout. For the SVdW model the fraction unphysically exceeds unity at $R \approx 2h^{-1}Mpc$ while the Vdn model conserves the total fraction of $F(0) \approx 0.3$ from the linear theory.[7]

spherical evolution model the mapping from linear to non linear regime is no more guided by the conservation of number density, like in the model of Sheth and van de Weygaert, but by the fixing of the volume fraction of the universe which is in voids assuming that the ones of a similar size merge as they expand by a factor of 1.7 to achieve a non linear density of $\rho_v = 0.2\rho_m$ today. While the model of Sheth and van de Weygaert greatly overpredicts the abundance of voids this new model matches the measured abundance to within 16 per cent for a range of void radii $1 < r(h^{-1}Mpc) < 15$. We will analyze more deeply this volume conserving model in the following section, considering that it works well at matching the number density of voids measured from the simulations at higher redshift as well as it correctly predicts the abundance of voids to within 25 per cent in a simulation of a matter dominated $\Omega_m = 1$ universe and that it is also able to examine the abundance of voids in the halo distribution finding fewer small voids, $r < 10h^{-1}Mpc$, and many more large voids, $r > 10h^{-1}Mpc$, compared to the underlying void distribution in the dark matter.
There is an other approach, developed in the last section of this chapter, that helps to improve one unphysical feature of the model by Sheth and van de Weygaert. Namely, while for Sheth and van de Weygaert the volume fraction covered by voids is $5\delta_c/(\delta_c + \mid \delta_v \mid) \approx 2$, taking $\delta_c = 1.686$ and $\mid \delta_v \mid = 2.7$, in [4] this fraction is smaller, because a smaller Eulerian volume is assigned to each of the Sheth and van de Weygaert void candidates: the associated void covering factor is 1.17, still greater than unity, but the problem now is only 20 per cent. Through this vision the calculation of the abundance of large voids and the analysis of their evolution is made improving how the void in cloud

process is incorporated into the excursion set approach, considering voids within a fully Eulerian, rather than purely Lagrangian, framework and accounting for correlations between different spatial scales in the initial conditions. We conclude this section discussing some details about voids evolution and their ancestral heritage.
The typical comoving void radius is larger at late times, because the threshold $\delta_v(a)$ decreases with time. Therefore the mass contained within a typical void is larger at late times. However we know that in an Einstein-de Sitter universe the mass fraction in voids, approximately one third of the total mass in the universe, does not evolve, so we deduce that the more massive voids present at later times are the result of the merging of the smaller voids present at early times. In fact, knowing that the thresholds δ_c and $\mid \delta_v \mid$ are higher at an earlier epoch, we can infer the fraction of mass M that was previously in voids that contained mass m making the following substitutions in equation 3.3: $\delta_c \to \delta_{c1} - \delta_{v0}$, $\delta_v \to \delta_{v1} - \delta_{v0}$, $\Lambda \to \Lambda(m) - \Lambda(M)$. We denote with the suffix 1 the quantities at the earlier time and with the suffix 0 the quantities at the present time. Integrating over all $0 < m < M$ we find the expression for the mass fraction of M that was in voids at the earlier epoch:

$$f_{void}(M) = 1 + \left(\frac{\delta_{v1} - \delta_{v0}}{\delta_{c1} - \delta_{v1}}\right) = \frac{\delta_{c1} - \delta_{v0}}{\delta_{c1} - \delta_{v1}}. \tag{3.14}$$

The fact that this fraction is less than unity reflects that at earlier times a part of the mass now contained in the void was not in the smaller ancestral voids, but, instead, in the walls and filaments which divided at earlier times the present larger void in its smaller constituent voids. Defining z as the redshift, in an Einstein-de Sitter universe we have

$$\delta_{c1} - \delta_{c0} = \delta_{c0}(z_1 - z_0), \qquad \delta_{v1} - \delta_{v0} = \delta_{v0}(z_1 - z_0), \tag{3.15}$$

therefore the mass fraction of void matter that was in voids at the earlier time can be expressed also as

$$f_{void}(M) = 1 - D_0 \frac{z_1 - z_0}{(1 + z_1)}, \tag{3.16}$$

with $D_0 \equiv \mid \delta_{v0} \mid /(\delta_{c0} - \delta_{v0})$ void and cloud parameter at the current epoch. At $z_1 \approx z_0$ the fraction is close to unity because the time of creation of the smaller voids constituent the larger one and the time of creation of this last one are almost the same. Whereas for large lookback times $z_1 \gg z_0$ the fraction tends to $1 - D_0$, equal to the global void mass fraction in the equation 3.5.

3.3 Volume conserving model

The Sheth and van de Weygaert model, based on the excursion set theory, makes a good prediction of the shape of the abundance of voids measured from simulations, but the predicted amplitude is incorrect because it is assumed an isolated spherical expansion which does not consider the merging of voids as they expand. Therefore in this section we show that fixing the volume fraction and shape of the abundance function of voids, rather

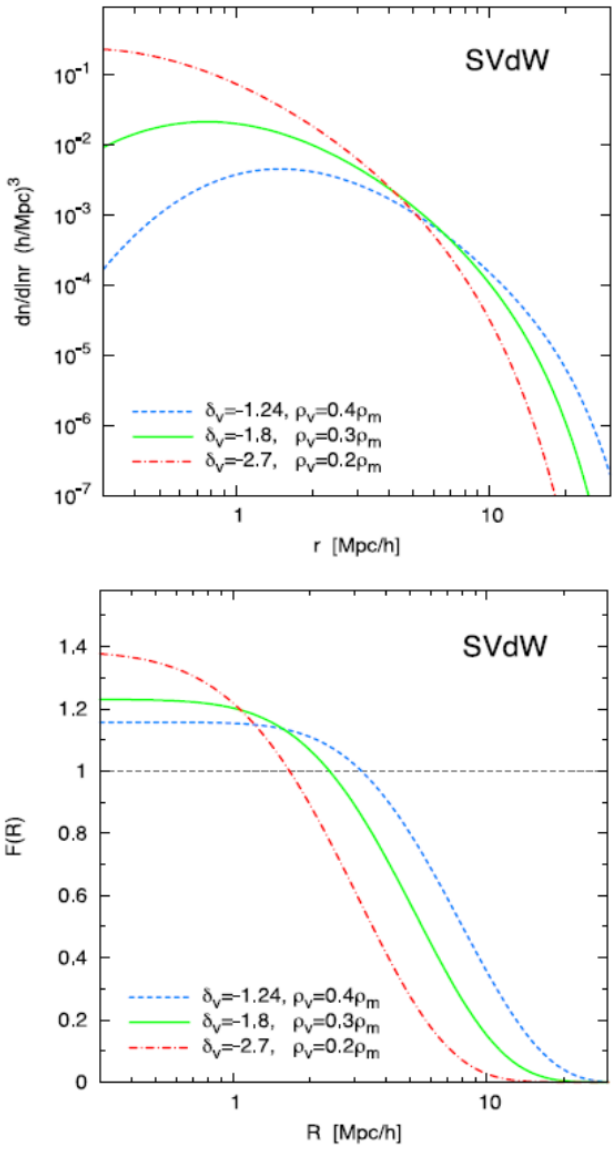

Figure 3.7: Upper: variation of the void underdensity ρ_v/ρ_m changes both the shape of the abundance through the linear barrier δ_v and the size of the voids or horizontal shift through $r/r_L = (\rho_v/\rho_m)^{-1/3}$. Decreasing $\mid \delta_v \mid$ increases the number of large voids and decreases that of small ones. Lower: the cumulative volume fraction in voids with radii larger than R decreases as $\mid \delta_v \mid \to 0$ and $R \to 0$ but making the larger voids more abundant. We use $\delta_c = 1.06$ throughout.[7]

than their total number density, during the expansion helps to avoid the unphysicality of the isolated spherical expansion model for voids and matches the measured abundances to within 16 per cent for void radii $1 < r(h^{-1}Mpc) < 15$. This model, which is also based on the excursion set approach with $\delta_v = -2.7$, works well and we have in the passage from the linear to the non linear regime that the number density of voids decreases by the same amount that the voids expand. In fact, if we define the volume fraction in linear theory as

$$F_L(R_L) = \int_{R_L}^{\infty} \frac{dr_L}{r_L} V(r_L) \frac{dn_L}{d\ln r_L}, \tag{3.17}$$

this fraction is conserved defining the non linear abundance as

$$V(r)dn = V(r_L)dn_L\Big|_{r_L(r)}. \tag{3.18}$$

A void combines with its neighbours conserving volume and not number in the expansion from the linear to the non linear regime, so the abundance becomes

$$\frac{dn}{d\ln r} = \frac{V(r_L)}{V(r)} \frac{dn_L}{d\ln r_L} \frac{d\ln r_L}{d\ln r}\Big|_{r_L(r)} = \frac{1}{V(r)} f(\Lambda, \delta_v, \delta_c) \frac{d\Lambda}{d\ln r_L} \frac{d\ln r_L}{d\ln r}\Big|_{r_L(r)}. \tag{3.19}$$

For this model, called Vdn model, we assume $r = 1.7r_L$ for voids with non linear density $\rho_v = 0.2\rho_m$ from N-body simulations as a starting point. We have $d\ln r_L = d\ln r = 1$ and we see in figure 3.5 that the impact of going from the linear to the non linear abundance is given by a shift in scale and a change in amplitude with no change in shape represented by the combination of the two arrows. Moreover in figure 3.6, since by construction the volume fraction is conserved, the cumulative volume fraction for this abundance function and that for linear theory defined in equation 3.17 differ only by a horizontal shift in scale.

The Vdn model is not the unique way of keeping the volume fraction below unity. For example, in the upper panel of figure 3.8, being in a $\sigma_8 = 0.8$ ΛCDM cosmology, we can change the parameters $\delta_v \to -2$ and $r/r_L = 1$ in the Sheth and van de Weygaert model to fit the Vdn model. Instead, for an other cosmology as the Einstein-de Sitter universe, in the lower panel of figure 3.8 the same change predicts different abundances than the Vdn model. In general simulations favour the Vdn model over universal changes in δ_v and r/r_L, in fact we will show now that the measured abundances of voids from ΛCDM simulations for different simulation box sizes are consistent with the predictions made by the Vdn excursion set model, also in case of variations in the critical void underdensity, redshift and cosmology. We test the robustness of both the volume conserving and the Sheth and van de Weygaert model comparing their results with the abundance of voids in the dark matter distribution calculated with a series of N-body simulations in box sizes $64h^{-1}Mpc$, $128h^{-1}Mpc$, $256h^{-1}Mpc$, $500h^{-1}Mpc$ on a side. The ΛCDM model used has the following cosmological parameters: $\Omega_m = 0.26, \Omega_{DE} = 0.74, \Omega_b = 0.044, h = 0.715$ and a spectral tilt of $n_s = 0.96$. The linear theory rms fluctuation in spheres of radius $8h^{-1}Mpc$ is set to be $\sigma_8 = 0.8$ for our main simulation set of eight independent realizations of the ΛCDM cosmology. We investigate the abundance of voids also in a ΛCDM

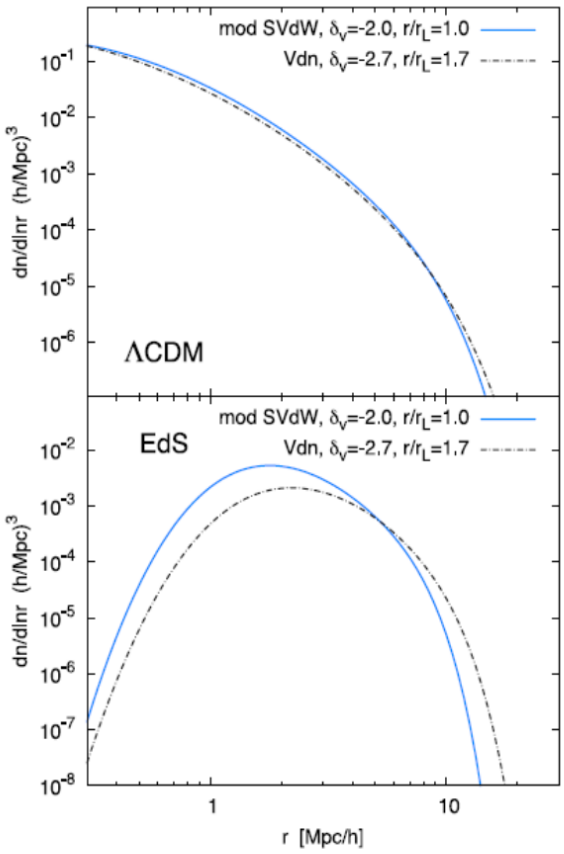

Figure 3.8: Void abundance in the Vdn model (black dot-dashed curve) and a modified SVdW model (blue solid line) with ad hoc variations designed to fit the ΛCDM Vdn model. In the upper panel we choose $\delta_v = -2$ and $r/r_L = 1$ in the SVdW model in violation of spherical expansion predictions in a $\sigma_8 = 0.8$ ΛCDM cosmology. In the lower panel we illustrate that the same set of parameters give not so much similarity between the SVdW model and the Vdn one in an Einstein-de Sitter universe. For all curves $\delta_c = 1.686$.[7]

cosmology with $\sigma_8 = 0.9$ and in an Einstein-de Sitter simulation with $\Omega_m = 1$. Even if the Einstein-de Sitter universe has already been ruled out by many observations we use it as a tool to examine how robust our void models are to power spectrum or cosmology changes. To estimate densities and find both voids and subvoids and to measure the abundance of spherical voids which have $\rho_v = 0.2\rho_m$ at $z = 0$ in our N-body simulations it is used the Voronoi tessellation method. It gives a local density estimate by dividing space into cells, where the cell around any given particle is the region of space closer to that particle than to any other. Therefore the goal of our void finder is to identify all spherical non overlapping underdense regions of average density $\rho = 0.2\rho_m$ in a dark matter simulation.
In figure 3.9 it is made a model comparison between the average number density of voids as a function of radius measured from eight different realizations of the ΛCDM cosmology in simulation box sizes $64h^{-1}Mpc$ (green), $128h^{-1}Mpc$ (purple), $256h^{-1}Mpc$ (red), $500h^{-1}Mpc$ (cyan) on a side and that one measured by the Vdn model and the Sheth and van de Weygaert model. Within the parameter interval $\delta_c = [1.06, 1.686]$ and assuming $\delta_v = -2.7$ and that the voids have expanded by a factor of 1.7 today in figure 3.9 the Sheth and van de Weygaert model and the Vdn one are represented respectively by the orange hatched region and by the grey shaded one. As we can see only the smallest voids of $r < 1h^{-1}Mpc$ are influenced by the range of different δ_c values accounting for the void in cloud process. The decrease in the void abundance at $r(h^{-1}Mpc) \sim 2.5, 1.5, 1$ for the $256, 128, 64h^{-1}Mpc$ boxes shows the resolution limit for each of these simulations where small voids are not fully resolved.
Assuming an isolated spherical expansion and the conservation of the linear theory number density are the conditions that constrain the Sheth and van de Weygaert model to overpredict the abundance of voids by a factor of 5, while, accounting for the fact that voids merge as they expand, the Vdn model agrees with the results measured from simulations in different box sizes to ~ 16 per cent across the range $1 < r(h^{-1}Mpc) < 15$. Therefore the Vdn model conserves the volume rather the number of voids since in going from the linear to the non linear regime the number density decreases by the same amount that the volume of the voids grow.
Now we test the goodness of the predictions of the Vdn model also with other choices of parameters in the simulations. Generally we adopt the shell crossing criterion $\rho_v = 0.2\rho_m$ for defining the void in both the Sheth and van de Weygaert and Vdn models. If we modify our void finder such that the largest non overlapping spherical regions recovered from the simulations have densities $\rho_v = 0.3\rho_m$ and $\rho_v = 0.4\rho_m$ in the left and right panels of figure 3.10 it is shown that the Vdn model continues preserving agreement with the simulations also with these density variations.
Changing the underdensity criteria in the spherical evolution model we consequently modify the linear threshold δ_v that alters the shape of the abundance function: in our case we have $\delta_v = -1.8$ for $\rho_v = 0.3\rho_m$ and $\delta_v = -1.24$ for $\rho_v = 0.4\rho_m$. Given $\rho_v = 0.2\rho_m$ for dark matter voids, the Vdn model has the same shape as the Sheth and van de Weygaert model but a factor of $\rho_v/\rho_m = 1.5$ lower amplitude. The number density of underdense regions with $\rho_v = 0.3\rho_m$ and $\rho_v = 0.4\rho_m$, which cannot be directly related

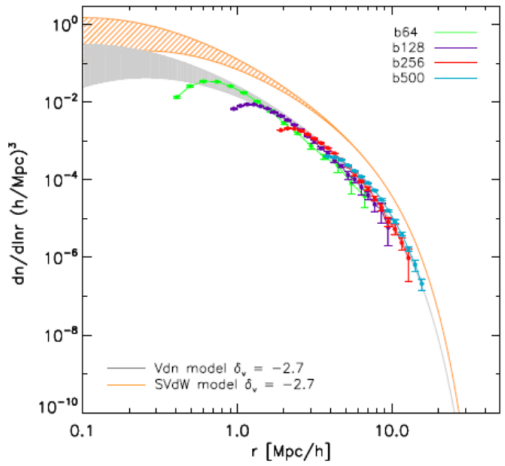

Figure 3.9: Void abundance in simulations versus predictions with $\rho_v = 0.2\rho_m$ in the dark matter distribution of the $\sigma_8 = 0.8$ ΛCDM cosmology in simulation box sizes $64h^{-1}Mpc$ (green), $128h^{-1}Mpc$ (purple), $256h^{-1}Mpc$ (red), $500h^{-1}Mpc$ (cyan) on a side. The error bars represent the scatter on the mean from eight different realizations of this cosmology in each box size. The range in predictions cover the parameter interval $\delta_c = [1.06, 1.686]$ with $\delta_v = -2.7$ and are consistent with simulations for Vdn (grey shaded) but not SVdW models (orange hatched).[7]

to shell crossing in the spherical expansion model, can be modelled by following the Vdn model obtained rescaling the Sheth and van de Weygaert model by ρ_v/ρ_m, equal to 0.3 and 0.4 in the two different cases, or by simply keeping this factor fixed at 1/5. The grey shaded regions in the two panels of figure 3.10 limited by the black dashed and dotted lines indicate the difference between the values of the rescaling factor ρ_v/ρ_m. However the simple phenomenological prescription of rescaling the amplitude by 1/5 is more similar to the simulation results than the ρ_v/ρ_m scaling prescription of the Vdn model that overpredicts the amplitude by ~ 1.5 for voids with $\rho_v = 0.3\rho_m$ and 2 for $\rho_v = 0.4\rho_m$. Therefore the excursion set models predict the overall shape of abundance function accurately and only the amplitude needs to be modified to fit the simulation results, here without the benefit of volume conservation as motivation. Moreover the preferred scaling of 1/5 is more than sufficient to bring the predictions to a physical void filling fraction for $\rho_v \geq \rho_m$. The robustness of the Vdn model in measuring the void abundance can be analyzed also with the variation of the redshift. In figure 3.11 the number density of voids is plotted as a function of radius at redshift $z = 0.5$ (blue) and $z = 1$ (red) measured from the ΛCDM, $\sigma_8 = 0.8$ simulation. The measured abundances from the three simulation box sizes $64, 128, 256h^{-1}Mpc$ are the volume weighted averages and errors over eight realizations. The volume conserving model using the parameter range $\delta_c = [1.06, 1.686]$ with $\delta_v = -2.7$ is represented as a black hatched region at $z = 0.5$ and as a grey shaded region at $z = 1$. In order to do not make confusion in this figure it is

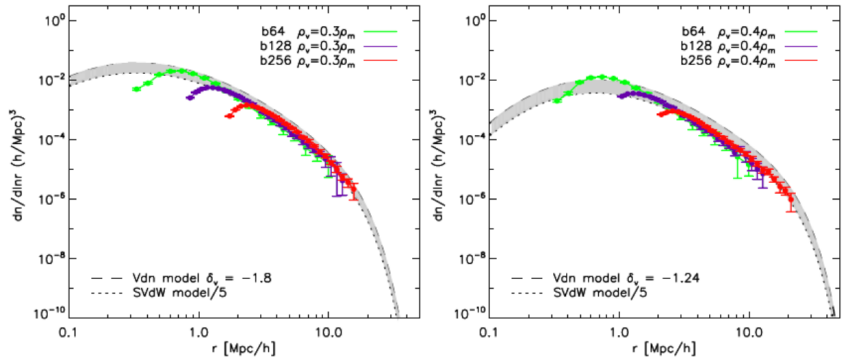

Figure 3.10: Void abundance for different defining underdensities $\rho_v = 0.3\rho_m$ (left handed panel), $\rho_v = 0.4\rho_m$ (right handed panel). The grey shaded region represents the excursion set predictions with varying amplitude. The amplitude rescaling versus the SVdW predictions ranges from ρ_v/ρ_m (Vdn; top black dashed curve) to 1/5 (bottom black dotted curve) both of which preserve agreement for $\rho_v = 0.2\rho_m$.[7]

used the same colour for the results from the three simulation boxes at each redshift. At both redshifts the Vdn model reproduces very well the abundance of voids in the dark matter in a ΛCDM universe. Instead, if we had plotted the Sheth and van de Weygaert model again it would have overpredicted the abundance of voids by approximately a factor of 5. Moreover in figure 3.11 we note that at $z = 1$ the smaller voids are more abundant than at $z = 0.5$, instead at $z = 0.5$ the larger voids are more abundant than at $z = 1$. From figure 3.12 it is evident that the volume conserving model, represented in both panels as a grey shaded region, works well not only in the standard ΛCDM cosmology with $\sigma_8 = 0.8$, but also in alternative universes as the ΛCDM cosmology with the modified value of $\sigma_8 = 0.9$, plotted in the upper panel, or the Einstein-de Sitter universe where the matter density parameter $\Omega_m = 1$, plotted in the lower panel. Even if the value δ_v for the linear perturbation theory underdensity is different in different cosmologies we have used the same quantity for this parameter, $\delta_v = -2.7$, because a small change in δ_v going from a ΛCDM universe to an Einstein-de Sitter one has a small impact on the predicted abundance of voids in the excursion set theory and the main differences arise from the change in the variance, $\sigma(R)$.
The volume conserving model fits well the number density of voids measured at $z = 0$ with $\rho_v = 0.2\rho_m$ by the simulations in the two different cosmologies within 25 per cent over the range $1 < r(h^{-1}Mpc) < 15$. For voids with small radii $r < 2h^{-1}Mpc$ we note, especially in the Einstein-de Sitter universe, an overall decrease in the abundance of voids in the dark matter distribution caused by the squashing of smaller voids due to the void in cloud process. We have already seen in figure 3.8 that a modification of the Sheth and van de Weygaert model in order to match the simulation results of ΛCDM with $\sigma_8 = 0.8$ would be useless for the Einstein-de Sitter cosmology. Finally we study voids in the galaxy population not through the dark matter density field but by the number density

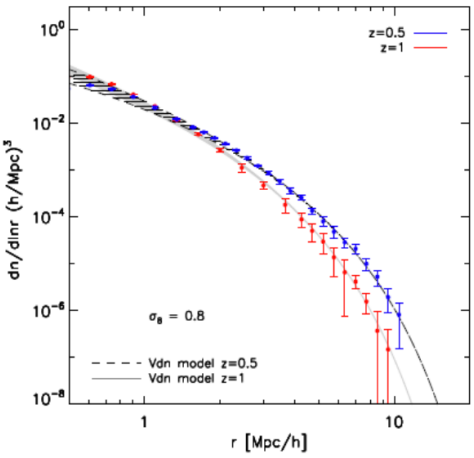

Figure 3.11: Redshift dependence of the voids abundance with $\rho_v = 0.2\rho_m$ at $z = 0.5$ (blue) and $z = 1$ (red) measured from the ΛCDM, $\sigma_8 = 0.8$ simulation. The black hatched (grey shaded) region represents the Vdn model using linear underdensity values of $\delta_v = -2.7$ at $z = 0.5(z = 1)$ for the range $\delta_c = [1.06, 1.686]$. Note that the measured average abundances and errors from the $256, 128, 64h^{-1}Mpc$ simulation boxes are the volume weighted values. Note also that in this figure we have plotted the results from the three simulation boxes using the same colour for clarity.[7]

field n_h of the dark matter haloes they populate. Our aim is to test how faithfully the abundance of voids in the dark matter matches that in the halo populations knowing that a comparison between voids in the dark matter and halo distributions accounts for the galaxy/dark matter halo biasing relation that makes galaxy voids larger than dark matter voids and voids selected using luminous galaxies larger than those using faint galaxies.

In figure 3.13 we use the void finder explained before to identify voids in the distribution of haloes which have $n_v = 0.2n_h$, where n_v is the average number density in the void whereas n_h is the average in the whole simulation, in order to show the abundance of voids at $z = 0$ in the haloes population using the FOF halo catalogues from the $128, 256, 500h^{-1}Mpc$ simulation boxes and the halo catalogues from the higher resolution Multidark and Bolshoi simulations, which have computational box sizes of $L = 1000$ and $250h^{-1}Mpc$ on a side, respectively. Our sample consists of 5768 voids using 1.7×10^6 haloes from the Multidark simulation and 4826 voids using 2.2×10^6 haloes from the Bolshoi simulation.

The number density of voids in the halo distribution cannot be modelled by the volume conserving approach assuming $n_v = 0.2n_h$ which corresponds to dark matter voids of $\rho_v = 0.2\rho_m$, represented by a solid grey line in figure 3.13. Moreover voids identified in this manner do not respect the predictions of the Sheth and van de Weygaert model which would have the same shape of the Vdn model but five times the amplitude of this last one. Instead, if we plot the Vdn model assuming that halo defined voids of $n_v = 0.2n_h$

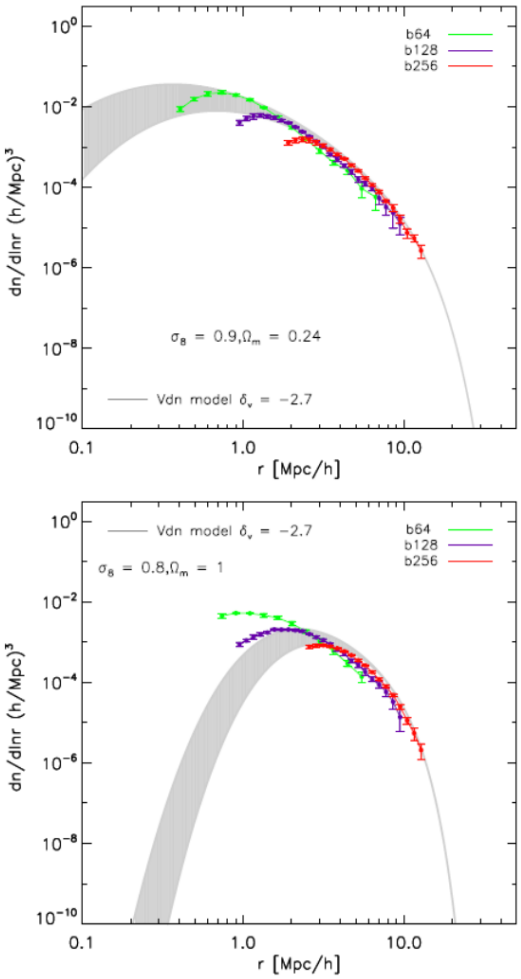

Figure 3.12: Void abundance for alternate cosmological parameters at $z = 0$. Upper: ΛCDM with the initial condition normalized to give $\sigma_8 = 0.9$. Lower: EdS model with $\sigma_8 = 0.8$ with $\Omega_m = 1$. The grey shaded region shows the Vdn model within the parameter interval $\delta_c = [1.06, 1.686]$ and using $\delta_v = -2.7$.[7]

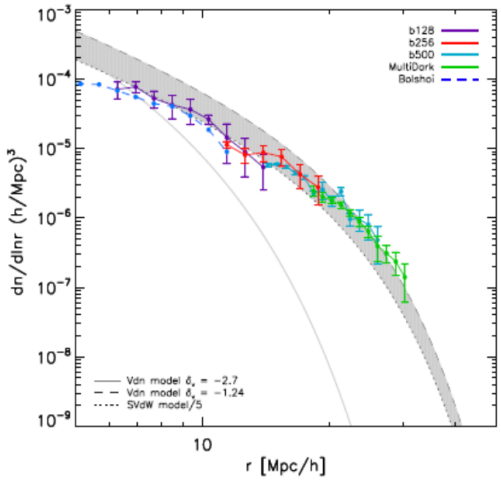

Figure 3.13: The number density of voids with $n_v = 0.2n_h$ in the halo distribution from the 128 (purple), 256 (red), 500 (cyan) h^1Mpc simulation boxes. The results from the Multidark and Bolshoi simulations are shown respectively in green and blue. The error bars represent the error on the mean from eight simulations. The errors on the Multidark simulation represent the Jackknife error on the mean. The grey shaded region bounded by the black dashed and dotted line represents the volume conserving model with $\delta_v = -1.24$ and varying amplitude as in figure 3.10. The grey solid line represents the Vdn model with $\delta_v = -2.7$.[7]

correspond to dark matter defined voids of $\rho_v = 0.4\rho_m$, the predictions of this volume conserving model plotted as a grey shaded region allowing the amplitude to vary from the predictions of the Vdn model which rescales the Sheth and van de Weygaert amplitude by $\rho_v/\rho_m = 0.4$ and the rescaling of 1/5, represented in figure 3.13 respectively by a black dashed line and a black dotted line, match well the abundance of voids in the halo populations though no single rescaling matches perfectly across the full range. These conclusions show that between voids in the dark matter and the dark matter halo distributions there is not always a perfect one to one correspondence and this is more evident if we define a void fixing the underdensity which is the same for dark matter and haloes. Analyzing figure 3.14 we understand the difference between the voids which we find in the dark matter and halo distributions using the same underdensity criterion, in fact more than one dark matter void overlaps with the halo void and moreover the radii of the dark matter voids at which $\rho_v = 0.2\rho_m$ are a lot smaller than the ones of the halo voids which satisfy the same criterion.
The Vdn predictions for the abundance of dark matter voids can be reconciled with that of the halo voids through a scale dependent modification to the barrier in the Vdn model useful to change the underdensity threshold used to find voids in the dark matter or fixing the underdensity threshold to define dark matter voids we can find a scaling of this threshold in order to define voids in the halo distribution.

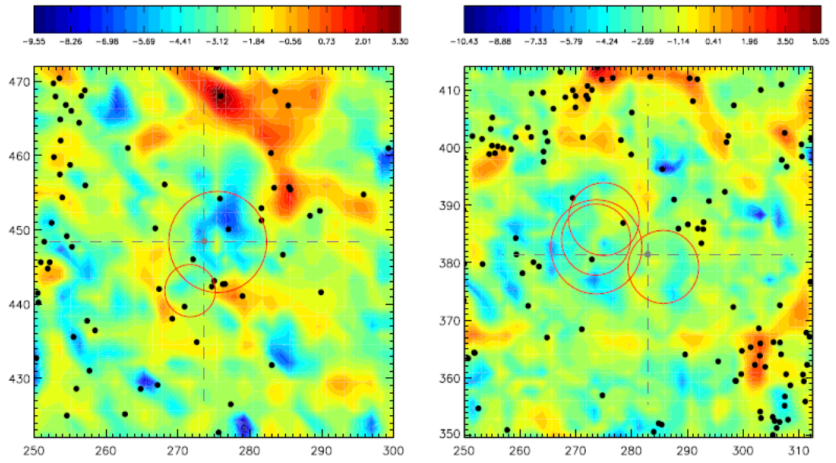

Figure 3.14: Left: a $10 \times 50 \times 50h^1 Mpc$ slice through the $500h^{-1}Mpc$ simulation box centred on a large, $r \sim 21h^{-1}Mpc$, void in the halo population (black dots). The diameter of the void is shown as a dashed grey line and the coloured contours represent the log of the density field which has been evaluated on a grid of 256^3 points. The red circles represent all the voids in the dark matter which have $\rho_v = 0.2\rho_m$ and whose centres are within $10h^{-1}Mpc$ of the centre of the void in the halo distribution. Right: a $60 \times 14 \times 60h^1 Mpc$ slice through the $500h^{-1}Mpc$ simulation box centred on a large, $r \sim 26h^{-1}Mpc$, void in the halo population (black dots). The red circles represent all the voids in the dark matter which have $\rho_v = 0.2\rho_m$ and whose centres are within $10h^{-1}Mpc$ of the centre of the void in the halo distribution. Note that these voids only appear to be overlapping due to the projection effect.[7]

3.4 New treatment of the void in cloud process

The Sheth and van de Weygaert model eliminates from the list of all possible voids all those that are surrounded by an initially larger region which is destined to have collapsed and formed a cluster before the void they surround would have formed. The problem is how to treat those underdense regions which will be squeezed, altering their sizes, by their surrounding overdense regions which will however possibly not have collapsed completely by the time the void inside them forms. Sheth and van de Weygaert tried to predict this effect changing the overdensity associated with cloud from collapse to turnaround, but while this change has a dramatic effect on smaller voids, the difference for big voids is small. This is a drawback because the largest voids are the most easily measured, and so most likely to place the most interesting constraints on cosmological models. Therefore in this section we present a formulation which resolves this problem of the Sheth and van de Weygaert model. Within the excursion set approach we show a more careful statement of the void in cloud process establishing the steps for being a void in terms of the late time field, namely a void is the largest region in the late time field which is sufficiently underdense, and to then determine what this needs from the initial field. This means that it is possible to relate Eulerian volumes in the late time field and Lagrangian ones in the initial field.
Before we compare the Lagrangian and Eulerian treatments and then analyze how the last one can be useful to develop a better model of the void in cloud process we make some considerations about the notation used in what follows: we call the Eulerian radius and volume of the void R and V, so that $V = 4\pi R^3/3$, and refer to Lagrangian length scales through the associated mass $m = \bar{\rho}(4\pi R_L^3/3)$, where $\bar{\rho}$ is the comoving background density and R_L is the Lagrangian radius which evolved into the Eulerian radius R; moreover we denote the variance of the linearly extrapolated density contrast when filtered on a Lagrangian scale corresponding to mass m as $\Lambda(m) = (2\pi^2)^{-1}\int_0^\infty dk k^2 P(k) W^2(kR_L)$, where $W(kR_L)$ is the filter and $P(k)$ the linearly evolved matter power spectrum.
We have two conditions that have to be respected in order to identify a region of size R as a void of Eulerian size R at some time t: this region of size R must be (a) less dense than some critical threshold (typically about 20 per cent of the background density) and (b) denser than this critical threshold value on all larger Eulerian scales. These same requests in the Lagrangian treatment used by Sheth and van de Weygaert are changed in: the Lagrangian region of mass scale M must be (a_L) less dense than some critical density initially (typically, linear theory overdensity of -2.7), (b_L) denser than this on all larger mass scales and (c_L) not dense enough on these larger Lagrangian scales for this to have influenced the evolution of the initial void candidate region sufficiently that it did not form a void at late times. The constraints $(a_L), (b_L)$ in the plane of initial overdensity versus scale correspond to the first crossing of a barrier of constant height δ_v, which assumes the value -2.7, approximately independent of the background density, at the present time using linear theory. This choice about the critical threshold value fixes the void mass as $M \approx 0.2\bar{\rho}V$. The third condition (c_L) is about how to eliminate from the list of all potential voids identified in the initial conditions those which cannot

be identified as voids also at later times, namely in the Eulerian field. In the excursion set approach by Sheth and van de Weygaert this constraint is followed by introducing a second barrier $B = \delta_c = constant$, where δ_c is the initial overdensity required for collapse at some time t. This threshold is useful because from the set of walks which first cross δ_v at the Lagrangian scale corresponding to the void mass M we must not consider those which crossed the barrier B, at some mass $m > M$, before they crossed δ_v. Using this method we exclude from the list of voids only those ones squeezed out completely by the regions which surround them and that have collapsed by time t.
In our new Eulerian treatment working on the conditions (a) and (b) stated in Eulerian space we analyze more carefully the voids which have only partially squeezed by the collapse of their surroundings, that where counted as voids at later times by Sheth and van de Weygaert.
The Eulerian overdensity $\Delta_{NL} \equiv m/(\bar{\rho}V)$, with m mass in a region of volume V at time t, and the linearly extrapolated density contrast $\delta(t)$ are related by the spherical evolution model in the following equation:

$$\Delta_{NL}(t) = \frac{m}{\bar{\rho}V} \approx \left(1 - \frac{\delta(t)}{\delta_c}\right)^{-\delta_c}, \tag{3.20}$$

with $V = 4\pi R^3/3$. If we rewrite this relation in an other way we can define a curve $B_V(m)$ which gives the value of the linearly extrapolated density contrast in a Lagrangian region containing mass m which develops into the Eulerian volume V at time t:

$$B_V(m) = \delta_c\left[1 - \left(\frac{m}{\bar{\rho}V}\right)^{-1/\delta_c}\right]. \tag{3.21}$$

$B_V(m)$ decreases monotonically as m decreases. At $m \gg \bar{\rho}V$ $B_V(m) \to \delta_c$, at $m = \bar{\rho}V$ $B_V(m)$ crosses 0 and at m sufficiently smaller than $\bar{\rho}V$ can cross δ_v. Moreover, setting $\delta(t) = -2.71$ and $\delta_c = 1.686$ makes $\Delta_{NL} \approx 0.2$ implying a void mass of $M \approx 0.2\bar{\rho}V$ for a void of Eulerian volume V.
In figure 3.15 the dotted lines represent a sequence of nested curves defined decreasing V, thought as a parameter. As $V \to 0$ we have that $B_{V\to 0}(m) \to \delta_c$ that is the constant barrier limit of nested curves. In figure 3.15 are also shown two random walks, plotted as blue and red solid lines, useful to understand, through the analysis of their paths, the differences between the Lagrangian and Eulerian treatments. For Sheth and van de Weygaert both walks have the same Lagrangian mass and Eulerian volume since both the two walks first cross δ_v at the same Lagrangian mass scale $\Lambda(M)$, so that $B_V(M) = \delta_v$, without overcoming δ_c before to first crossing δ_v. In our new treatment the mass inside Eulerian V at time t is given by the value of $\Lambda(m)$ at which the associated barrier $B_V(m)$ is first crossed. Therefore the two walks are different void candidates. We assign the void with the blue walk which decreases monotonically the same mass and volume assigned by Sheth and van de Weygaert, because the monotonicity in Lagrangian δ translates directly into a monotonicity in Δ_{NL}, so to satisfy either the conditions $(a), (b)$ or the conditions $(a_L), (b_L), (c_L)$. In this specific case the Eulerian volume would lie between

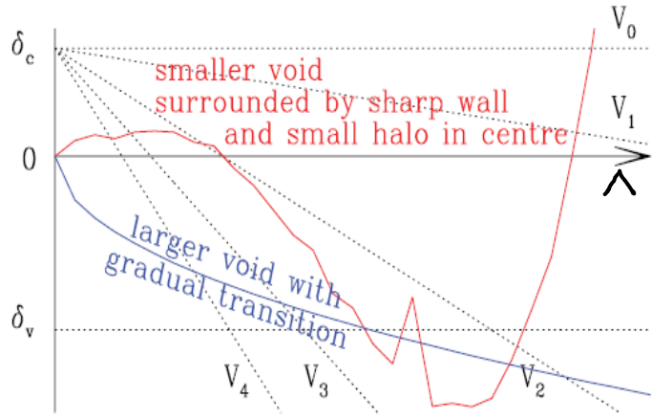

Figure 3.15: Excursion set model of voids plotted as overdensity versus scale. Dotted lines show the barriers associated with Eulerian volumes $V_4 > V_3 > V_2 > V_1 > V_0$: barriers for larger volumes fall more steeply. The horizontal line at $\delta_v = -2.71$ shows the critical linearly extrapolated initial density for voids. Solid lines show two examples of random walks; both first cross δ_v on the same mass scale Λ. Since neither walk crossed δ_c prior to crossing δ_v, Sheth and van de Weygaert would have assigned the same void mass and Eulerian volume to both walks. In our prescription, instead, the blue one which falls monotonically with Λ is associated with a larger Eulerian volume (between V_3 and V_2), because its evolution is not modified by the void in cloud process: we would have assigned the same mass and volume to it as they did; the red one represents an overdensity on the Eulerian scale V_2, because it is first crossed at $\delta > 0$, but a void on the Eulerian scale just smaller than this, because the first crossing of the next shallower barrier will be at $\delta < \delta_v$. The evolution of this void has been modified by the collapse of the overdensity surrounding it. We would assign a larger mass to the wall which surrounds the void, and a smaller mass and volume to the void itself, compared to Sheth and van de Weygaert. Finally, note that for this walk the first crossing of δ_v is not so important.[4]

V_3 and V_2.
On the contrary for the void represented by the red walk the two different treatments do not correspond. The non monotonicity of δ means that Δ_{NL} is not monotonic either. Although the predicted mass decreases monotonically with Eulerian volume V, it need not do so smoothly. When B_V is tangent to the walk on scales V the predicted mass must jump downwards as $V \to V - \Delta V$, because the value of Λ on which $B_{V-\Delta V}$ is first crossed can be substantially larger than that on which B_V was first crossed. In the figure this happens at about V_2. This means that for the entire portion of the walk from the value of Λ at which the walk itself is tangent to the barrier B_V to the next larger value of Λ at which the walk overcomes the same barrier translating the Lagrangian δ to an Eulerian Δ_{NL} using equation 3.20 will not yield the correct answer. So we must use Eulerian qunatities and not only Lagrangian ones. We can interpret physically as a dense wall surrounding the underdense void this sharp transition in mass and Eulerian density at nearly constant Eulerian volume. In the current case of the red walk, we would assign the void an Eulerian volume that is substantially V_2, with mass interior to the void given by the value of Λ at which the walk intersects B_{V_2} and mass at the void wall given by the value of Λ at which the walk was tangent to B_{V_2}. The two masses can be quite different indicating that the Eulerian void should be rather well delineated by the surrounding Eulerian overdensity. This indicates that these kind of voids, for which our algorithm differ from that of Sheth and van de Weygaert, are the easiest to observe, while the voids analyzed in the same way by the two algorithms, which are those associated with walks similar to the monotonically decreasing one in figure 3.15, have no obvious defining wall and therefore they are the hardest to observe.
First crossings of δ_v are fundamental especially for Sheth and van de Weygaert, because they give the superset of Lagrangian void candidates from which one discards those which first crossed δ_c, on the basis that they represent voids that would have been crushed out of existence by Eulerian evolution. We account for the squeezing rather than complete crushing of these regions due to Eulerian evolution, but this is not only a modification useful to reduce the predicted volumes of the voids which remain. If the first crossing of δ_v happens to lie in a region where the $\delta - \Delta_{NL}$ mapping of equation 3.20 does not apply, then it is simply not as important as subsequent crossings of δ_v. If we suppose an higher spike in the walk sufficient to cross above B_{V_2} for a while before dropping down to and zigzagging around δ_v a few times the Eulerian region just within V_2 would not be a void, because the walk crossed B_{V_2} above δ_v, but one of the subsequent zigzags around δ_v might actually be the one which first crosses an Eulerian B_V representing a squeezed Eulerian void. A void like this would have a volume smaller than that given to the initial first crossing candidate by Sheth and van de Weygaert, therefore the first crossing of δ_v is no more the most relevant one, although it still plays an important role.
The fact that the first crossing of δ_v is no longer so important is one reason why we have to implement our algorithm numerically in order to study the distribution of void volumes associated with our new approach to the void in cloud process.
In figure 3.16 we show the results of our and Sheth and van de Weygaert Monte Carlo algorithms compared to the associated analytic expressions for walks with uncorrelated steps

and the corresponding results for walks with correlated steps. In case of uncorrelated steps we accumulate independent Gaussian random numbers g_i with a fixed variance, $\delta_j^{(uncorr)} = \sum_{i=1}^{j} g_i$, and record the step at which the barrier δ_v is first crossed as well as the step at which δ_c is first crossed. The black filled and open symbols in figure 3.16 represent the numerical results of the Sheth and van de Weygaert algorithm, respectively the solution to the void in void problem represented by the distribution of Λ at which δ_v is first crossed and the solution to the void in cloud problem by Sheth and van de Weygaert considering of walks for which δ_v is first crossed the subset which has never crossed δ_c before crossing δ_v. The fact that the solid black curves representing the associated analytic expressions go through the symbols indicates that we have agreement with the numerical algorithm.
The numerical algorithm implementing the new model of the void in cloud problem presented in this section is represented in figure 3.16 by the cyan histogram. For a walk that crosses δ_v at least once, we choose all steps prior to the first crossing. Each step j corresponds to a pair (δ_j, Λ_j) which together define an Eulerian volume V_j. The smallest value of V_j associated with the walk, called V_{min}, if it is zero it means that the walk exceeded δ_c, so the void candidate has been crushed out of existence and the walk is eliminated from the list of void walks. Instead, if $V_{min} > 0$ we have a void setting a mass of $M_{min} \approx 0.2\bar{\rho}V_{min}$ and determining a Λ_{min}, a value that is larger than the value of the variance on which the walk first crosses below δ_v. If the walk remains below the barrier $B_{V_{min}}(m)$ for all $\Lambda(m) < \Lambda_{min}$ we store this value and proceed to the next walk. Otherwise we select the first of all steps larger than Λ_{min} which are below δ_v, and repeat the algorithm above until a void is identified, or until Λ_{min} becomes sufficiently large that the associated void size is negligibly small.
In the large mass regime our algorithm predicts about a factor of 2 fewer voids than the two distributions of no void in cloud and Lagrangian void in cloud, that instead are similar for big voids. On smaller scales our prediction predicts more voids than the Lagrangian void in cloud, but still about a factor of 3 less than the distribution of voids when the void in cloud problem has been ignored altogether.
It is interesting to classify the voids identified by our new algorithm in terms of the number of times we have to loop through the algorithm, because we can make an analytic estimate of the fraction of walks $f_0(\Lambda)$ for which a void is identified after only a single pass through the algorithm. Multiplying this analytic estimate by a factor 2 we have

$$\Lambda f(\Lambda) = 2\Lambda f_0(\Lambda) = \frac{\Lambda f_0(\Lambda)}{1 - 1/2} = \Lambda f_0(\Lambda) \sum_{n=0}^{\infty} 2^{-n}, \tag{3.22}$$

that is an excellent description of the full set of void walks represented by the cyan histogram, and this is shown as the dashed line in figure 3.16.
The corresponding results for walks with correlated steps are shown in figure 3.16 through the magenta symbols and histogram. Each walk with uncorrelated steps is transformed into one with correlations by applying smoothing filters of different scales: the filter $W(kR_L)$ is applied to the same set of numbers g_i as above to get $\delta_j^{corr} = \sum_i g_i W(kR_{L_j})$, where R_L is the Lagrangian length scale related to mass m by $m = (4\pi/3)\bar{\rho}R_L^3$. The

correlation depends on the form of the filter and on the shape of the initial linear theory power spectrum $P(k)$, since one needs to know which values of k_j and R_{L_j} to associate with the jth step. This can be done, once we have specified a power spectrum and a filter, by inverting the relations $j\Delta\Lambda = (2\pi^2)^{-1}\int_0^{k_j} dk k^2 P(k)$ and $j\Delta\Lambda = (2\pi^2)^{-1}\int_0^{\infty} dk k^2 P(k) W^2(kR_{L_j})$. We used a Gaussian smoothing filter $W(kR) = e^{-(kR)^2/2}$ and $P(k) \propto k^{-1.2}$. Then we make the same analysis for each correlated walk as for the uncorrelated case.
Differently from the case of uncorrelated steps, here the three ways of estimating void abundances, namely the no void in cloud, the Lagrangian void in cloud and the better model of the void in cloud problem, represented in figure 3.16 respectively by the black filled symbols, the open ones and the magenta histogram, all give almost the same results. It is easy to understand the differences appearing at small masses: to ignore the void in cloud problem overestimates the abundances relative to the Lagrangian void in cloud treatment. In case of a correlated walk this is only a small effect because respect to an uncorrelated walk if it crosses δ_v it is difficult that it has crossed δ_c before. In practice, in a correlated case most walks which cross δ_v do not go into the disallowed region ($> \delta_c$), so removing them makes little difference. Instead, the Lagrangian void in cloud analysis overestimates the abundances relative to our Eulerian void in cloud algorithm because it only does not consider the voids that got completely crushed, but does not alter the sizes of those that got squeezed a little. So the consequence is that it tends to overestimate the sizes of voids, but this effect is important only for small voids.
The first crossing analytic distribution for walks with completely correlated steps in the presence of a single constant barrier of height δ_v, given by the expression

$$\Lambda f_v^{(cc)}(\Lambda) = \frac{1}{2}\frac{\mid \delta_v \mid}{\sqrt{2\pi\Lambda}} e^{-\delta_v^2/2\Lambda}, \tag{3.23}$$

represented by the dotted curve in figure 3.16, provides a good description of our predicted distribution. This prediction considers the walks perfectly smooth, therefore the void in cloud problem becomes no significant. First crossing distributions for walks with correlated steps are relatively insensitive to shape of the underlying power spectrum or the smoothing filter.

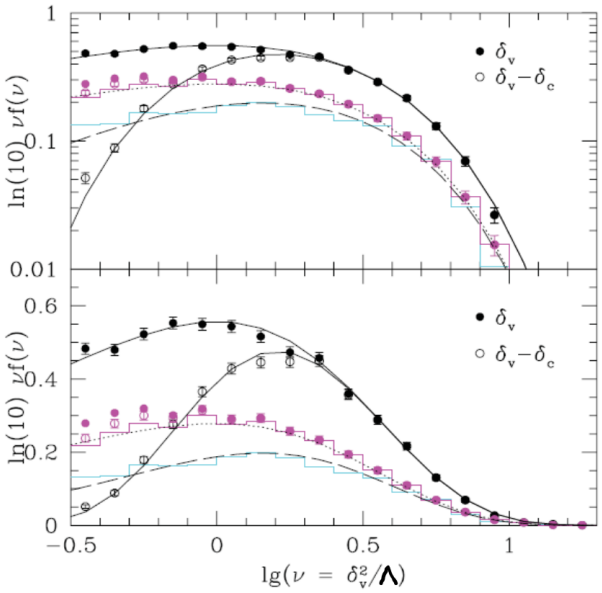

Figure 3.16: Monte Carlo solution of various excursion set based predictions for void abundances. Filled circles show the first crossing distribution of a single barrier of height δ_v; open circles show the distribution of the subset of walks which did not first cross δ_c; histograms show the distribution associated with our new algorithm. The black symbols and cyan histogram are for walks with uncorrelated steps, while the magenta symbols and histogram are for walks with correlated steps. The differences between the symbols and the corresponding histograms are much more pronounced for walks with uncorrelated steps. Solid curves show the corresponding analytic solutions for walks with uncorrelated steps, for the single barrier and two constant barrier cases. Dashed curve shows two times the distribution $\Lambda f_0(\Lambda)$ and provides an excellent description of the cyan histogram. Dotted curve shows the expected solution for walks with completely correlated steps, which describes our results for correlated steps rather well.[4]

Chapter 4

Clustering properties of dark matter haloes

The stochastic approach to the spatial clustering of dark matter haloes in Lagrangian space presented here will be useful to develop our analytical description of the two-point correlation function of voids. The formalism used is based on the formulation of the excursion set approach mentioned in the second chapter. Then we calculate correlation functions of haloes in Lagrangian space using either a multidimensional Fokker-Planck equation with suitable boundary conditions, or an array of Langevin equations with spatially correlated random forces. Our results can be compared with theoretical predictions for the halo autocorrelation function and with numerical simulations and analytical treatments of halo clustering finding more or less agreement depending on the mass range and on the correlation distance. Finally we present a deeper analysis of the halo correlation function considering also Eulerian correlation functions of objects not only with the same mass and the same collapse redshift, but also with combinations of same collapse redshift and different mass, different collapse redshift and same mass and different collapse redshift and mass.

4.1 From joint upcrossing distribution to two-point correlation function

We can exploit the excursion set approach to derive clustering properties of dark matter haloes. We take density fluctuation processes at different spatial locations and then, as the smoothing scale shrinks, we study their evolution. The result is that we notice correlation between the trajectories and from the joint distribution of the first upcrossing filtering radii we calculate the Lagrangian halo-halo correlation function.
Selecting two different points $\boldsymbol{x_1}$ and $\boldsymbol{x_2} = \boldsymbol{x_1} + \boldsymbol{r}$ in Lagrangian space we study as the stochastic processes $\delta_1(\Lambda) = \delta(\boldsymbol{x_1}, \Lambda)$ and $\delta_2(\Lambda) = \delta(\boldsymbol{x_2}, \Lambda)$ evolve, varying Λ. Integrating the system of correlated Langevin equations that describe the evolution of the processes $\delta_1(\Lambda)$ and $\delta_2(\Lambda)$ we obtain the joint probability distribution $f_2(\Lambda_1, \Lambda_2; \boldsymbol{x_1}, \boldsymbol{r})$ of those

variances (Λ_1, Λ_2) that correspond to the first upcrossing scales of the threshold t_f by the two processes $\delta_1(\Lambda)$ and $\delta_2(\Lambda)$, that is fundamental in order to compute the halo-halo correlation function. The probability density f_2 cannot depend on the vector $\boldsymbol{x_1}$ because of the homogeneity and on the orientation of $\boldsymbol{r}$ because of the isotropy, so we can express it as $f_2(\Lambda_1, \Lambda_2; \boldsymbol{x_1}, \boldsymbol{r}) = f_2(\Lambda_1, \Lambda_2; r)$. Now, defined a class of objects selected by the mass interval corresponding to the Λ-range $I \equiv [\Lambda_{min}, \Lambda_{max}]$, it is important to express the probability of determining two points separated by r contained within collapsed objects of class I, namely within two haloes with mass in the intervals $(M_1 - dM_1, M_1)$ and $(M_2 - dM_2, M_2)$, as fixed by the corresponding variance ranges $(\Lambda_1, \Lambda_1 + d\Lambda_1)$ and $(\Lambda_2, \Lambda_2 + d\Lambda_2)$, integrating $f_2(\Lambda_1, \Lambda_2; r) d\Lambda_1 d\Lambda_2$:

$$P_{II}(r) = \int_I \int_I d\Lambda_1 d\Lambda_2 f_2(\Lambda_1, \Lambda_2; r). \tag{4.1}$$

Consequently the probability of finding a point contained in an object of type I is $P_I = \int_I d\Lambda f(\Lambda)$. Then we obtain, for points in the same class I,

$$\xi^{pts}_{II}(r) = \frac{P_{II}}{P_I^2} - 1 = \frac{\int_I \int_I d\Lambda_1 d\Lambda_2 f_2(\Lambda_1, \Lambda_2; r)}{[\int_I d\Lambda f(\Lambda)]^2} - 1. \tag{4.2}$$

Considering disjoint classes I_1 and I_2 the correlation function is

$$\xi^{pts}_{I_1 I_2}(r) = \frac{\int_{I_1} \int_{I_2} d\Lambda_1 d\Lambda_2 f_2(\Lambda_1, \Lambda_2; r)}{\int_{I_1} d\Lambda f(\Lambda) \int_{I_2} d\Lambda f(\Lambda)} - 1. \tag{4.3}$$

The quantities $\xi^{pts}_{II}(r)$ and $\xi^{pts}_{I_1 I_2}(r)$ are the Lagrangian correlations of the mass elements contained in the collapsed haloes. But we have to make a step more in order to calculate the halo correlations $\xi^{hh}_{II}(r)$, weighing properly the statistical contribution for each extended halo. In fact the sets of points where the first upcrossings occur at the same Λ are point like disconnected regions and statistically they form collapsed haloes, each contributing by $1/V(\Lambda)$, where $V(\Lambda) \equiv M/\rho_b$ is the typical Lagrangian volume of an object of mass M associated with the variance Λ. Therefore, knowing that the mean number density of collapsed objects of scale Λ is $n(\Lambda) = f(\Lambda)/V(\Lambda)$, similarly, for the distribution of pairs at distance r, we have

$$n_2(\Lambda_1, \Lambda_2; r) = \frac{f_2(\Lambda_1, \Lambda_2; r)}{V(\Lambda_1) V(\Lambda_2)}. \tag{4.4}$$

Considering haloes with a finite size the halo-halo correlation functions are

$$\begin{cases} \xi^{hh}_{II}(r) = \frac{n_{II}}{n_I^2} - 1 = \frac{\int_I \int_I d\Lambda_1 d\Lambda_2 n_2(\Lambda_1, \Lambda_2; r)}{[\int_I d\Lambda n(\Lambda)]^2} - 1, & \text{for points in the same class } I, \\ \xi^{hh}_{I_1 I_2}(r) = \frac{\int_{I_1} \int_{I_2} d\Lambda_1 d\Lambda_2 n_2(\Lambda_1, \Lambda_2; r)}{\int_{I_1} d\Lambda n(\Lambda) \int_{I_2} d\Lambda n(\Lambda)} - 1, & \text{for points in disjoint classes } I_1 \text{ and } I_2. \end{cases} \tag{4.5}$$

At this point we have seen how to obtain the expressions for the halo correlation functions in case of one class of objects or in case of two disjoint classes of objects. Now we want

to emphasize how to calculate the joint distribution $f_2(\Lambda_1, \Lambda_2; r)$, solving the system of Langevin equations and the two dimensional Fokker-Planck equation associated, either in case of perfectly uncorrelated processes or in case of perfectly correlated ones, but mainly in order to obtain an approximate general solution of $f_2(\Lambda_1, \Lambda_2; r)$.
The evolution of the pair of correlated processes $\delta_1(\Lambda)$ and $\delta_2(\Lambda)$ is governed by the Langevin system of equations

$$\begin{cases} \frac{\partial \delta_1(\Lambda)}{\partial \Lambda} = \zeta_1(\Lambda), & \delta_1(0) = 0, \\ \frac{\partial \delta_2(\Lambda)}{\partial \Lambda} = \zeta_2(\Lambda), & \delta_2(0) = 0, \\ < \zeta_1(\Lambda) >=< \zeta_2(\Lambda) >= 0, & \zeta_1(\Lambda),\ \zeta_2(\Lambda) \text{ Gaussian processes}, \\ < \zeta_1(\Lambda)\zeta_1(\Lambda') >=< \zeta_2(\Lambda)\zeta_2(\Lambda') >= \delta_D(\Lambda - \Lambda'), & \\ < \zeta_2(\Lambda)\zeta_2(\Lambda') >= \frac{\partial \xi(r;\Lambda)}{\partial \Lambda}\delta_D(\Lambda - \Lambda'). & \end{cases} \tag{4.6}$$

The last equation is useful to complete, introducing sharp k-space filtering, the definition of the stochastic field $\zeta(\boldsymbol{x}, \Lambda)$ given in the second chapter. $\xi(r; \Lambda)$ is the linear two-point correlation function for the mass density fluctuations smoothed on the scale $R_f \equiv 1/k_f$ associated with the variance Λ:

$$\xi(r;\Lambda) \equiv \frac{1}{2\pi^2}\int_0^{k_f(\Lambda)} dk k^2 P(k) j_0(kr), \qquad with \quad \frac{\partial \xi(r;\Lambda)}{\partial \Lambda} = j_0[k_f(\Lambda) r]. \tag{4.7}$$

The unconstrained probability distribution for δ_1 and δ_2 at a given Λ that solves the two-dimensional Fokker-Planck equation

$$\frac{\partial F_r(\delta_1, \delta_2; \Lambda)}{\partial \Lambda} = \frac{1}{2}\left[\frac{\partial^2}{\partial \delta_1^2} + \frac{\partial^2}{\partial \delta_2^2} + 2\frac{\partial \xi(r;\Lambda)}{\partial \Lambda}\frac{\partial^2}{\partial \delta_1 \partial \delta_2}\right] F_r(\delta_1, \delta_2; \Lambda), \tag{4.8}$$

with initial condition $F_r(\delta_1, \delta_2; \Lambda = 0) = \delta_D(\delta_1)\delta_D(\delta_2)$ and natural boundary conditions, vanishing at infinity, associated with the Langevin system of equations 4.6, is obtained by integrating the above differential equations and averaging over the ensemble:

$$F_r(\delta_1, \delta_2; \Lambda) = \frac{1}{2\pi\sqrt{\Lambda^2 - \xi(r;\Lambda)^2}} \exp\left\{-\frac{\Lambda(\delta_1^2 + \delta_2^2) - 2\xi(r;\Lambda)\delta_1\delta_2}{2[\Lambda^2 - \xi(r;\Lambda)^2]}\right\}. \tag{4.9}$$

As in the case of the one-dimensional Fokker-Planck equation, adopting the absorbing barrier approach, we impose proper boundary conditions to the equation 4.8 in order to calculate the distribution of the first upcrossings of the threshold t_f by the binary process $\{\delta_1, \delta_2\}$. But in the two-dimensional case the distribution $f_2(\Lambda_1, \Lambda_2; r)$ cannot be found from $F_r(\delta_1, \delta_2; \Lambda)$ simply by differentiation, because the whole binary system automatically disappears as soon as when one Brownian particle is first absorbed. However, a two-step procedure is the key to solve the problem of expressing in the right way the joint distribution $f_2(\Lambda_1, \Lambda_2; r)$.
With the same initial condition, the Fokker-Planck equation is solved with absorbing barriers at $\delta_1 = t_f$ and $\delta_2 = t_f$. Then we find the survival probability density for the

pairs which have never crossed the thresholds and, having found this quantity, we can compute the probability current through each point:

$$\boldsymbol{J}(\delta_1, \delta_2; \Lambda) = -\frac{1}{2}\left[\frac{\partial F_r}{\partial \delta_1} + \frac{\partial \xi(r;\Lambda)}{\partial \Lambda}\frac{\partial F_r}{\partial \delta_2}, \frac{\partial \xi(r;\Lambda)}{\partial \Lambda}\frac{\partial F_r}{\partial \delta_1} + \frac{\partial F_r}{\partial \delta_2}\right]. \tag{4.10}$$

On a boundary wall, $\delta_1 = t_f$, where $F_r(t_f, \delta_2; \Lambda) = 0$, implying $\partial F_r/\partial \delta_2 = 0$, this reduces to

$$\boldsymbol{J}(t_f, \delta_2; \Lambda) = -\frac{1}{2}\left[\left.\frac{\partial F_r}{\partial \delta_1}\right|_{t_f}, \frac{\partial \xi(r;\Lambda)}{\partial \Lambda}\left.\frac{\partial F_r}{\partial \delta_1}\right|_{t_f}\right]. \tag{4.11}$$

Then the flux through any point of the barrier $\delta_1 = t_f$ is given by the scalar product $\boldsymbol{J} \cdot \boldsymbol{n}$, with $\boldsymbol{n} \equiv (1, 0)$ unit vector perpendicular to the absorbing wall:

$$\Phi_r(t_f, \delta_2; \Lambda) = -\frac{1}{2}\left.\frac{\partial F_r}{\partial \delta_1}\right|_{t_f}. \tag{4.12}$$

The quantity $\Phi_r(t_f, \delta_2; \Lambda)d\delta_2$ represents the probability that the pair of processes (δ_1, δ_2) leave the permitted region passing through the gate $[(t_f, \delta_2), (t_f, \delta_2 + d\delta_2)]$ at the time Λ. Once δ_1 has crossed the barrier at Λ_1, for $\Lambda > \Lambda_1$ only the study of the evolution of the surviving process δ_2 up to its first upcrossing through the boundary $\delta_2 = t_f$ is crucial. Therefore, since our case regards Brownian trajectories, free of correlations along the Λ axis, the evolution of δ_2 is simply governed by its own Langevin equation for $\Lambda > \Lambda_1$, and its probability distribution is derived from the one-dimensional Fokker-Planck equation 2.5, considering absorbing boundary $\delta_2 = t_f$ and initial condition $\delta_{2*} \equiv \delta_2(\Lambda_1|\delta_1 = t_f)$ at $\Lambda = \Lambda_1$. Therefore the distribution of the variances Λ_2 associated to first upcrossing events of the threshold by the process δ_2, given that δ_1 upcrossed the critical level at Λ_1, is similar to the equation 2.7 modified considering these new boundary and initial conditions:

$$f(\Lambda_2 - \Lambda_1, t_f - \delta_{2*}) = \frac{(t_f - \delta_{2*})}{\sqrt{2\pi}(\Lambda_2 - \Lambda_1)^{3/2}} \exp\left[-\frac{(t_f - \delta_{2*})^2}{2(\Lambda_2 - \Lambda_1)}\right]. \tag{4.13}$$

With these elements we can construct the joint distribution $f_2(\Lambda_1, \Lambda_2; r)$ by a convolution:

$$\begin{aligned} f_2(\Lambda_1, \Lambda_2; r) = &\int_{-\infty}^{t_f} d\delta_2 \Phi_r(t_f, \delta_2; \Lambda_1) f(\Lambda_2 - \Lambda_1, t_f - \delta_2) + \\ &+ \int_{-\infty}^{t_f} d\delta_1 \Phi_r(\delta_1, t_f; \Lambda_2) f(\Lambda_1 - \Lambda_2, t_f - \delta_1), \end{aligned} \tag{4.14}$$

where the integrals represent the contributions of those pairs for which respectively $\Lambda_2 \geq \Lambda_1$ and $\Lambda_2 < \Lambda_1$.

This expression of the joint distribution is formal. Therefore in order to have a useful quantity for the expression of the joint distribution we have to solve for the probability density F_r, considering different correlation distances, namely two trivial cases for $r \to \infty$ and $r \to 0$ and the general case of finite non zero lag r.

In a perfectly uncorrelated situation with $r \to \infty$ the two processes become independent and the solution of the Fokker-Planck equation is

$$F_\infty(\delta_1, \delta_2; \Lambda) = F(\delta_1, \Lambda; t_f) F(\delta_2, \Lambda; t_f), \tag{4.15}$$

with $F(\delta, \Lambda; t_f)$ probability distribution for a one-dimensional process with absorbing boundary at t_f as denoted in equation 2.6. We present here a solution that is a linear superposition of four unconstrained independent density distributions deriving from different initial conditions, so we have to consider the real distribution $\delta_D(0,0)$, an image source $\delta_D(t_f, t_f)$ and two image sinks $\delta_D(t_f, 0)$ and $\delta_D(0, t_f)$ in order to express $F_\infty(\delta_1, \delta_2; \Lambda)$ as

$$\begin{aligned} F_\infty(\delta_1, \delta_2; \Lambda) = G_\infty(\delta_1, \delta_2; \Lambda) - G_\infty(\delta_1 - 2t_f, \delta_2; \Lambda) + \\ - G_\infty(\delta_1, \delta_2 - 2t_f; \Lambda) + G_\infty(\delta_1 - 2t_f, \delta_2 - 2t_f; \Lambda), \end{aligned} \tag{4.16}$$

with $G(\delta_1, \delta_2; \Lambda) = (2\pi\Lambda)^{-1} \exp[-(\delta_1^2 + \delta_2^2)/2\Lambda]$ solution of the two-dimensional Fokker-Planck equation with boundary conditions: $\lim_{\delta_i \to \infty} G(\delta_1, \delta_2; \Lambda) = 0$, $i = 1, 2$. Therefore, using expression 4.14, we obtain

$$f_2(\Lambda_1, \Lambda_2; r \to \infty) = f(\Lambda_1, t_f) f(\Lambda_2, t_f), \tag{4.17}$$

which inserted in the equations relative to the correlation function gives $\xi^{pts} = \xi^{hh} = 0$ for infinite lag.
Instead, in case of a perfectly correlated situation with $r \to 0$ we will obtain a one-dimensional vision of the quantities needed to express the correlation function. In fact the solution of the Fokker-Planck equation is

$$F_0(\delta_1, \delta_2; \Lambda) = F(\delta_1 + \delta_2, 4\Lambda; t_f) \delta_D(\delta_1 - \delta_2), \tag{4.18}$$

with $F(\delta, \Lambda; t_f)$ probability distribution as given in equation 2.6. Using the method of image sources we can expand this expression as a superposition of Green's functions:

$$F_0(\delta_1, \delta_2; \Lambda) = G_0(\delta_1, \delta_2; \Lambda) - G_0(\delta_1 - 2t_f, \delta_2 - 2t_f; \Lambda), \tag{4.19}$$

where $G_0(x, y; \Lambda) \equiv F(x, \Lambda; t_f) \delta_D(x - y)$. Therefore the joint distribution of first upcrossing variances in this special case becomes

$$f_2(\Lambda_1, \Lambda_2; r \to 0) = f(\Lambda_1, t_f) \delta_D(\Lambda_1 - \Lambda_2). \tag{4.20}$$

Finally we have in case of correlation functions for points in the same class I $\xi_{II}^{pts} \to [1/\int_I d\Lambda f(\Lambda)] - 1$ and $\xi_{II}^{hh} \to \{\int_I d\Lambda f(\Lambda)/V(\Lambda)^2 / [\int_I d\Lambda f(\Lambda)/V(\Lambda)]^2\} - 1$, whereas for points in disjoint classes I_1 and I_2 $\xi_{I_1 I_2}^{pts} \to 0$ and $\xi_{I_1 I_2}^{hh} = 0$.
Contrarily to the limiting cases just discussed, we cannot write a general analytical solution by simply applying the image method, because the position and the sign of the image sources of probability come out dependent on the correlation between the processes, namely on r. Anyway, we will write a simple function that satisfies the absorbing

boundary conditions being also an accurate approximation for the solution of the correlated diffusion equation.
The perfectly correlated solution for small separations $r \ll R_f$ represents a good approximation to the true one. Taken two statistically independent Gaussian processes $\Sigma(\Lambda) = \delta_1(\Lambda) + \delta_2(\Lambda)$ and $\Delta(\Lambda) = \delta_1(\Lambda) - \delta_2(\Lambda)$, $< \Sigma(\Lambda)\Delta(\Lambda) >$, the variances of their unconstrained probability distribution are, respectively, $\sigma^2_\Sigma = 2[\Lambda + \xi(r;\Lambda)]$ and $\sigma^2_\Delta = 2[\Lambda - \xi(r;\Lambda)]$. Therefore, for $r \ll R_f$, that is $\Lambda \ll \sigma^2(r)$, where $\sigma^2(r)$ reflects the variance of the mass density fluctuations smoothed on the scale r, we have $\xi(r;\Lambda) \simeq \Lambda$ and, consequently, $\sigma^2_\Delta = 0$, so that the probability distribution of the variable Δ is practically a Dirac delta function centered on the zero value. But, however, the perfectly correlated situation is not interesting for computing the halo correlation function, because the points in which we follow the trajectories that upcross the threshold are involved in the collapse of the same halo.
For large separations $r \gg R_f$, that is for $\Lambda \gg \sigma^2(r)$, we can finally obtain an approximate solution for the joint distribution transforming the equation 4.8, through the substitution of $\xi(r;\Lambda)$ with the unsmoothed linear mass correlation $\xi_m(r)$, into the uncorrelated two-dimensional diffusion equation that can be solved using the image method. Therefore keeping the analytic form of the solution just obtained for $r \gg R_f$, but inserting in it the correlation function $\xi(r;\Lambda)$ to replace its large lag limit $\xi_m(r)$, we express $F_r(\delta_1, \delta_2; \Lambda)$ as

$$\begin{aligned} F_r(\delta_1, \delta_2; \Lambda) = G_r^+(\delta_1, \delta_2; \Lambda) - G_r^-(\delta_1 - 2t_f, \delta_2; \Lambda) + \\ - G_r^-(\delta_1, \delta_2 - 2t_f; \Lambda) + G_r^+(\delta_1 - 2t_f, \delta_2 - 2t_f; \Lambda), \end{aligned} \tag{4.21}$$

where

$$G_r^\pm = \frac{1}{2\pi\sqrt{\Lambda^2 - \xi(r;\Lambda)^2}} \exp\left\{ -\frac{\Lambda(\delta_1^2 + \delta_2^2) - 2[\pm\xi(r;\Lambda)]\delta_1\delta_2}{2[\Lambda^2 - \xi(r)^2]} \right\}. \tag{4.22}$$

If the term

$$\frac{\partial\xi(r;\Lambda)}{\partial\Lambda} \frac{\partial^2}{\partial\delta_1\partial\delta_2} [G_r^-(\delta_1 - 2t_f, \delta_2; \Lambda) + G_r^-(\delta_1, \delta_2 - 2t_f; \Lambda)] \tag{4.23}$$

can be neglected compared to the Λ-derivative of the expression 4.21 the probability distribution expressed as in 4.21 will be a valid approximation to the proper one. Therefore, using the expression 4.21, thanks to the equation of the flux and to the convolution from which is obtained the joint distribution $f_2(\Lambda_1, \Lambda_2; r)$, this last one can be expressed approximately well as

$$\begin{aligned} f_2(\Lambda_1, \Lambda_2; r) = \frac{t_f^2\Lambda_1\Lambda_2 + [\Lambda_1\Lambda_2 - t_f^2(\Lambda_1 + \Lambda_2)]\xi(r;\Lambda_m) + t_f^2\xi(r;\Lambda_m)^2 - \xi(r;\Lambda_m)^3}{2\pi[\Lambda_1\Lambda_2 - \xi(r;\Lambda_m)^2]^{5/2}} \times \\ \times \exp\left[-\frac{t_f^2}{2} \frac{\Lambda_1 + \Lambda_2 - 2\xi(r;\Lambda_m)}{\Lambda_1\Lambda_2 - \xi(r;\Lambda_m)^2} \right], \end{aligned} \tag{4.24}$$

where $\Lambda_m = \min(\Lambda_1, \Lambda_2)$. By using this expression to compute the halo-halo correlation function between objects selected in infinitesimal mass ranges, we obtain

$$\xi^{obj}(r) \equiv \xi^{pts}(r) = \frac{f_2(\Lambda_1, \Lambda_2; r)}{f(\Lambda_1) f(\Lambda_2)} - 1. \tag{4.25}$$

This last equation, given the values of $f_2(\Lambda_1, \Lambda_2; r)$ and $f(\Lambda, t_f)$ through, respectively, the equations 4.24 and 2.7, represents an approximation to the exact form of the halo-halo correlation function which can be obtained by numerically integrating the correlated Langevin equations.
The halo two-point function can be expressed in powers of the filtered mass autocorrelation function,

$$\xi^{obj}(r) \equiv \xi^{pts}(r) = \sum_{n=1}^{\infty} \frac{1}{n!} b_n(\Lambda_1) b_n(\Lambda_2) \xi(r; \Lambda_m)^n, \tag{4.26}$$

where the factors $b_n(\Lambda)$ coincide with the Lagrangian bias coefficients. The linear bias term dominates the halo correlation at large separation for $M \neq M_*$. In this limit haloes with $M > M_*$ have $b_1 > 0$ and are biased with respect to the mass distribution in Lagrangian space. When $M \gg M_*$ b_1 can be very large. Objects with $M < M_*$ have $-1/t_f < b_1 < 0$ and are moderately antibiased. The limiting case $M = M_*$ have $b_1 = 0$ and the leading term of ξ^{hh} is proportional to $\xi(r; \Lambda_m)^2$, implying much lower halo correlations compared with different mass ranges.

4.2 Lagrangian halo correlation function in the Einstein-de Sitter universe

In order to check the validity of the approximate solution of the joint distribution and, consequently, of the halo correlation function given in the previous section we compare these results with an other analytical model of prediction called 'counting field' and with the numerical solution of the joint distribution of first upcrossing variances and of the halo correlation function by integrating our spatially correlated Langevin equations, outcomes that are better described in detail by the introduction of the 'best fitting models'. In figure 4.1 we see the differences between these models in two different scale free power spectra $P(k) \propto k^n$, with $n = -1$ and $n = -2$, in an Einstein-de Sitter universe, depending on the mass range analyzed and on the distance of correlation. The evolution of clustering is self-similar, and the results obtained at a particular epoch are representative of the whole history. Any physically reliable power spectrum on scales relevant for galaxy formation in a hierarchical scenario presents the two values of the spectral index just indicated. We choose the threshold $\delta_c = 1.686$ and in order to have the linear mass variance as measured in $8h^{-1}Mpc$ spheres equal to 1 we normalize the power spectrum. As regards the mass ranges that identify different classes of haloes, we regulate their broadness in order to optimize the balance between the time required to statistically populate them, that have not to be too narrow, and a precise description of clustering. Therefore we select three different classes of objects for each power spectrum by requiring that they are equipopulated in terms of first upcrossing events, not number of haloes, in order to explore all these regimes of Lagrangian clustering, the $M \gg M_*$ one concerning a biased halo distribution, the $M = M_*$ if we are interested in an almost unclustered distribution, and the $M \ll M_*$ presenting an antibiased distribution, with approximately equal number accuracy. Table 1 gives the parameters that define the three classes of haloes

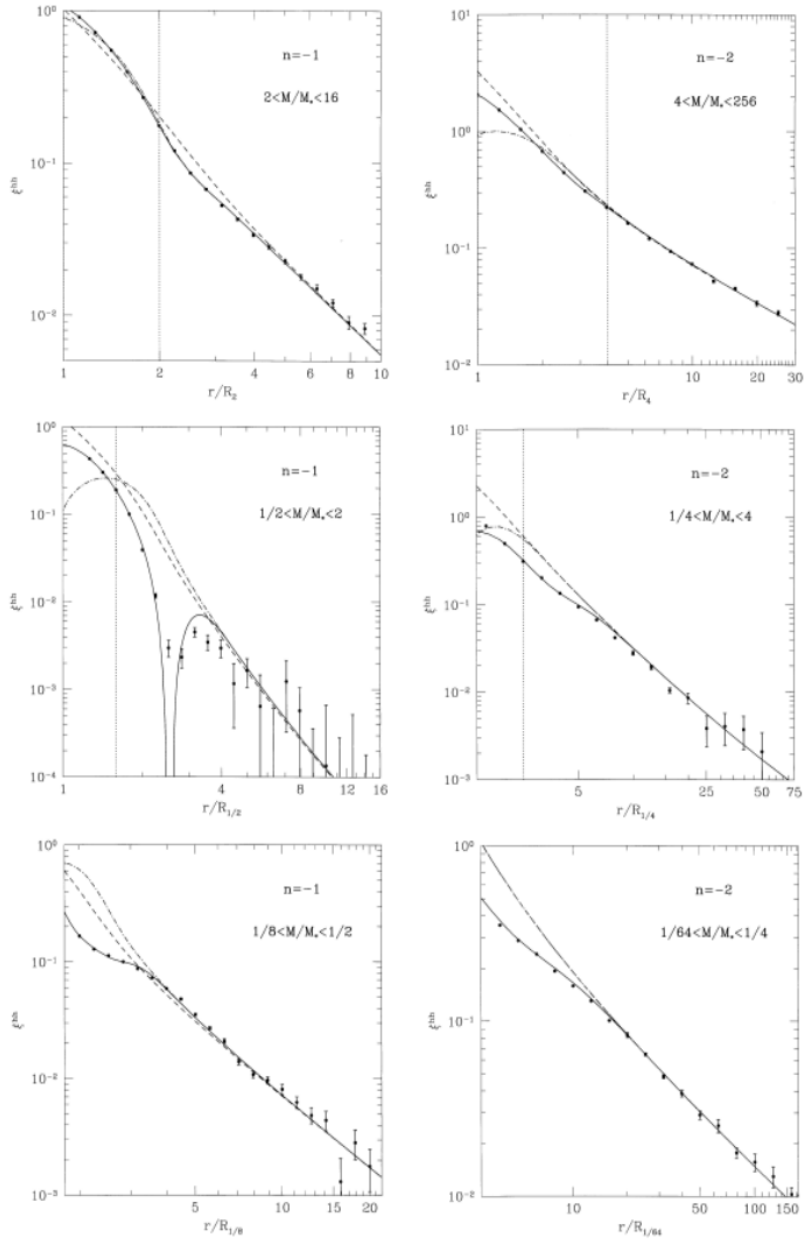

Figure 4.1: The Lagrangian halo correlation function ξ^{hh} in the Einstein-de Sitter universe with two different scale free power spectra, $n = -1$ and $n = -2$, is shown for three halo mass ranges. The object separation r is scaled to the Lagrangian radius of the least massive halo in each range. The vertical dotted lines, where shown, are placed at the Lagrangian radius of the most massive halo in each range. The points represent the mean value of different realizations obtained by numerically solving our correlated Langevin equations, while the error bars represent the scatter of the mean. The continuous lines refer to the best fitting models of equations 4.31 and 4.32. The dashed lines are obtained from our approximated solution of the Fokker-Planck equation 4.29, while the dot-dashed lines show the predictions of the counting field model of equation 4.30. Top-hat filtering is used in all cases.[5]

we choose and the probability of occurrence of first upcrossing events in each of them.

Class	Λ_{min}	Λ_{max}	M_{min}/M_* $(n=-1)$	M_{max}/M_* $(n=-1)$	M_{min}/M_* $(n=-2)$	M_{max}/M_* $(n=-2)$	$\int_{\Lambda_{min}}^{\Lambda_{max}} f(\Lambda)d\Lambda$
I_1	0.45	1.79	2	16	4	256	0.20
I_2	1.79	4.51	1/2	2	1/4	4	0.22
I_3	4.51	11.37	1/8	1/2	1/64	1/4	0.19

***Table*1.** *Parameters that identify the classes of haloes.*

The numerical method gives the exact halo-halo correlation function in the excursion set approach. The stochastic differential equations 4.6 are equivalent to the integral equations

$$\begin{cases} \delta_1(\Lambda+\gamma)-\delta_1(\Lambda)=\int_\Lambda^{\Lambda+\gamma} d\Lambda'\zeta_1(\Lambda'), & \delta_1(0)=0, \\ \delta_2(\Lambda+\gamma)-\delta_2(\Lambda)=\int_\Lambda^{\Lambda+\gamma} d\Lambda'\zeta_2(\Lambda'), & \delta_2(0)=0, \end{cases} \tag{4.27}$$

where the statistical properties of the Gaussian processes ζ_1 and ζ_2 are given in equation 4.6. To solve numerically a stochastic differential equation we replace the equivalent integral equation by its expansion in power series of $\sqrt{\gamma}$, truncate the series after a selected number of terms and give a rule for computing each term that is considered. An extrapolation of the results for $\gamma \to 0$ is required in order to control the effect of the temporal discretization. All this procedure is simplified in the case of a set of Wiener processes, therefore, by integrating over a finite time step γ, the integrals defined in equations 4.27 give

$$\begin{cases} \delta_1(\Lambda+\gamma)-\delta_1(\Lambda)=a_{11}(\gamma,\Lambda)G_1, & \delta_1(0)=0, \\ \delta_2(\Lambda+\gamma)-\delta_2(\Lambda)=a_{21}(\gamma,\Lambda)G_1+a_{22}(\gamma,\Lambda)G_2, & \delta_2(0)=0, \\ a_{11}(\gamma,\Lambda)^2=a_{21}(\gamma,\Lambda)^2+a_{22}(\gamma,\Lambda)^2=\gamma, & \\ a_{11}(\gamma,\Lambda)a_{21}(\gamma,\Lambda)=\xi(r;\Lambda+\gamma)-\xi(r;\Lambda), & \end{cases} \tag{4.28}$$

where the G_i are independent Gaussian variables with zero mean and unit variance. Iterating this set of equations at each time step by modelling the G_i terms with Gaussian pseudo random numbers we produce the trajectories.

The first passage time distribution is determined by solving the discretized stochastic equation starting at the initial point and, then, terminating the simulation of a trajectory as soon as the boundary is reached, accounting for possible intrastep crossings if we do not want to have the resulting distribution being influenced by the temporal discretization. In fact, we are not sure that the process δ has ever crossed the threshold during the time interval γ even with the conditions $\delta(\Lambda) < t_f$ and $\delta(\Lambda+\gamma) < t_f$. Therefore, unless one uses very small time steps, it is inaccurate choosing as first crossing time that corresponding to the first step at which $\delta > t_f$. We solve the problem performing a small Monte Carlo test at each time step, obtaining high precision even using larger time steps, therefore reducing the CPU time.

Given a power spectrum and selected the value for the critical threshold t_f and the lag r, a pair of first upcrossing variances for each realization of the processes δ_1 and δ_2 is determined by the introduction of the algorithm we have just explained. Therefore

through a large number of realizations we calculate the joint probability $f_2(\Lambda_1, \Lambda_2; r)$ and the halo-halo correlation function.
A technical problem we have to deal with in the Langevin formalism, generated by the unsmoothed Bessel function in the system of equations 4.6, which is unavoidable in the Wiener process approach, is the rising of spurious oscillations in the correlation function induced by the sharp k-space filter. In fact, the occurrence of oscillations in the mass correlations computed in the Fourier conjugate space is made unavoidable by the sharpness of the smoothing kernel. This is evident by convolving the linear density field with $\tilde{W}_{SKS}(k, R_f)$, since we obtain $\xi_m(r;\Lambda) \propto \{1 - \cos[k_f(\Lambda)r]\}/r^2$ for $n = -1$ and $\xi_m(r;\Lambda) \propto Si[k_f(\Lambda)r]/r$ for $n = -2$, where $Si(z) \equiv \int_0^z dy j_0(y)$. This oscillating behaviour affects also ξ^{hh}. This problem can be avoided by substituting in equation 4.28 the sharp k-space filtered correlation $\xi(r;\Lambda)$ with the one obtained with top-hat smoothing, analogue to the technique used to compare the mass function predicted by the excursion set approach to the outputs of N-body simulations. Thus we are able to reduce the oscillations, but for $n > -2$ we cannot eliminate them completely at separations comparable to the halo size. Taking as an example the situation for $n = -1$ and top-hat filtering, the term a_{21}, positive for $\Lambda \leq \sigma^2(r)$, negative for $\Lambda \geq \sigma^2(r)$ and rapidly approaching zero for $\Lambda \gg \sigma^2(r)$, after assuming a minimum negative value, defined in equations 4.28, provokes the appearance of oscillations, when r is comparable to the Lagrangian radius of the given halo, in the computation of the halo correlation function. In the following, adopting from the beginning a more realistic window function, namely non sharp k-space, we will always use equations 4.28 and top-hat smoothing to obtain ξ^{hh} in order to solve the problem of oscillations with the only drawback of dealing with a space correlated set of coloured stochastic processes.
Calculating numerically the correlation function, for each physical separation r taken in the range $1 \leq r/R_* < 12$ for $n = -1$ and $1 \leq r/R_* < 40$ for $n = -2$, where R_* is the Lagrangian radius associated to the characteristic halo mass M_*, many realizations of the stochastic processes δ_1 and δ_2 are studied until we have the crossing of the threshold by 10^6 pairs of trajectories at resolutions Λ_1 and Λ_2, both contained in one of the three mass ranges mentioned before. We repeat the simulations 20 times for $n = -1$ and 8 times for $n = -2$ using different sequences of pseudo random numbers to build up the trajectories. All this work is made in order to obtain the right values for $\xi^{hh}(r)$ plotted in figure 4.1 once averaged over the different simulations. The error bars represent the standard deviation of the mean. Through the confrontation with the numerical values for $\xi^{hh}(r)$ we check the validity of two different analytic expressions of the correlation function, our approximated solution of the Fokker-Planck equation,

$$1 + \xi^{hh}(r) = \frac{t_f^2\Lambda_1\Lambda_2 + [\Lambda_1\Lambda_2 - t_f^2(\Lambda_1 + \Lambda_2)]\xi(r;\Lambda_m) + t_f^2\xi(r;\Lambda_m)^2 - \xi(r;\Lambda_m)^3}{\Lambda_1^{-3/2}\Lambda_2^{-3/2}[\Lambda_1\Lambda_2 - \xi(r;\Lambda_m)^2]^{5/2}}$$
$$\exp\left[-\frac{t_f^2}{2}\frac{(\Lambda_1 + \Lambda_2)\xi(r;\Lambda_m)^2 - 2\Lambda_1\Lambda_2\xi(r;\Lambda_m)}{\Lambda_1\Lambda_2[\Lambda_1\Lambda_2 - \xi(r;\Lambda_m)^2]}\right], \tag{4.29}$$

and the correlation function calculated in the 'counting field' model,

$$1+\xi_{cf}^{hh}(r)=\frac{1}{\sqrt{1-\omega^2}}\Bigg\{1+\frac{\sigma_2^2}{(1-\omega^2)}\left(\frac{1}{\sigma_1}-\frac{\omega}{\sigma_2}\right)\frac{\partial\omega}{\partial\sigma_2}+\frac{\sigma_1^2}{(1-\omega^2)}\left(\frac{1}{\sigma_2}-\frac{\omega}{\sigma_1}\right)\frac{\partial\omega}{\partial\sigma_1}+\frac{\sigma_1^2\sigma_2^2}{t_f^2}\frac{\partial^2\omega}{\partial\sigma_1\partial\sigma_2}$$
$$+\frac{\sigma_1^2\sigma_2^2}{t_f^2(1-\omega^2)^2}\left[\omega(1-\omega^2)+(1+\omega^2)\frac{t_f^2}{\sigma_1\sigma_2}-\omega t_f^2\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)\right]\frac{\partial\omega}{\partial\sigma_1}\frac{\partial\omega}{\partial\sigma_2}\Bigg\}$$
$$\exp\left[-\frac{t_f^2}{2}\frac{\omega^2\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)-2\frac{\omega}{\sigma_1\sigma_2}}{(1-\omega^2)}\right], \tag{4.30}$$

analyzing the different plots in figure 4.1.
In the expression of the correlation function in the 'counting field' model $\sigma_i=\Lambda_i^{1/2}$ and $\omega=\xi(\Lambda_1,\Lambda_2;r)/\sigma_1\sigma_2$, with $\xi(\Lambda_1,\Lambda_2;r)$ being the correlation function between the linear mass density fluctuation field smoothed with two different resolutions Λ_1 and Λ_2. For sharp k-space filtering $\xi(\Lambda_1,\Lambda_2;r)=\xi(r;\Lambda_m)$. Moreover, equation 4.30 can be reduced to equation 4.29 for separations larger than the smoothing lengths when the derivatives of $\xi(\Lambda_1,\Lambda_2;r)$ with respect to Λ_1 and Λ_2 are negligible.
Concentrating in the confrontation between the different models we notice that the two analytical models give rise to the same Lagrangian bias factors and to the same clustering regime, if the halo separation is a few times larger than their Lagrangian size. Asymptotically all the models tend to the lowest non vanishing term of the series expansion in equation 4.26. From the different plots in figure 4.1 it is evident that both analytical models agree with the results of the Monte Carlo simulations, except for lags of order the halo Lagrangian size, and the discrepancy becomes larger as the ratio M/M_* decreases. In particular, the details of the numerical solution are not shared by the two models at small separations, where, at least for $n=-1$, the spurious oscillations given by the adoption of top-hat smoothing in equations 4.28, which have been derived after sharp k-space filtering, are relevant.
Introducing an extra modulation through a decaying sinusoidal term in equation 4.30 we can attempt to describe better the numerical outcomes, but they are reproduced in the most precise way, as we can see in figure 4.1, by the 'best fitting models':

$$\frac{\xi^{hh}(r)-\xi_{cf}^{hh}(r)}{1+\xi_{cf}^{hh}(r)}=C_1\cos\left(C_2\frac{r}{R_*}+C_3\right)\exp\left[-C_4\left(\frac{r}{R_*}\right)^2\right], \tag{4.31}$$

$$\frac{\xi^{hh}(r)-\xi_{cf}^{hh}(r)}{1+\xi_{cf}^{hh}(r)}=C_1\cos\left(C_2\frac{r}{R_*}+C_3\right)\exp\left[-C_4\frac{r}{R_*}\right], \tag{4.32}$$

respectively for $n=-1$ and $n=-2$. The problem is that the coefficients $C_\alpha(\alpha=1,2,3,4)$, which are found using the Levenberg-Marquardt non linear least squares method in each mass range, depend both on the shape of the power spectrum and on the halo masses.

4.3 A better analytical description of Lagrangian bias

The wide range of mechanical, chemical and radiative feedback processes happening throughout the universe at all times provoke a lot of nonlocal cosmological problems during structures formation.
Analytic methods were developed by Press and Schechter, Bond, Sheth and others in order to determine the distribution of haloes, but they are limited in predictions about only the average number density of virialized haloes, therefore nothing about their relative positions, because the highly nonlocal effects influencing structures formation in the universe are far outstripping the power of approximate analytic techniques. Even the combination with numerical simulations seems to be useless, because we can already investigate only a limited range of possible scanarios and parameters, and moreover, in cases when simulations are more efficient, frequently we have the lack of an analytical counterpart. However, recently, analytical studies of the two-point correlation function and detailed numerical investigation of the spatial correlation function of the initial positions of cosmological dark matter haloes overcome these limitations. Both in mildly nonlinear and highly nonlinear contexts the modeling of the biased clustering of haloes at a given epoch improved when they are observed in the Eulerian coordinate system. Satisfactory models have also been created in the Lagrangian coordinate system, where the haloes cell centers were originally located. In this reference frame it is much more important to know the total mass of material between two objects rather their final comoving distance in order to calculate the distance between them. Such reference frame is the natural frame for Press and Schechter analytical calculations, which associate peaks in the initial density field with collapsed objects at various redshifts, and makes possible an accurate treatment of the propagation of disturbances between objects, because it is more important the value of the total column depth of material separating two perturbations than the value of their precise distance in physical space.
However, we need to analyze the Lagrangian clustering between haloes forming at different epochs in order to model nonlocal effects, because cosmological disturbances take time to move from one object to the other, thus in this section we will elaborate on an approximate analytical model by Scannapieco and Barkana that considered the collapse of two neighboring points of arbitrary mass and formation redshift, separated by an arbitrary Lagrangian distance, on other analytical models and on detailed numerical Lagrangian halo-halo correlation functions at different masses and redshifts comparing all their results in order to achieve the best possible analysis of the behaviour of the Lagrangian correlation function, considering also nonlinear cosmological influences.
In the following comparisons with analytic correlation functions we use two different detailed numerical simulations given by a parallel OpenMP-based version of the HYDRA code in order to achieve high accuracy at small separations and to minimize the effects of box mode damping. For the cosmological parameters we use the measurements of the cosmic microwave background, the number abundance of galaxy clusters and high redshift supernova distance estimates, namely $h = 0.65$, σ_8^2 the variance of linear fluctuations on the $8h^{-1}Mpc$ scale with $\sigma_8 = 0.87$, $n = 1$ the tilt of the primordial power

spectrum, the total matter density $\Omega_0 = 0.3$, the vacuum density $\Omega_\Lambda = 0.7$, the barionic density $\Omega_b = 0.05$. All these three last values of density are expressed in units of the critical density.
The two different runs we choose are the first (run A), carried out in a cubic volume 78.5 comoving Mpc on a side, populated with 350^3 dark matter particles, more adapt in describing small distances, and the second (run B), carried out in a cubic volume 113 comoving Mpc on a side, populated with 400^3 dark matter particles, more adapt in describing large distances. Since we consider only masses above $10^{11} M_\odot$, for run A, in which each mass particle is $3.9 \times 10^8 M_\odot$, we have over 250 particles in each group used in the study, instead, for run B, in which each mass particle is $7.9 \times 10^8 M_\odot$, we have over 125 particles for each group. Both simulations, performed with 64 bit precision, are integrated from an initial redshift of $z = 49$ down to $z = 1$ and use fixed physical Plummer softening lengths of $5.7kpc$ and $6.9kpc$ for run A and B respectively.
The different groups of particles are identified by the HOP algorithm, which is able, using the local density for each particle, to establish at which point the particle is assigned to the group defined by the nearest local maximum along a path of increasing density. But we have to pay attention to do not assign all particles to groups requiring an outer threshold density δ_{outer}. Moreover, we need the merging of groups, for which the boundary density between them exceeds δ_{saddle}, that are accepted as final groups if they have one particle whose density overcomes δ_{peak}. Therefore, in our case, the HOP parameters, which are $N_{dens} = 48$, $N_{hop} = 20$, $N_{merge} = 5$, $\delta_{peak} = 160$, $\delta_{saddle} = 140$, $\delta_{outer} = 80$, created at redshifts $z = 5, 3, 2, 1$, $1548, 4841, 6179, 6739$ groups, respectively, with masses above $10^{11} M_\odot$ in run A and $4276, 13118, 17063, 18424$ at the same redshifts in run B.
From the resulting cumulative mass functions for both runs we expect our groups to be a good sample of most of the recent N-body simulations.
If we want to define the Lagrangian coordinates of a group we have to trace back the position of each of the particles contained within it to the start of the numerical simulation and then compute the center of mass of this distribution. These positions are fundamental for the numerical construction of the distribution function of haloes at any two given masses M_1 and M_2, output redshifts z_1 and z_2 and initial comoving separation r:

$$\frac{d^2 n}{dM_1 dM_2}(M_1, z_1, M_2, z_2, r). \tag{4.33}$$

However, in the following process it will be useful to compare the numerical and analytical results at different masses and redshifts using the Lagrangian correlation function ξ_L, namely the excess probability of finding two haloes that are initially separated by a comoving distance r, rather than the distribution function of haloes just mentioned. That is

$$\xi_L(M_1, z_1, M_2, z_2, r) + 1 \equiv \left[\frac{d^2 n}{dM_1 dM_2}(M_1, z_1, M_2, z_2, r)\right] \left[\frac{dn}{dM_1}(M_1, z_1) \frac{dn}{dM_2}(M_2, z_2)\right]^{-1}, \tag{4.34}$$

where dn/dM is the overall distribution of haloes at a single mass and redshift.
Now we begin to analyze the behaviour of the Lagrangian correlation function, and also

of the Eulerian one in comoving coordinates, comparing the numerical simulation values with three different analytic models, the Mo and White model, 1996, hereafter MW96, the Scannapieco and Barkana model, 2002, hereafter SB02, the Porciani model, 1998, hereafter P98, that will result more or less adapt to follow the numerical simulation values depending on the correlation distance, on the density threshold ν, on the different ranges of mass and redshift, namely if we are considering dark matter haloes with different masses or with the same mass and with different or same formation redshift, and on the kind of correlation function we study in each particular case, Lagrangian or Eulerian. We will present some figures at the end of this section in order to make evident through plots the results of these comparisons.
Before to concentrate on the study of the different situations, it is useful to divide our sample of groups into three bins at each redshift, containing objects of total mass $10^{11} \leq M < 10^{11.5} M_{\odot}$, $10^{11.5} \leq M < 10^{12} M_{\odot}$, $10^{12} \leq M < 10^{12.5} M_{\odot}$. Since the majority of groups in each bin are dominated by the smallest values, the mean is approximate in the right way considering it equal to the minimum value plus a third of the width, therefore, in the following, we refer to these subsets as bins of mass $10^{11.15} M_{\odot}$, $10^{11.65} M_{\odot}$, $10^{12.15} M_{\odot}$, comparing them with analytic results for these values.
In the figures at the end of the section in constructing ξ_L we take a bin width of $1 Mpc$ and we calculate the error bars considering the standard Poissonian error, which depends on the number of pairs in a given bin, and the additional scatter given by the finite sample size used to build the correlation function.
The variance in the number of pairs in a given bin i in a single simulation is

$$\sigma^2(N_{i,1,2}) = N_{i,1,2} + N_{i,1,2}^2 \left(\frac{2}{N_1} + \frac{2}{N_2} \right), \tag{4.35}$$

where $N_{i,1,2}$, N_1, N_2 are, respectively, the number of pairs in a bin i, the total numbers of objects of each of the two types measured. The total error in each bin i for objects of two different types 1,2 corresponding to the blending of the relative measurements of ξ_L with weights w^A and w^B is

$$\begin{aligned} \sigma^2(\xi_{L,i,1,2}^{tot}) = (w_{1,2}^A)^2 (\xi_{L,i,1,2}^A + 1)^2 (N_{i,1,2}^A)^{-2} \sigma^2(N_{i,1,2}^A) \\ + (w_{1,2}^B)^2 (\xi_{L,i,1,2}^B + 1)^2 (N_{i,1,2}^B)^{-2} \sigma^2(N_{i,1,2}^B), \end{aligned} \tag{4.36}$$

where the weights are based on the total number of pairs with masses 1 and 2 in each simulation at any given redshift z:

$$w_{1,2}^A = N_1^A N_2^A / (N_1^A N_2^A + N_1^B N_2^B), \qquad w_{1,2}^B = N_1^B N_2^B / (N_1^A N_2^A + N_1^B N_2^B). \tag{4.37}$$

An important value that arises in the analytical mass functions is the density threshold

$$\nu \equiv 1.69 D(z)^{-1} \sigma(M)^{-1}, \tag{4.38}$$

where $D(z)$ is the linear growth factor. The density threshold is used by MW96 for the computation of the standard geometrical bias estimates present in their expression of the correlation function:

$$\xi_L(M_1, z_1, M_2, z_2, r) \approx b_L(M_1, z_1) b_L(M_2, z_2) D(z_1) D(z_2) \xi_{DM}(r), \tag{4.39}$$

where $b_L \equiv (\nu^2 - 1)/1.69$ and ξ_{DM} is the underlying matter correlation function, linearly extrapolated to the present time. In the plots showed at the end of this section the analytical estimates of ξ_L by MW96, SB02 and P98 are represented, respectively, by short dashed lines, solid lines and dotted lines. P98 is similar to SB02 but works only for two haloes with the same collapse redshift.
Studying first the Lagrangian correlation function of objects with the same mass at the same redshift, utilizing also the information coming from figure 4.2, we see that there is good correspondence between the numerical results and the two analytical models by SB02 and P98 at all masses and redshifts, even if in the case of the SB02 model we notice a discrepancy from the numerical simulations in the $1 - 2Mpc$ bins, and the difference is more significant in the $10^{12.15} M_\odot$ case, in which the $2Mpc$ bin is missing from all plots. The group finding process causes this phenomenon. Chosen the cosmological parameters we have previously listed, a spherical Lagrangian region, encompassing a mass M, has a comoving radius

$$R_L(M) = 1.9[M/(10^{12} M_\odot)]^{1/3} \quad Mpc. \tag{4.40}$$

This value implies that for the $10^{12.15} M_\odot$ plot the second bin is contained entirely within the value of R_L, and that the same bin is contained within R_L for almost a half for the $10^{11.65} M_\odot$ case. The problem is that any group finding algorithm considers objects at these separations as a single higher mass group, without considering them in the numerical calculation of ξ_L. Also the innermost bin is not interesting for what we are studying, since an object of mass M at a redshift z is always found at a distance of $r = 0$ from an object of equal mass and redshift, namely itself. Therefore the correlation function at zero separation is formally infinite.
The standard model MW96, derived in the limited case of large separations and high values of the density threshold, for $\nu < 1$ becomes negative and cannot reproduce ξ_L within a distance of ~ 4 comoving Mpc or for values of the density threshold less than 1.5. Therefore for these cases we do not consider this model.
More standard estimates need to compare analytical models and numerical simulations through the study of the Eulerian correlation function in comoving coordinates. In figure 4.3 it is plotted the Eulerian correlation function of objects with the same mass at the same redshift. In this case we introduce in the determination of the quantities of interest the fact that haloes move toward each other, therefore the bias factor appearing in the MW96 prediction is modified to $b_E = (\nu^2 - 1)/1.69 + 1 = b_L + 1$. Our numerical sample is confirmed by the agreement with the estimates at large separations. Moreover, a numerical study of several 256^3 simulations found that the $b_E = b_L + 1$ mapping between Lagrangian and Eulerian coordinates provided a good fit to these correlation functions at large distances. However, plotting in figure 4.3 the bias estimated as

$$b_E(z, r) = 1 \pm \sqrt{\mid \xi_L(r) D(z)^{-2} \xi_{DM}(r)^{-1} \mid}, \tag{4.41}$$

where $\xi_L(r)$ is the Lagrangian correlation function as calculated by SB02 or P98 and the $\pm$ sign is taken to be positive if $\nu \geq 1$ and negative if $\nu < 1$, without considering the innermost bin, we notice that for the redshift $z = 5, 3, 2$ this Eulerian to Lagrangian mapping seems to hold, but as ν falls down approaching the one value this approximation

breaks down, making necessary to find a more precise mapping in order to estimate these correlation functions at small separations.
Now we concentrate on the Lagrangian correlation function of two dark matter haloes with different masses at the same redshift. As we can see from figure 4.4 all analytical models agree with the numerical results at large separations and high values of the density threshold. Moreover, also in this case MW96 prediction is wrong when ν approaches 1, while P98 and SB02 models are more adapt to this range of parameters. However, differently from the case of the same mass and redshift, at small separations only the SB02 model is appropriate, because at these distances the MW96 model is too simplified and the P98 model implements in the wrong way the barriers that exclude the formation of a density peak within a larger peak in the excursion set formalism. These same barriers, instead, are properly accounted for in the SB02 model, so that in this formalism it is impossible the presence of two dark matter haloes of different masses at the same position and redshift.
Also for the case of Eulerian correlation function of objects with different masses at the same redshift, which results are plotted in figure 4.5, at large separations, namely for $\nu > 1$, our simulations are correctly estimated by the MW96 model, namely considering $b_E = (\nu^2 - 1)/1.69 + 1$, and by the P98 and SB02 models, whose results are mapped to this coordinate system by equation 4.41. In this case, unlike the Lagrangian one, no exclusion is seen in the numerical results at small separations. Two dark matter haloes whose initial centers of mass are separated can move toward each other to very close distances as collapse and virialization decrease the radius of a perturbation by a factor $\sim 1/6$. Therefore, since in the Eulerian case no exclusion to be close to each other for two different objects is imposed, the P98 and the MW96 models are better approximations for these separations than the SB02 model. In particular, the MW96 model fits best the data, while the other two models fail at the redshift $z = 1$. Despite this agreement, the relationship between Eulerian and Lagrangian spaces is always complex for low ν values and small separations.
All that we have just said regards our analytical modeling of single redshift collapsing haloes, and the comparisons coming from the analysis of the situation with these conditions of formation of the dark matter haloes give us confidence. In reality, cosmological disturbances take a certain amount of time to propagate from their sources to neighboring objects. When feedback effects reach the neighboring perturbations before their virializing they are more efficient. Any analytical model of nonlocal effects in structure formation has to adequately treat these situations. Therefore the study of the Lagrangian correlation function of dark matter haloes formed at different redshift is fundamental.
The positions of galaxies measured at a redshift z_1 can be correlated with the ones measured at an earlier redshift z_2. But if the correlation distance between the galaxies is $r > c[t(z_1) - t(z_2)]$ the light from the z_2 galaxy, as we see it today, cannot reach the z_1 galaxy, as we observe it, because such distances, ≥ 1000 comoving Mpc, are too long for our output times and well beyond our ability to simulate. Therefore, we consider distances $r \ll c[t(z_1) - t(z_2)]$ much more useful for the propagation of cosmological disturbances, although not directly observable.

In figure 4.6 we concentrate on two dark matter haloes with the same mass and different formation redshifts. The evolution of the numerical simulations at large separations is almost the same as in figure 4.2. Both the MW96 and SB02 models reproduce the numerical values of the correlation, stronger for rarer objects and weaker for smaller masses and later redshifts, between haloes, even if only the SB02 model can be used for $\nu_1 \leq 1$ or $\nu_2 \leq 1$. The P98 model is not represented in figure 4.6, because it is adapt only at a single redshift.
Instead, the two redshifts case is completely different from the single redshift case when we study the system at smaller separations. Since it is impossible to find an object with the same mass at two different redshifts, at small separations each numerical value approaches $\xi_L = -1$. Thus at this regime the MW96 model and the SB02 one behave differently. The first one overestimates the small distance clustering of two haloes with the same mass and different collapse redshifts, while in the previous case of two haloes of the same mass and redshift it underestimates the formally infinite small distance clustering. On the other hand, the second one considers impossible the formation of a halo of the same mass at the same position at different redshifts. That is why the resulting curves follow the simulations, turning over at small separations. In fact, for the SB02 model, there is a good correspondence between the analytical modeling and the numerical results at all masses, redshifts and separations, even if in the plot relative to the cross correlation between haloes formed at redshifts $z_1 = 1$ and $z_2 = 5$ at the low mass regime the SB02 prediction underestimates the numerical values.
As last case we put our attention on the determination of the Lagrangian correlation function of objects with different masses and redshifts, that analytically it is the most difficult situation to study. First we analyze the case in which the smaller dark matter halo forms at higher redshift. In practice, we are interested in the probability that an object of mass M_{progen} is found at a redshift z_{progen} at a position at which an object with a mass M_{final} is known to exist at a redshift z_{final}. Lacey and Cole (1993) were the first to build analytically this distribution in the excursion set formalism context. In figure 4.7 we notice, as in the previous cases, that the SB02 model reproduce well the numerical values at large separations for all combinations of masses and redshifts. In particular it coincides at $r = 0$ with the Lacey and Cole estimates. Moreover, also in this case we find different calculated quantities by the numerical simulation and analytical model at small separations. This is primarily due to the fact that at small separations we consider correlation distances $r \leq R_L(M)$. Bins with a $1Mpc$ width mean that changes in the position of the Lagrangian center of mass of the smaller dark matter halo within the larger one can move power between the leftmost bins. Therefore numerically the position of the smaller objects matters. Instead, the analytical estimates reproduce the total number of M_2 haloes merging into M_1, without considering of where they are in the final halo. Therefore, choosing an inner bin width that is representative of the size of the final halo is the only way to fairly compare our analytical and numerical results.
The shaded regions presented in figure 4.7 represent the range of $\xi_L(0)$ values given by the recalculation of the numerical correlation function at zero separation over a range of bin widths from $R_L(M_1)$ to $R_L(M_1) + R_L(M_2)/2$, with M_1 mass of the larger halo.

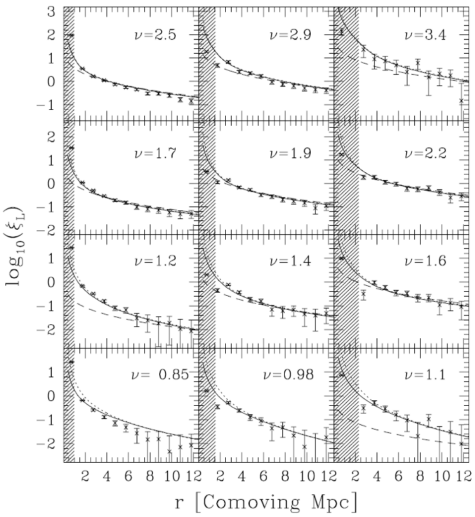

Figure 4.2: Lagrangian correlation function of objects with the same mass at the same redshift. From top to bottom $z = 5$, $z = 3$, $z = 2$, $z = 1$, and from left to right the mass bins are centred on $10^{11.15}M_{\odot}$, $10^{11.65}M_{\odot}$, $10^{12.15}M_{\odot}$. The points are the simulation values, the solid lines are the SB02 model, the dotted lines are the P98 model, the dashed lines are the simple MW96 geometric bias estimates. The shaded regions are within the spherical radius of the halo as defined by equation 4.40. We omit the MW96 estimate in cases in which $\nu \leq 1$.[6]

Larger the bin width is, weaker the overall correlation function is. Comparing these shaded regions with the $r = 0$ predictions of the progenitor distribution made by Lacey and Cole, represented in figure 4.7 by the horizontal dotted lines, we notice a general correspondence if the difference between masses and redshifts is small. Instead, for big differences in mass and redshift these models are not so similar any more. Therefore, in figure 4.7 are represented the limits of trying to express the standard progenitor outside of the M_1/M_2 and $D(z_1)/D(z_2)$ ranges in which originally worked Lacey and Cole. We need more advanced models for progenitor distributions in order to understand this behaviour.
Finally, in figure 4.8 we present the same plots that in figure 4.7, but inverting the progenitor problem, assigning an earlier formation redshift to the dark matter halo with the bigger mass. No such pair of structures can exist at $r = 0$, because the gravitation only increases the mass of any structure with time. Therefore ξ_L approaches -1 at small separations. In figure 4.8 what provokes the behaviour at $r = 0$ in figure 4.7 is replaced by simple exclusion, and the SB02 results are adapt to model the simulations for any parameters chosen. This turnover needs $\xi_L(r)/\xi_{DM}(r)$ to vary with radius. At the same time, MW96 estimates are widely discrepant for small correlation distances, because they cannot follow these conditions.

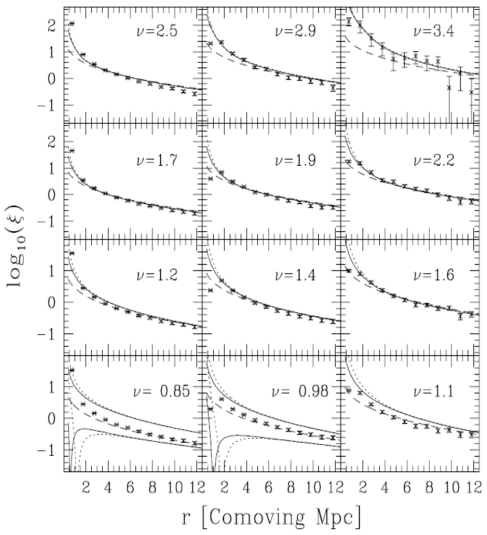

Figure 4.3: Eulerian correlation function of objects with the same mass at the same redshift. Panels are as in figure 4.2. The points are the simulation values, the dashed lines are the simple MW96 Eulerian geometric bias estimates. The solid and dotted lines are the SB02 and P98 models, respectively, mapped to Eulerian coordinates as described in the text. In the cases in which $\nu \leq 1$, both possible values for b_E as in equation 4.41 are shown.[6]

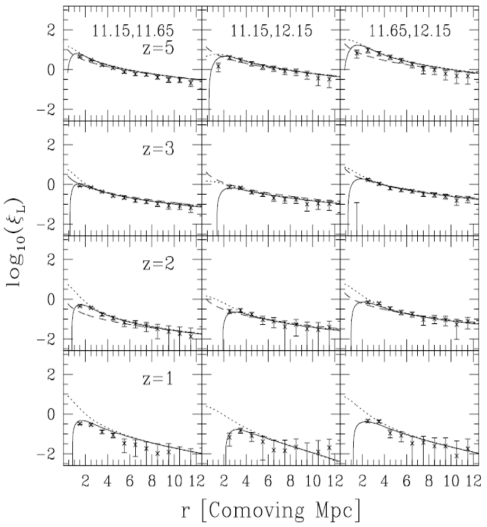

Figure 4.4: Lagrangian correlation function of objects with different masses at the same redshift. From top to bottom $z = 5$, $z = 3$, $z = 2$, $z = 1$, and from left to right the mass bin pairs are centered on $(10^{11.15}M_\odot, 10^{11.65}M_\odot)$, $(10^{11.15}M_\odot, 10^{12.15}M_\odot)$, $(10^{11.65}M_\odot, 10^{12.15}M_\odot)$, respectively. Lines and points are as in figure 4.2, with the MW96 estimate omitted if $\nu \leq 1$.[6]

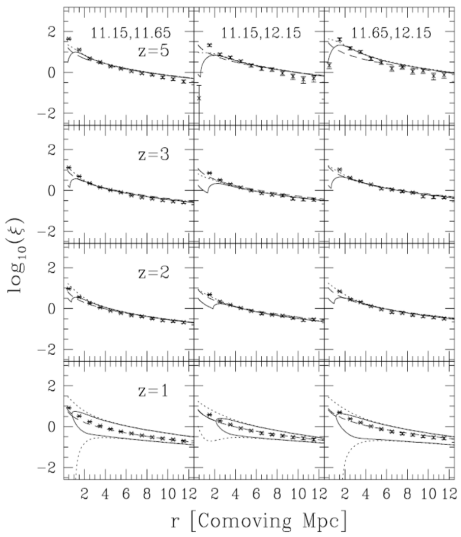

Figure 4.5: Eulerian correlation function of objects with different masses at the same redshift. Panels are as in figure 4.4. The points are the simulation values, the dashed lines are the simple MW96 Eulerian geometric bias estimates. The solid and dotted lines are the SB02 and P98 models, respectively, mapped to Eulerian coordinates as described in the text. In the $z = 1$ cases, both possible values for b_E as in equation 4.41 are shown.[6]

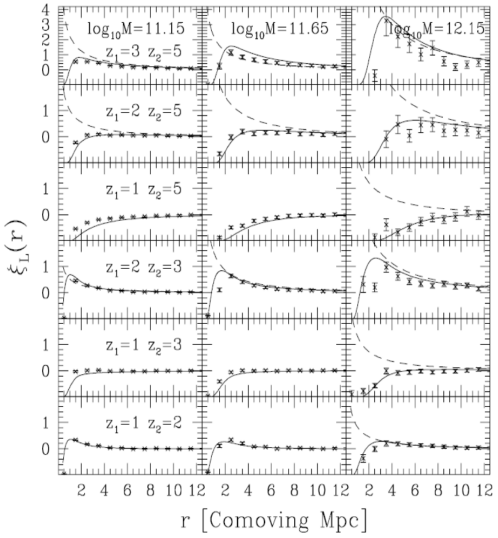

Figure 4.6: Lagrangian correlation function of objects with the same mass at different redshifts. Each row is labeled by its z values, and each column is labeled by its M value. In all panels the solid lines are the SB02 model, the dashed lines are the MW96 model in the cases in which $\nu \geq 1$ for both haloes. Note that we plot ξ_L rather than $\log_{10}(\xi_L)$ as the correlation function becomes negative at small separations.[6]

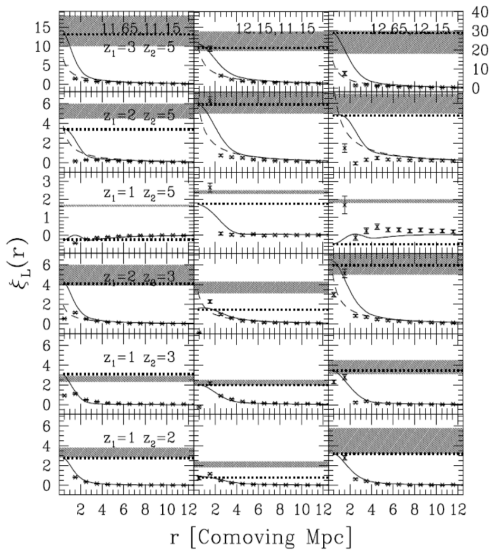

Figure 4.7: Lagrangian correlation function of objects with different masses and redshifts, with the smaller mass at the higher redshift. In this figure, the points are the simulation results, the dashed and solid curves are the MW96 and SB02 models, respectively, and the horizontal dotted lines are the $r = 0$ estimates as given by the Lacey and Cole (1993) progenitor distribution. Finally, the shaded bands are the range of progenitor number densities measured from the simulations. Each row is labeled by its z values, and each column from left to right corresponds to mass bins centered on $(10^{11.15}M_{\odot}, 10^{11.65}M_{\odot})$, $(10^{11.15}M_{\odot}, 10^{12.15}M_{\odot})$, $(10^{11.65}M_{\odot}, 10^{12.15}M_{\odot})$. The y axis label on the right only applies to the upper panel at the extreme right.[6]

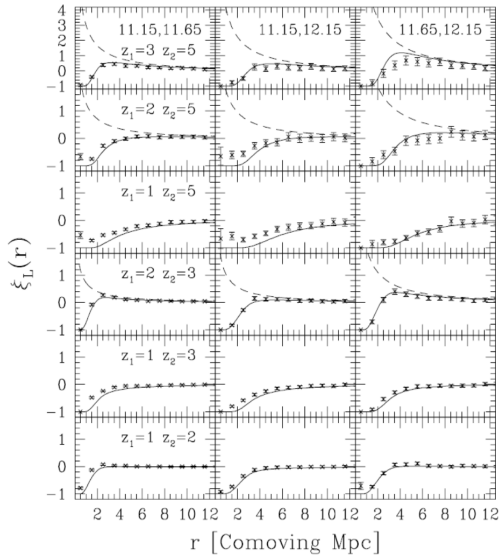

Figure 4.8: Lagrangian correlation function of objects with different masses and redshifts, with the larger mass at the higher redshift. Rows and columns are as in figure 4.7.[6]

Chapter 5

Description of the two-point correlation function of voids

Cosmic voids can be an interesting probe of large-scale structure, accounting for the bulk of the cosmic web. Moreover, voids are easily observable in modern galaxy surveys. Thanks to their low matter content, voids are good laboratories in order to study general relativity, dark energy models or inflationary non Gaussianities. Contrarily to haloes, voids evolve simpler and virialize less, therefore they maintain better the initial conditions. But for many years voids have been studied less than haloes, because voids occupy large volumes and so galaxy surveys need to cover both larger volumes and reach higher sampling densities at the same time respect to haloes. In fact, ranging from a few to over 100 Mpc, void size influence significantly the clustering statistics. However, recent galaxy surveys like the SDSS have produced void catalogs suitable for statistical analyses and through various possibilities of void identification, based for example on the watershed algorithm, useful because it does not need any prior morphology of voids and is parameter free, we have the possibility to study the clustering properties of voids, with the aim of extracting all the cosmological information we can from void measurements. For this purpose we try to build a two-point correlation function of voids through the definition of a joint probability distribution, which corresponds to the first upcrossing scales of the threshold by the two stochastic processes, whose evolution creates the relative voids, following the idea developed in [5] for haloes, considering that in the case relative to the voids we have two barriers, so we have to face some changes in the expression for the joint probability distribution obtained by the convolution, making important considerations about which value of the threshold associated with the collapse of clouds is better to choose. As second method of analyses of the two-point correlation function of voids we will model the void power spectrum knowing that the distribution of voids is biased relative to the underlying dark matter distribution, like in halo clustering studies. But for voids we have to include exclusion effects in their clustering statistics, since voids are spatially more extended than haloes.
Understand how to handle these effects in order to build a plausible two-point correlation function of voids will be essential to extract cosmological information from the large-scale

clustering of voids in the future.

5.1 A model for the two-point correlation function of voids

We have already mentioned how the analysis of voids is important in order to study cosmological characteristics that are difficult to express through the results derived from the calculations of the clustering of haloes.
Therefore, now we try to understand how to explicit an analytical expression of the two-point correlation function of voids, that is a good physical instrument to describe the clustering of voids and consequently to make possible the determination of a series of interesting cosmological information, using the idea presented in the chapter 4 for the calculation of the two-point correlation function of haloes from the determination of the joint probability distribution, but keeping in mind all the complications arising for voids respect to haloes, about the introduction of a second barrier and the determination of the threshold relative to the collapse depending on how we interpret the phenomenon of the void in cloud and the distribution of void sizes, in order to build a plausible analytical expression for the joint probability distribution of voids necessary to finally determine the two-point correlation function of voids.
In an Einstein-de Sitter universe the total mass fraction in voids does not evolve, but singularly the mass fraction of voids depends on the value of the density $\bar{\rho}(1+\delta)$, that fluctuates from cell to cell. In the excursion set approach, voids in denser cells are smaller, have a narrower size distribution, and account for a smaller fraction of mass in the cell they inhabit. For a cell of volume V, within which the density is $\bar{\rho}(1+\delta)$, the mass is $M = \bar{\rho}V(1+\delta)$. In the spherical evolution model the initial and final perturbations are related:

$$\delta_0(\delta) = \frac{\delta_{sc}}{1.68647}\left[1.68647 - \frac{1.35}{(1+\delta)^{2/3}} - \frac{1.12431}{(1+\delta)^{1/2}} + \frac{0.78785}{(1+\delta)^{0.58661}}\right], \tag{5.1}$$

with δ_0 and δ having the same sign. Initially dense regions become denser, whereas the comoving density in underdense regions decreases with time.
In the void model studied here random walks do not start from the origin $[\Lambda = 0, \delta_0 = 0]$, but from the position $[\Lambda(M), \delta_0(\delta)]$. Therefore, making the following substitution in equation 3.3, that is $\delta_c \to \delta_c - \delta_0(\delta)$, $\delta_v \to \delta_v - \delta_0(\delta)$, $\Lambda \to \Lambda(m) - \Lambda[\bar{\rho}V(1+\delta)]$, and integrating the resulting distribution over $0 \leq m \leq M$ we obtain the fraction of the total mass $M = \bar{\rho}V(1+\delta)$ that is in voids of mass m:

$$f_{void}(\delta) = \frac{\delta_c - \delta_0(\delta)}{\delta_c - \delta_v}. \tag{5.2}$$

This means that the mass fraction $f_{void}(\delta)$ decreases as the density δ of the cell increases. Instead, $f_{void}(\delta) \to 1$ if $\delta_0(\delta) \to \delta_v$ as the density we associate with a void $\delta \to -0.8$. Thus this analysis explains why dense regions have a smaller fraction of their mass in voids.
The typical void size scales as $\Lambda(m) \approx \Lambda(M) + \mid \delta_v - \delta_0(\delta) \mid$, where the void size $R(m)$

decreases as $\Lambda(m)$ increases, and, moreover, is larger in regions of lower density because $\mid \delta_v - \delta_0(\delta) \mid$ increases as δ increases.
The relation $\delta_c / \mid \delta_v \mid$, representing how the process of void formation is subjected to the void in cloud demolition, influences the sharpness of the peak in the void size distribution, that becomes more sharply peaked as void in cloud assumes more importance. Therefore, in dense regions ($\delta > 0$), where voids are more likely to be squeezed by collapsing clouds, the distribution of void sizes is expected to be narrower.
The distribution of void sizes is extremely influenced by the behaviour of the random walks, so also the determination of the two barriers, fundamental to understand the evolution of voids including the processes of the void in void and the void in cloud, is characterized by the variation of the path of the density contrast $\delta(t)$, depending on time. In fact, in the last section of the chapter 3, through the Eulerian treatment of the spherical evolution of voids we have determined that the threshold δ_c, useful to understand how the collapse of clouds influences the distribution and the evolution of the relative voids they surround, varies depending on the mass m and on the respective volume $V = 4\pi R^3/3$ at time t of the region considered:

$$\Delta_{NL}(t) = \frac{m}{\bar{\rho}V} \approx \left(1 - \frac{\delta(t)}{\delta_c}\right)^{-\delta_c}, \tag{5.3}$$

$$B_V(m) = \delta_c \left[1 - \left(\frac{m}{\bar{\rho}V}\right)^{-1/\delta_c}\right]. \tag{5.4}$$

Also the role played by the threshold δ_v is different consequently to the fact that void candidates are different depending on the path followed by the random walks. We have seen that thanks to this new more complete approach to the evolution of voids the first crossing of δ_v is not necessarily the most relevant one, but it is important as the subsequent. However, all these considerations make impossible to derive an analytic expression for the distribution of void volumes associated with this new formulation of the void in cloud problem and for the joint probability distribution useful to calculate the two-point correlation function of voids. In fact, the process to derive these quantities, as we have already seen in chapter 4 for haloes, is based to the fact that we consider fixed values for the two barriers, with the determination of the first crossing of δ_v fundamental in order to write analytically an expression for the joint probability distribution corresponding to the first upcrossing scales of the threshold and, consequently, for the two-point correlation function of voids. Therefore, in the expression of the joint probability distribution we use the value δ_v, corresponding to the initial density for voids, and, moreover, we adopt a value for the barrier δ_c, important to determine the influence of the void in cloud, that is a weighted average between the two values that delimitate the range of all possible δ_c, as Sheth and van de Weygaert observed, namely between the linearly extrapolated overdensity at turnaround and at virialization, or, according to the spherical collapse model, between 1.062 and 1.686, respectively:

$$\bar{\delta}_{mean} = w_c \delta_c + w_{ta} \delta_{ta}, \tag{5.5}$$

where we choose our weights equal to

$$w_c = \frac{N_c}{N_c + N_{ta}} \qquad w_{ta} = \frac{N_{ta}}{N_c + N_{ta}}, \tag{5.6}$$

with N_c, N_{ta} respectively the number of voids completely squeezed by the collapse of their surrounding clouds and the number of voids formed before the complete contraction of their surrounding clouds, thus when they begin to turnaround.
This way of considering the two barriers is an idea in order to find analytical expressions for the first upcrossing distribution of voids and for the joint probability distribution useful to obtain a two-point correlation function of voids as plausible as possible, following the way of proceeding for haloes we have already described in chapter 4. Therefore, taking the pair of processes (δ_1, δ_2), if it is δ_1 the first process that crosses the barrier at Λ_1, we write the first upcrossing distribution of voids with initial condition $\delta_{2*} \equiv \delta_2(\Lambda_1 \mid \delta_1 = \delta_v)$ as:

$$\begin{aligned} f(\Lambda_2 - \Lambda_1, \delta_v - \delta_{2*}, \bar{\delta}_{mean}) &= f(\Lambda_2 - \Lambda_1, \delta_v - \delta_{2*}) \\ &\quad - \int_{\Lambda_1}^{\Lambda_2} f(\Lambda', \bar{\delta}_{mean}) f(\Lambda_2 - \Lambda_1, \delta_v - \delta_{2*} \mid \Lambda', \bar{\delta}_{mean}) d\Lambda'. \end{aligned} \tag{5.7}$$

If, instead, it is δ_2 the first process that crosses the barrier at Λ_2, we write the first upcrossing distribution of voids with initial condition $\delta_{1*} \equiv \delta_1(\Lambda_2 \mid \delta_2 = \delta_v)$ as:

$$\begin{aligned} f(\Lambda_1 - \Lambda_2, \delta_v - \delta_{1*}, \bar{\delta}_{mean}) &= f(\Lambda_1 - \Lambda_2, \delta_v - \delta_{1*}) \\ &\quad - \int_{\Lambda_2}^{\Lambda_1} f(\Lambda', \bar{\delta}_{mean}) f(\Lambda_1 - \Lambda_2, \delta_v - \delta_{1*} \mid \Lambda', \bar{\delta}_{mean}) d\Lambda'. \end{aligned} \tag{5.8}$$

This quantity, together with the expression for the flux $\Phi_r(\delta_v, \delta_2; \Lambda_1) d\delta_2$, that represents the probability that the pair of processes (δ_1, δ_2) leave the permitted region passing through the gate $[(\delta_v, \delta_2), (\delta_v, \delta_2 + d\delta_2]$ at the time Λ_1, in case the first process that crosses the barrier is δ_2 we have the same quantities exchanging only the term 1 with 2, we obtain all the information we need for the computation of the joint probability distribution of voids $f_2(\Lambda_1, \Lambda_2; r)$. In fact, once one of the two processes has crossed the barrier, we are interested in studying only the evolution of the surviving process up to its first upcrossing through the boundary. Therefore, since we are considering Brownian trajectories, free of correlations along the Λ axis, the evolution of each process is governed by its own Langevin equation.
The joint probability distribution is obtained by a convolution:

$$\begin{aligned} f_2(\Lambda_1, \Lambda_2; r) &= \int_{-\infty}^{\delta_v} d\delta_2 \Phi_r(\delta_v, \delta_2; \Lambda_1) f(\Lambda_2 - \Lambda_1, \delta_v - \delta_{2*}, \bar{\delta}_{mean}) \\ &\quad + \int_{-\infty}^{\delta_v} d\delta_1 \Phi_r(\delta_v, \delta_1; \Lambda_2) f(\Lambda_1 - \Lambda_2, \delta_v - \delta_{1*}, \bar{\delta}_{mean}), \end{aligned} \tag{5.9}$$

where the first and second integrals on the right-hand side represent, respectively, the contributions of those pairs for which $\Lambda_2 \geq \Lambda_1$ and $\Lambda_2 < \Lambda_1$.
By using this expression to compute the two-point correlation function of voids we obtain

$$\xi_{vv}(r) = \frac{f_2(\Lambda_1, \Lambda_2; r)}{f(\Lambda_1) f(\Lambda_2)} - 1. \tag{5.10}$$

5.2 Clustering of cosmic voids

The clustering of void centres can be estimated also writing the two-point correlation function of voids that contain mass m_1 and m_2 as

$$\xi_{vv}(r \mid m_1, m_2) = b(m_1) b(m_2) \xi_{dm}(r), \tag{5.11}$$

where ξ_{dm} is the correlation function of the dark matter, and the bias factor $b(m)$ depends on the mass or size of the voids. Knowledge of the number density of objects is sufficient to estimate their spatial distribution, at least on large scales. Therefore, $b(m)$ depends on which estimate of $n_v(m)$ we use.
What we have just said indicates that before to analyze the two-point correlation function of voids through the bias respect to the distribution of dark matter, either analytically with the void bias from the peak-background split (PBS) or numerically with simulations, we compare the numerical void size distribution resulting from the simulations we use with the theoretical one, presented in chapter 3 about voids, obtained from the excursion set formalism.
The simulations used in this section in order to determine the numerical results relative to the void size distribution and to the void cross power and auto power spectrum, which are more useful than the cross and two-point correlation function in the process of analysis of the clustering properties of voids because they are easier to handle, are two different ones with box sizes $1500(h^{-1}Mpc)$ (six realizations) and $250(h^{-1}Mpc)$ (three realizations), abbreviated as L1500 and L250 later on. In each simulation there are 1024^3 particles. The cosmology is a flat ΛCDM model, with the WMAP 7 cosmological parameters adopted: $\Omega_m = 0.272$, $\Omega_\Lambda = 0.728$, $\Omega_b = 0.0455$, $\sigma_8 = 0.81$. Thus, for the large box each particle carries a mass of $2.37 \times 10^{11} M_\odot h^{-1}$, while this value is $1.10 \times 10^9 M_\odot h^{-1}$ for the small box. The combination of large and small box sizes enables us to capture a wide range in void sizes and to conduct a resolution study. We use Gaussian initial conditions with a spectral index of $n_s = 0.967$. Voids are identified in the dark matter distribution and as void finder we use a Voronoi tessellation method already mentioned in section 3 of chapter 3 about the volume conserving model. The tracer sampling density is a crucial quantity of the void finding process. We randomly exclude tracer particles from our simulations to achieve different degrees of subsampling in order to understand how much void properties depend on the tracer sampling density. For L1500 we use $0.02(h^{-1}Mpc)^{-3}$. For L250 we use three different sampling densities, namely $2, 0.2, 0.02(h^{-1}Mpc)^{-3}$. The sample relative to the low sampling density $0.02(h^{-1}Mpc)^{-3}$ is dominated by top level voids, while for the sample with $2(h^{-1}Mpc)^{-3}$ tracer density the contribution from subvoids

is much more important. The hard sphere model that we consider is no longer effective when subvoids are included. Thus we need to discard them in that case.
The modelling of the abundance of voids is fundamental to derive the peak background split bias parameters for voids that we compare with the ones we measure numerically from the cross power spectrum and the auto power spectrum of voids individually. All the expressions we use for the mass and size description of voids are the results found by Sheth and van de Weygaert already described in the second section of chapter 3.
In observational data it is more relevant the void size distribution than the void mass function. We use the spherical collapse model to convert the mass in Lagrangian space to size in Eulerian space, taking the non linear density contrast of voids equal to -0.8. Therefore, we can estimate the Lagrangian size of voids R_L as $R_L = 0.58R$. Assuming, as Sheth and van de Weygaert, that voids conserve their number density evolving from Lagrangian to Eulerian space, we have that the void size distribution in Eulerian space is

$$\frac{dn}{d\ln R} = \frac{dn}{d\ln R_L} = 3\frac{dn}{d\ln M}. \tag{5.12}$$

We have already seen in the third and fourth section of the chapter 3 that there are some extensions of the Sheth and van de Weygaert model which describe better the void in cloud problem, as, for example, the volume conserving model that considers large voids formed by merging of the smaller ones, obtaining

$$\frac{dn}{d\ln R} = \frac{V_L}{V}\frac{dn_L}{d\ln R_L}, \tag{5.13}$$

where the Lagrangian quantities are denoted by the subscript L, while the Eulerian ones are without subscript. Even if this prescription results in much better agreement with simulation data about the abundance of voids than the original Sheth and van de Weygaert model we are mostly interested in larger voids, therefore, we can use the results of the simpler Sheth and van de Weygaert model because it is the best fit to our simulation data about bias parameters relative to cross and auto power spectrum of voids.
Since the void abundance depends on the void definition, we shall treat δ_v as a free parameter to model the data. On the other hand, since we are interested mainly in larger voids, the effect of δ_c is negligible, unless its value is much smaller than the spherical collapse threshold of 1.68. Thus we fix $\delta_c = 1.68$ throughout. In this regime, that is of larger voids, even the one barrier distribution, approximation for $\delta_c \gg \mid \delta_v \mid$ of the equation 3.6, can be used as first crossing distribution:

$$\nu f(\nu) \approx \sqrt{\frac{\nu}{2\pi}}\exp\left(-\frac{\nu}{2}\right). \tag{5.14}$$

The void size distributions measured in our N-body simulations at $z = 1, 0.5, 0$ are plotted in figure 5.1. The data points from the L1500 and L250 simulations agree reasonably well in general, especially on the abundance of small voids for a sampling density of $0.02(h^{-1}Mpc)^{-3}$. However, the L1500 catalogs are dominated by top level voids and

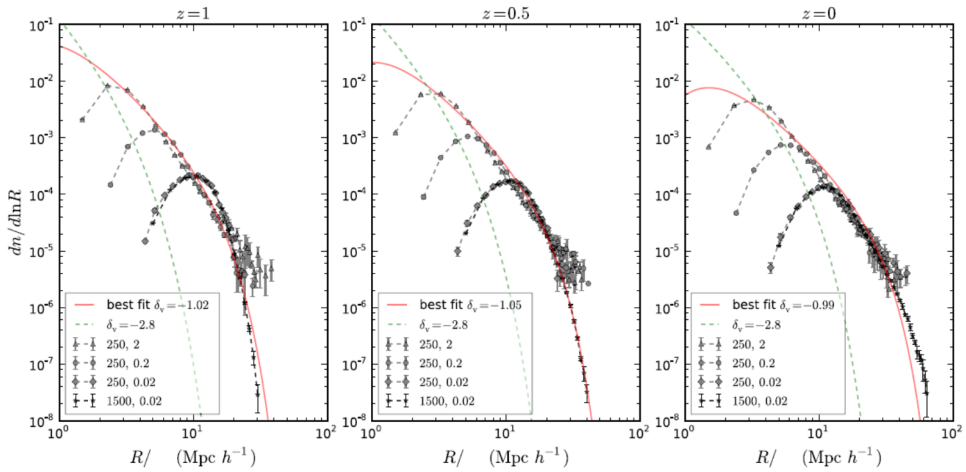

Figure 5.1: The void size distribution at $z = 1, 0.5, 0$ (from left to right) measured in simulations. The data are obtained from L1500 with sampling density $0.02(h^{-1}Mpc)^{-3}$ (star black), and L250 with sampling densities $2, 0.2, 0.02(h^{-1}Mpc)^{-3}$, respectively (triangle, gray), (circle, gray), (diamond, gray). The SvdW void size distribution is shown with $\delta_v = -2.8$ (dashed line, green) and with the best fit value δ_v (solid line, red) as stated in the inset of each panel.[8]

thus sample the large voids much better and extend the void size distribution to larger void radii, while the L250 catalogs with increased sampling density are more influenced by subvoids and sample better voids with small radii.
In figure 5.1 we also plot the Sheth and van de Weygaert prediction with $\delta_v = -2.8$ together with the best fit void size distribution obtained varying freely δ_v. Since the higher sampling densities are more influenced by subvoids, that are not included in the Sheth and van de Weygaert model, we only fit the L1500 data with $R > 20(h^{-1}Mpc)$. We fit the data separately at each redshift and obtain best fit values of $\delta_v = -1.02, -1.05, -0.99$ for $z = 1, 0.5, 0$, respectively. These values are consistent with each other, but are quite different from the canonical spherical collapse value of $\delta_v = -2.8$, that is the value of the threshold when we consider shell crossing as the condition of construction of voids, because the algorithm we use in our simulations defines voids of arbitrary geometries and density profiles in various tracer sampling densities, that not always agree with the shell crossing estimate.
Now we use the void mass function and the void size distribution found by adopting the excursion set formalism in the Sheth and van de Weygaert model to derive the peak background split bias parameters for voids. Supposing there is a long wavelength perturbation δ_L in the Lagrangian space that shifts the thresholds δ_v and δ_c as $\delta_v \to \delta_v - \delta_L$ and $\delta_c \to \delta_c - \delta_L$ the bias parameters in Eulerian space are

$$b_i = \frac{1}{n_0}\frac{\partial^i}{\partial\delta^i}[(1+\delta)n(\delta_L)]|_{\delta=0}, \tag{5.15}$$

where δ denotes the corresponding perturbation in Eulerian space and n_0 and $n(\delta_L)$ represent the mass function with zero and δ_L background perturbation, respectively. The factor $1+\delta$ maps the mass function from Lagrangian space to Eulerian space and we use spherical collapse to relate the Lagrangian and Eulerian background perturbations:

$$\delta_L = \delta - \nu_2\delta^2 + \nu_3\delta^3 + \dots, \tag{5.16}$$

with $\nu_2 = 12/21$ and $\nu_3 = 341/567$. Therefore, we can explicit the expression for the first bias parameter:

$$b_1 = 1 + \frac{\nu - 1}{\mid \delta_v \mid} - \frac{(\delta_v/\delta_c)^2}{4\nu(\delta_c + \mid \delta_v \mid)}. \tag{5.17}$$

Given these tools we measure first the cross power spectrum between voids and dark matter and then the void auto power spectrum in our simulations to extract the large scale bias parameters from them and to compare them with the theoretical ones in order to have the main elements to understand some characteristics of the clustering properties of voids. In fact, studying the power spectrum is almost the same as analyzing the two-point correlation function, because the first one is essentially the Fourier transform of the second, with the advantage that it is easier to study, therefore the conclusions about the physical features of the system are the same.
Defined the void density contrast δ_{void} as

$$\delta_{void} \equiv \frac{n_v - \bar{n}_v}{\bar{n}_v}, \tag{5.18}$$

where n_v and $\bar{n}_v$ are the number density and the mean number density of voids, respectively, the cross power spectrum P_c between voids and dark matter is defined as

$$< \delta_{void}(\boldsymbol{k_1})\delta_{dm}(\boldsymbol{k_2}) >= P_c(k_1)\delta_D(\boldsymbol{k_1} + \boldsymbol{k_2}), \tag{5.19}$$

where δ_{dm} is the dark matter density contrast and δ_D the Dirac delta function. Using the cross power spectrum between voids and dark matter and the dark matter auto power spectrum P_{dm} we define the cross bias parameter

$$b_c \equiv \frac{P_c}{P_{dm}}. \tag{5.20}$$

We calculate this parameter from simulations and we plot the numerical results of the cross bias parameter depending on the wave number k at redshift $z = 0$ in figure 5.2. The behaviour of b_c at the different scales for each void of different size gives us information about the relations between voids and dark matter. The cross bias of the larger voids with radius $R \geq 20(h^{-1}Mpc)$ is less influenced by variations in the sampling density, instead for void radii $R < 20(h^{-1}Mpc)$ the behaviour of b_c changes with the sampling density at low k. Overall, b_c converges at zero at high $k(k \geq 10/R)$ whether for big voids or small ones, whereas it exhibits oscillations on intermediate scales. At low $k(k \leq 1/R)$ b_c is roughly constant, showing biased values for small voids, almost unbiased values for voids with radius about $17(h^{-1}Mpc)$ and antibiased ones for big voids.

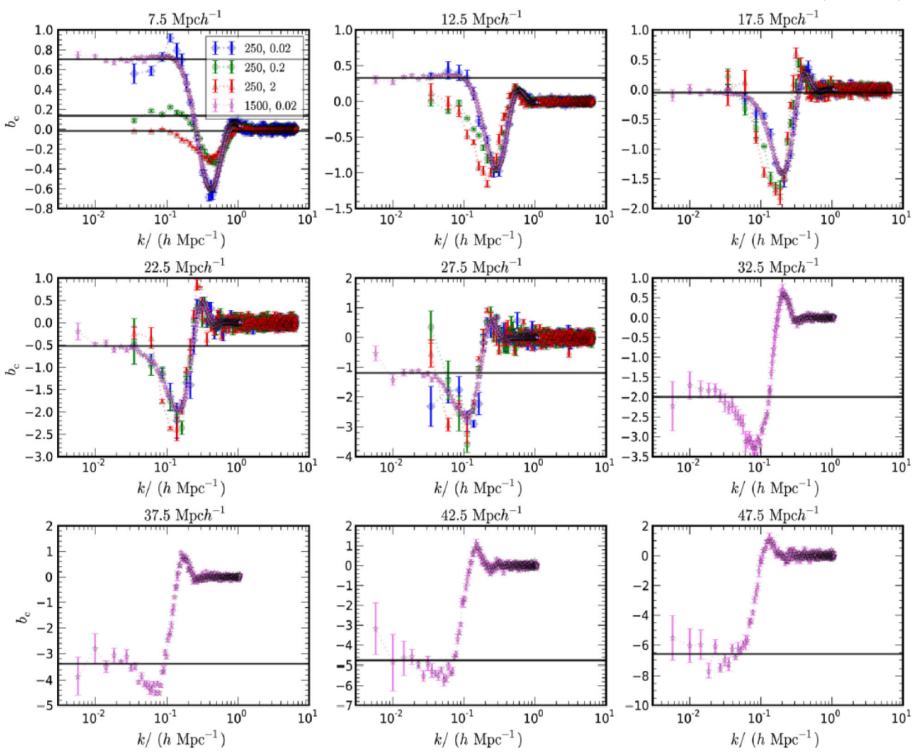

Figure 5.2: Void matter cross bias b_c as a function of wave number k for various void sizes R at $z = 0$. The results are shown for L1500 with sampling density $0.02(h^{-1}Mpc)^{-3}$ (star, purple), L250 with sampling density $0.02, 0.2, 2(h^{-1}Mpc)^{-3}$, respectively (diamond, blue), (circle, green), (triangle, red). Horizontal lines show best fits to the linear large scale regime of b_c whenever the fit is feasible (solid line, black).[8]

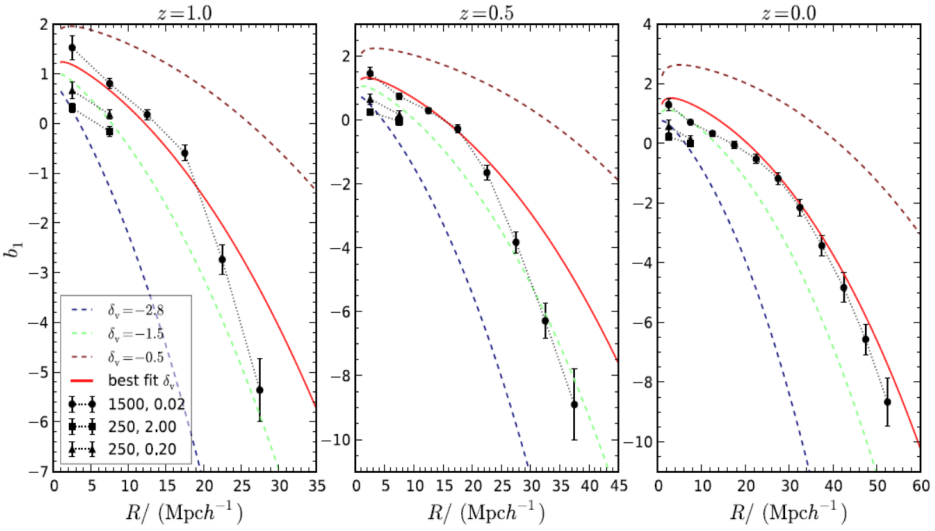

Figure 5.3: The large scale best fit to the void matter cross bias b_c as a function of void radius R from L1500 (circles), L250 with sampling density 2 (square) and $0.2(h^{-1}Mpc)^{-3}$ (triangle) at redshifts $z = 1, 0.5, 0$ (from left to right). The curves show b_1 computed from equation 5.17 for various values of δ_v, in particular the best fit δ_v from the void size distribution (solid line, red).[8]

To better analyze the dependence of bias of voids on their radius only at large scales we plot in figure 5.3 the best fit large scale value of b_c, as a function of void radii R for $z = 1, 0.5, 0$, taken from the results of the numerical simulations corresponding to the different sampling density, comparing them with the prediction of peak background split bias of equation 5.17, assuming different values of δ_v, also plotted in figure 5.3. We notice that $\delta_v = -2.8$ underestimates the data, whereas a better agreement is achieved when is used the best fit δ_v derived from the void size distribution, but only for voids with $R > 20(h^{-1}Mpc)$ at $z = 0$, because the agreement slightly deteriorates at higher redshifts. The contribution of subvoids is not negligible in the higher sampling densities, therefore the bias results from L250 with higher sampling densities are below the PBS prediction. The oscillations in b_c are related to the structure of the void density profile, that describes the distribution of matter conditioned on having a void center at $r = 0$. It can be shown that it is the same as the cross correlation function $\xi_c(r)$ between void centers and dark matter particles, in fact we have

$$\rho_v(r) = \bar{\rho}_m[1 + \xi_c(r)], \tag{5.21}$$

$$\Delta_v(r) \equiv \frac{\rho_v(r)}{\bar{\rho}_m} - 1 = \xi_c(r). \tag{5.22}$$

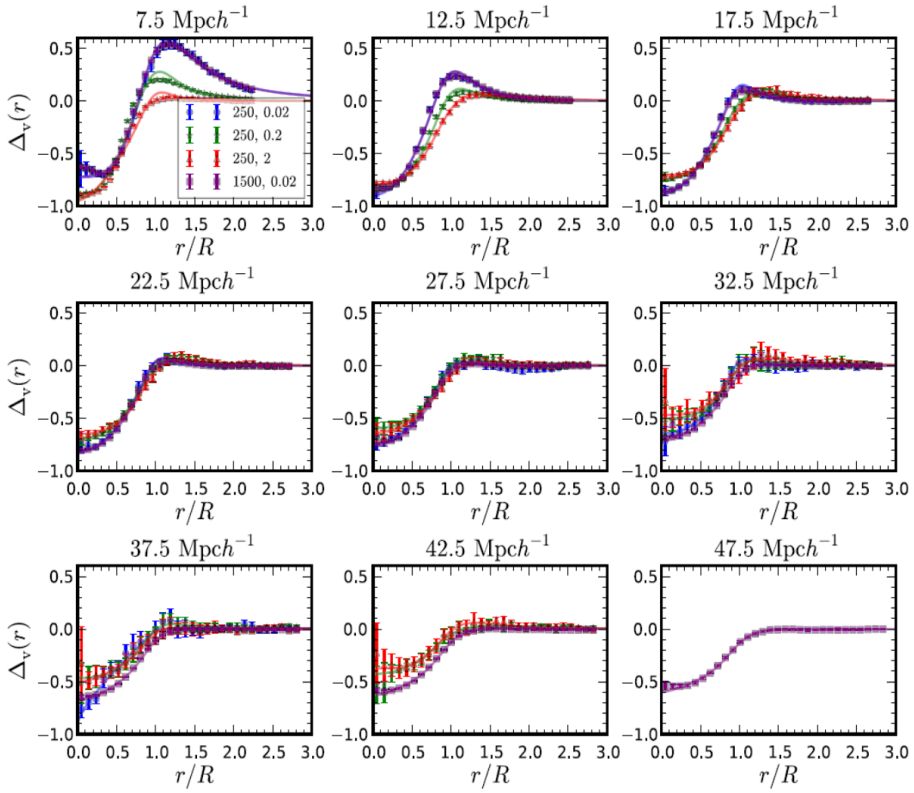

Figure 5.4: Void density profiles for different void radii at $z = 0$ from L1500 with sampling density $0.02(h^{-1}Mpc)^{-3}$ (square, violet) and L250 with sampling densities 0.02 (circle, blue), 0.2 (star, green), $2(h^{-1}Mpc)^{-3}$ (triangle, red). Solid lines show the best fits using equation 5.23 (same colors).[8]

We investigate the numerical results of our simulations about the void density profile fitting them with an accurate formula for $\Delta_v(r)$, that is:

$$\Delta_v(r) = \delta_{cen}\frac{1-(\frac{r}{r_s})^{\alpha}}{1+(\frac{r}{R})^{\beta}}, \tag{5.23}$$

where, since we are probing a wide range of sampling densities and redshifts, we will allow the central density fluctuation δ_{cen}, the scale radius r_s, namely the radius at which $\Delta_v(r)$ vanishes, the two parameters α, β to vary freely when we fit them to our simulation data.
Through our simulations we also measure the spherically averaged void density profile. In figure 5.4 we plot the numerical results noticing good agreement between L1500 and L250 with identical sampling density of $0.02(h^{-1}Mpc)^{-3}$ and the fits using equation 5.23 are a good description of the simulation data. Moreover, voids from higher sampling densities exhibit lower ridges at fixed void radius. However, although the equivalence between $\Delta_v(r)$ and $\xi_c(r)$ is mathematically exact, the empirical formula in equation 5.23

is only accurate out to a few times the void radius R and does not include the large scale correlation regime. In fact, even the Fourier transform of $\Delta_v(r)$ is formally equal to the void matter cross power spectrum P_c resulting from our simulations:

$$P_c(k) = \int \frac{4\pi r^2 dr}{(2\pi)^3} \frac{\sin(kr)}{kr} \Delta_v(r), \tag{5.24}$$

this equation is accurate only in the high k regime. To make easier the comparison with our simulation results, we furnish an estimate for the void matter cross bias $b_c(k)$ normalizing $\Delta_v(k)$ with respect to the nonlinear dark matter power spectrum. The results for b_c are qualitatively similar to those displayed in figure 5.2, but not completely, since in the case of the Fourier transform we have a scale dependence at low k which is particularly large for voids with high ridges. This strong scale dependence as $k \to 0$ originates from the fact that equation 5.23 is not accurate in the analysis of large scale correlations. Usually, in order to predict the cross power spectrum to low values of k using the void profile fitted by equation 5.23 we extrapolate a small scale quantity to large scales. To explicit the differences between the $z = 0$ void cross bias b_c from simulations and the one obtained from Fourier transform of the best fit void density profile shown in figure 5.4 we compare in figure 5.5 the last one with the void matter cross bias b_c from the L1500 simulation. We notice that the agreement is only qualitative, especially at low k, where the profile from equation 5.23 often causes a strong scale dependence, even if it improves for larger voids, because we measure their density profiles out to larger distances.
This discussion encourages to express the void matter cross power spectrum P_c considering a scale dependence at very low k. Therefore, we can split equation 5.24 into two contributions:

$$P_c(k) = \int_0^{r_*} \frac{4\pi r^2 dr}{(2\pi)^3} \frac{\sin(kr)}{kr} [\Delta_v(r) - b_1 \xi_{dm}(r)] + b_1 \int_0^{\infty} \frac{4\pi r^2 dr}{(2\pi)^3} \frac{\sin(kr)}{kr} \xi_{dm}(r), \tag{5.25}$$

where the scale r_*, with magnitude equal to a few void radii, is determined with the simplifying assumption that $\Delta_v(r) = b_1 \xi_{dm}(r)$ for $r > r_*$, that means that at large r the void matter cross correlation function is expected to be proportional to the dark matter correlation function. The crucial point is that if void bias is different from a simple k independent linear contribution the term $\Delta_v(r) - b_1 \xi_{dm}(r)$ does not vanish. Therefore, having at small k $\sin(kr)/kr \sim 1$, in the limit $k \to 0$ we have:

$$P_c(k) = const + b_1 P_{dm}(k) \qquad b_c(k) = \frac{const}{P_{dm}(k)} + b_1. \tag{5.26}$$

The first term in both these two equations generates a residual k dependent bias at very low k, but, as we have just assumed, $\Delta_v(r) = b_1 \xi_{dm}(r)$ for $r > r_*$. Therefore, we can consider any residual scale dependence at low k equal to zero, as confirmed by the fact that in figure 5.2 we ignore the constant term in the expression of $P_c(k)$ in the first of the two equations 5.26. The dependence on the wave number k is generally more significant in the description of the void center, but towards larger distances, especially for bigger voids, the PBS approximation, that is k independent, results most accurate. In fact, if

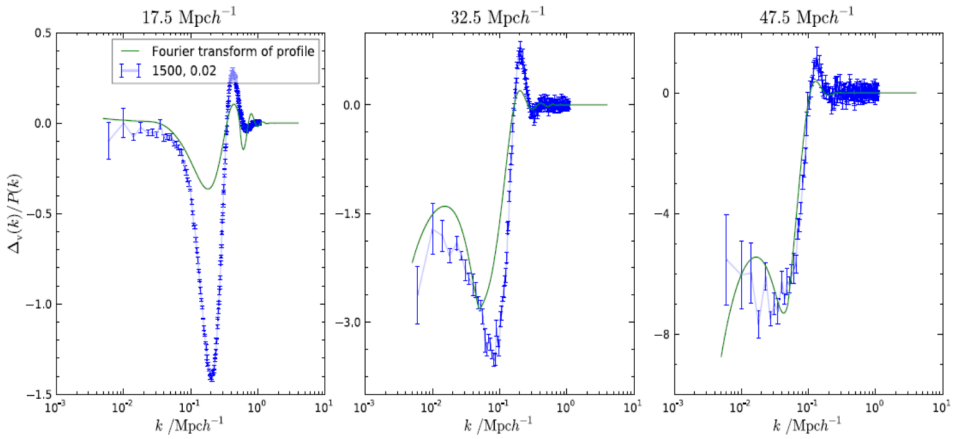

Figure 5.5: Comparison of the void matter cross bias b_c from the L1500 simulation (blue with error bars) with the Fourier transform of the corresponding best fit void density profile derived by the matter power spectrum (green lines) at $z = 0$.[8]

we parametrize $b_c(k)$ including a filter function, since voids trace a smoothed version of the mass density field, while b_c is defined relative to the unsmoothed mass distribution, through the following expression:

$$b_c(k) = (b_0 + b_2k^2 + b_4k^4)\exp\left[-\frac{1}{2}(kR_G)^2\right], \tag{5.27}$$

where b_0, b_2, b_4, R_G are free parameters, we notice that comparing this k dependent bias parametrization with our numerical simulations at different sampling densities we have the best agreement for the high k regime.
To complete our discussion about clustering of voids we investigate the bias parameters of the void auto power spectrum comparing our numerical results to the PBS predictions. The sampling densities dependence of the void auto power spectrum at $z = 0$ is shown in figure 5.6, subtracting the scale independent Poisson shot noise, that affects the void auto power spectrum due to the discrete nature of voids:

$$P_{poi} = \frac{1}{(2\pi)^3\bar{n}_v}. \tag{5.28}$$

The results obtained L1500 and L250 with the same sampling density of $0.02(h^{-1}Mpc)^{-3}$ agree. However, as in the case of the void cross power spectrum, the power spectra from different sampling densities show larger discrepancies in case of small voids, due to their high subvoid fraction, instead, for big voids the differences among various sampling densities reduce. Voids are biased tracers of the mass, allowing us to infer information about the dark matter power spectrum, but they are also biased tracers of the density field. In fact, voids can also be defined in the spatial distribution of galaxies without knowledge of the underlying dark matter density field. Therefore, the void auto power spectrum

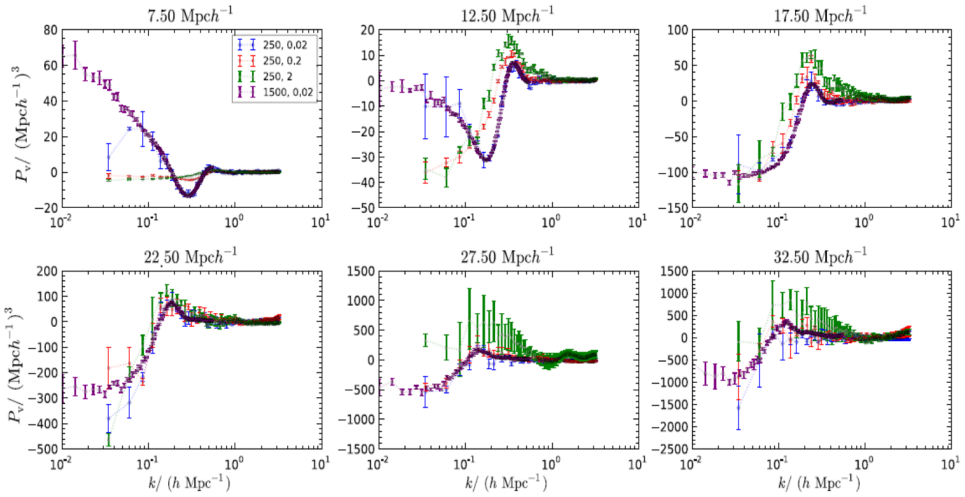

Figure 5.6: Void auto power spectrum from L1500 (square, violet) and L250 with sampling densities 0.02 (diamond, blue), 0.2 (circle, red), $2(h^{-1}Mpc)^{-3}$ (triangle, green) for voids of different size at $z = 0$. Poisson shot noise has been subtracted.[8]

Table 5.1: Different models for P_{excl}.

Model#	P_{excl}
1	$P_{HS}(k;D)$
2	$P_{TH}(k;D)$
3	$P_{tanh}(k;D,\sigma_1)$
4	$W(k;\sigma_G)P_{HS}(k;D)$
5	$W(k;\sigma_G)P_{TH}(k;D)$

is more useful in the analysis of clustering of voids than the void matter cross power spectrum, because it is more closely related to observational data.
On large scales it is sufficient to consider a linear bias model for the void auto power spectrum. However, extending our analysis to $k \geq 0.1(Mpc^{-1}h)$, we need to introduce a renormalized bias approach to include higher order bias up to the 1-loop order. In fact, even at low k linear bias is not so good at describing the behaviour of the void auto power spectrum.
In both the linear and renormalized bias models it is important to introduce in the resulting expressions for the void auto power spectrum a term representing the fact that voids are much more extended than haloes, thus void exclusion plays an important role in modelling the void auto power spectrum. In the table 5.1 are listed five different approximations for the void auto power spectrum exclusion term P_{excl}. The first model is a power spectrum corresponding to a correlation function described by the so called Percus-Yervick equation. The second model is a power spectrum corresponding to a hard sphere

correlation approximated by a top hat window in configuration space. The third one is a soft sphere model that smoothly interpolates the transition region in the correlation function of the second model. Finally, the fourth and the fifth are P_{HS} and P_{TH} multiplied by a Gaussian damping factor of width σ_G, namely $W_G(k) = \exp[-1/2(\sigma_G k)^2]$.
We first analyze the simple linear bias model, in which

$$P_v = P_{1,1} + P_{excl}, \tag{5.29}$$

$$P_{1,1} = b_1^2 P_{dm}, \tag{5.30}$$

where P_{dm} denotes the non linear dark matter power spectrum.
In figure 5.7 we plot the best fits of equation 5.29 to the void auto power spectrum from L1500 for different void radius bins at $z = 0$ and the individual terms from equation 5.29 separately to underline their relative importance. Poisson shot noise has been subtracted from the numerical results, and only data points up to $k = 0.2(Mpc^{-1}h)$ are included in the fit. For the power spectrum exclusion term is used only the model 1, because the others lead to similar results. Also here the fit given by the theoretical model, that in this case corresponds to equation 5.29, is poor for the smallest voids, but improves the description of the power spectra for larger voids. In fact, towards smaller scales linear bias is not sufficient and oscillations from the hard sphere model are too strong. In the low k regime $P_{1,1}$ and P_{excl} are of opposite sign and, since the second one is comparable or even larger than the first one, the exclusion term plays a fundamental role for these big scales. Overall, the model of equation 5.29 fails for the small voids with $R \leq 15(h^{-1}Mpc)$ whose power spectrum, unlike voids of larger radii, increases as k decreases and can become positive. It is unclear which one between the two different terms of biasing and exclusion is the most influent for the failure in the modelling, but certainly we can understand more in the configuration space, where they are disentangled in the correlation function expression, with the effects of the exclusion term confined at short distances while the linear biasing is more important at large r.
In figure 5.8 we plot as a function of the void size R the best fit values for b_1, D, the diameter of the hard spheres, and the χ^2 per degree of freedom for our five models, that yield similar results. In the most left panel relative to b_1 we plot also the b_1 from the PBS formalism, namely equation 5.17 with the best fit δ_v from the void size distribution, and the large scale b_c measurements. We notice that the agreement is much better between the b_c measurements and the PBS predictions than between these last ones and the data. Instead, the χ^2 per degree of freedom shows that the linear bias approximation is not so accurate in the fitting of numerical results about the void auto power spectrum, especially at small R. It is evident that the hard sphere exclusion model considering a linear bias demonstrates to have problems in the description of the void auto power spectrum for small voids. Therefore, now include higher orders of bias parameters in order to study the system through a renormalized bias model.
We know that voids are biased tracers of the density field, so to pass to the renormalized bias model we include contributions up to the third order for the value δ_v from the void size distribution:

$$\delta_v = b_1\delta + \frac{b_2}{2}\delta^2 + \frac{b_3}{6}\delta^3, \tag{5.31}$$

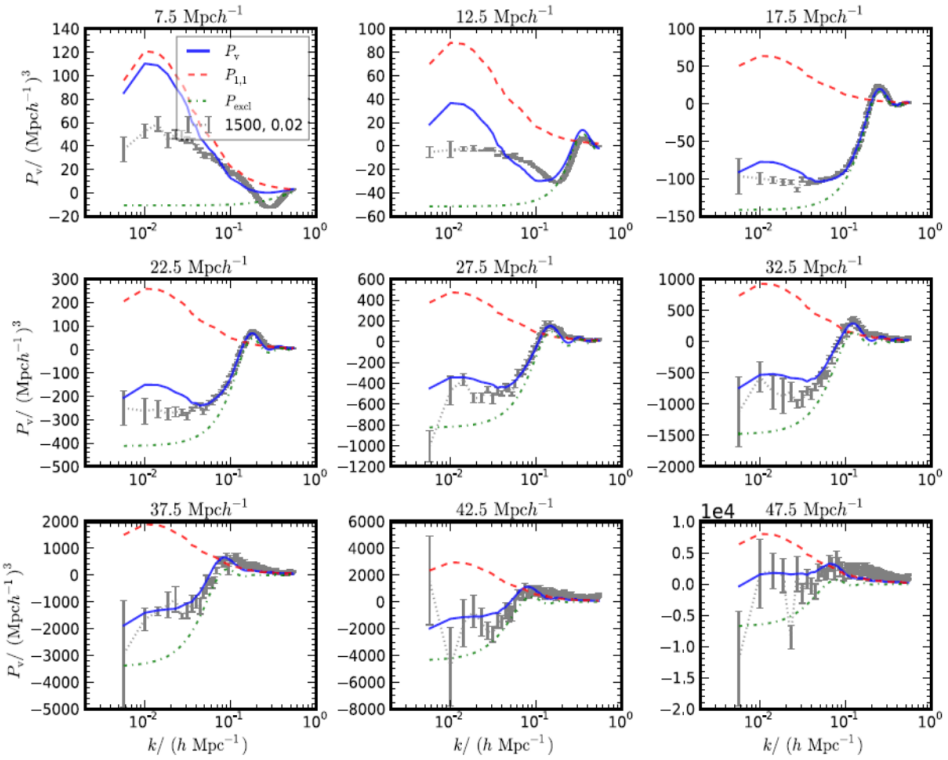

Figure 5.7: Best fits from equation 5.29 (solid line, blue) to the void auto power spectrum from L1500 at $z = 0$ (gray data points with error bars). The individual components of the best fit are also shown: the linear bias term $P_{1,1}$ (dashed line, red) and the void exclusion term P_{excl} (dotted dashed line, green). Here, model 1 is used for P_{excl} and Poisson shot noise has been subtracted from the data.[8]

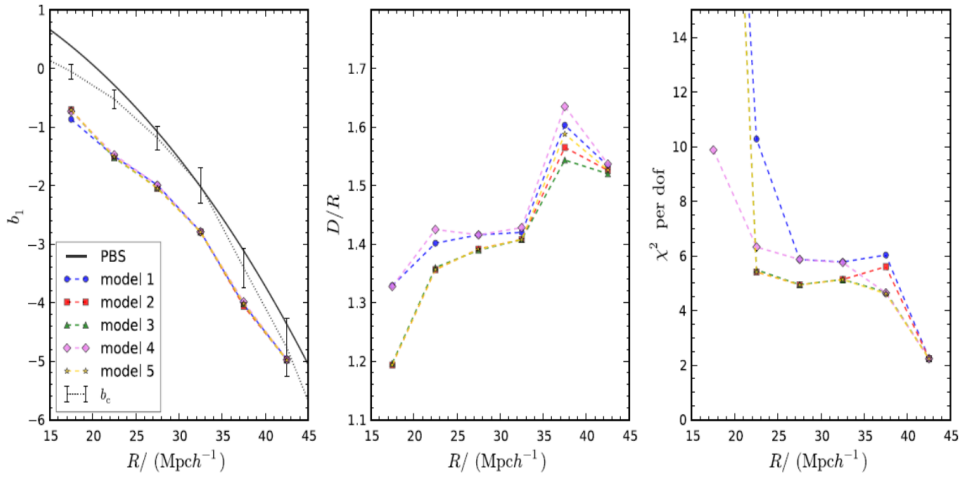

Figure 5.8: The best fit values for b_1 and D using equation 5.29 to fit the void auto power spectrum and the corresponding χ^2 per degree of freedom. The results from model 1 (circle, blue), 2 (square, red), 3 (triangle, green), 4 (diamond, violet), 5(star, yellow) are shown. The solid line depicts b_1 as predicted from the PBS formalism. We also show the large scale measurement of b_c (dotted line, black) for comparison.[8]

where b_2 is the second order bias parameter, not like the term b_2 in equation 5.27, that it is only a coefficient of k^2, and b_3 is the third order bias parameter. In the standard 1-loop expansion of the void auto power spectrum with local bias four new terms appear, but, since two of these terms are proportional to $P_{dm}(k)$, we renormalize b_1 including them in $P_{1,1}$. Hence, there are effectively only two new terms in the expression for the void auto power spectrum considering a renormalized bias model:

$$P_v = P_{1,1} + P_{2,11} + P_{11,11} + P_{excl}, \tag{5.32}$$

where the two new terms $P_{2,11}$ and $P_{11,11}$ depend only on the parameters b_1 and b_2, because b_3 only appears in one of the terms proportional to $P_{dm}(k)$ included in $P_{1,1}$.
We repeat the same plot of figure 5.7, but, considering in this case a renormalized bias model, the resulting best fits are obtained after the inclusion of the b_2 terms and, mostly, the new figure 5.9 shows the behaviour of the two new terms $P_{2,11}$ and $P_{11,11}$ separately to highlight how the inclusion of these two quadratic bias terms in the determination of the void auto power spectrum improves significantly the agreement with the numerical data. If we analyze deeper the individual components of the best fit power spectrum we notice that the term $P_{2,11}$ is negligible for the entire range of k shown. Instead, the other new term $P_{11,11}$, which is nearly constant at low k, is comparable to, or even larger than $P_{1,1}$ for the biggest voids. Therefore, it is this term that improves the fit to the void auto power spectrum respect to the previous one considering a linear bias model, even if also following a renormalized bias model the fits are still inaccurate for the smallest voids ($R = 7.5(h^{-1}Mpc), R = 12.5(h^{-1}Mpc)$).

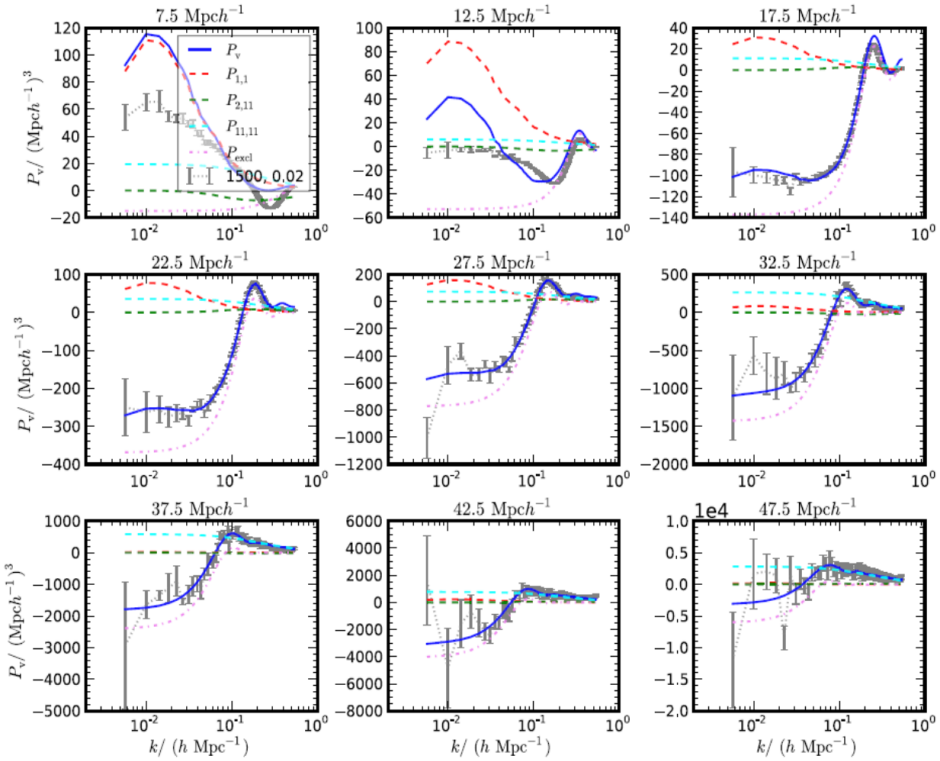

Figure 5.9: Best fits from equation 5.32 (solid line, blue) to the void auto power spectrum from L1500 at $z = 0$ (gray data points with error bars). The individual components of the best fit are also shown: the linear bias term $P_{1,1}$ (dashed line, red), $P_{2,11}$ (dashed line, green), $P_{11,11}$ (dashed line, cyan), and the void exclusion term P_{excl} (dotted dashed line, violet). Here, model 4 is used for P_{excl} and Poisson shot noise has been subtracted from the data.[8]

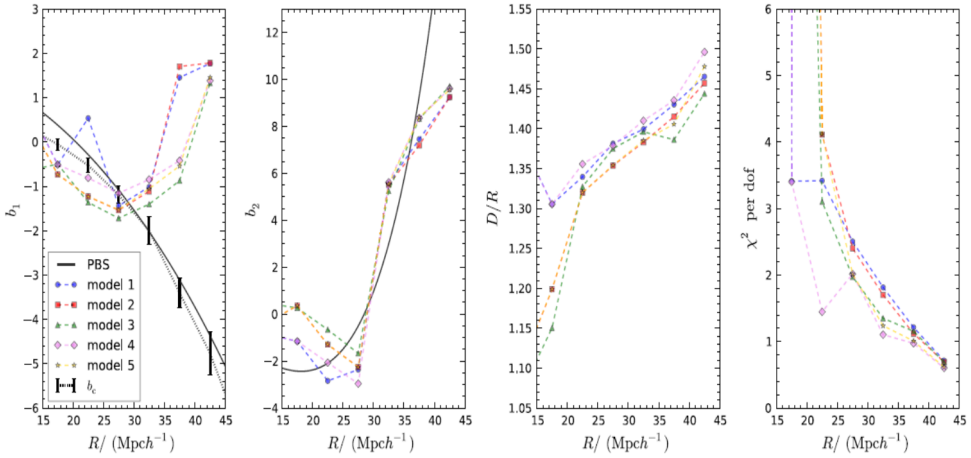

Figure 5.10: The best fit values for b_1, b_2 and D using equation 5.32 to fit the void auto power spectrum and the corresponding χ^2 per degree of freedom. The results from model 1 (circle, blue), 2 (square, red), 3 (triangle, green), 4 (diamond, violet), 5(star, yellow) are shown. The solid line depicts b_1 and b_2 as predicted from the PBS formalism. We also show the large scale measurement of b_c (dotted line, black) for comparison.[8]

We plot in figure 5.10 as a function of the void size R the best fit values for b_1, D, and the χ^2 per degree of freedom for our five models, that yield similar results, as we have already made for the linear bias model in figure 5.8, but here we plot also an other panel relative to the parameter b_2. For b_1 and b_2 we plot the corresponding PBS prediction for comparison. The best fit for b_1 agrees with the PBS prediction, especially for models 1 and 4, for $R \leq 30(h^{-1}Mpc)$, while, at larger void sizes, the best fit b_1 and the PBS prediction begin to differentiate from each other, because the first one turns over and starts to increase, since it must compensate the important contribution of $P_{1,1}$ to the fit, instead, the second keeps on decreasing. On the other hand, between the best fit b_2 and the PBS prediction there is a better agreement, despite the first one is smaller in magnitude at $R \geq 40(h^{-1}Mpc)$. We observe that the coupling term $P_{2,11}$ is negligible, while $P_{1,1}$ and $P_{11,11}$ are the dominant ones, dependent on the quadratic bias parameters b_1^2 and b_2^2, respectively. For D, the best fit values are similar to those in figure 5.8. Looking at the χ^2 per degree of freedom behaviour, we notice that the inclusion of b_2 allows to achieve a big improvement in the fitting of the numerical data respect to the modelling in the linear bias model, since it reaches a lower value respect to the χ^2 per degree of freedom in figure 5.8, namely ~ 1. Whether in figure 5.8 or in figure 5.10 the best fit b_1 is systematically slightly lower than the PBS and large scale fit of b_c measurements, but in the case of the renormalized bias model only for $R \leq 30(h^{-1}Mpc)$, because for larger R the best fit b_1 is higher than its theoretical predictions. If the model is self consistent we would expect that the best fit b_1 from the void auto power spectrum agrees with the large scale b_c measurement. Unfortunately, the construction of voids is affected by systematics, like the exclusion effect modeled by the hard sphere model, that devi-

ate voids from being simple biased tracers of the underlying dark matter density field. Cross correlating dark matter with the void density field tries to eliminate some of the systematics, but the remaining ones influence the system to have inaccurate bias fitting. Hence, we can explain why the b_c measurements agree with the PBS results much better than those from the void auto power spectrum.

In conclusion it is evident that voids are very sensitive to the identification procedure. So are their abundance and biasing. On the other hand, for haloes the agreement between theory and numerical results is generally more encouraging than what we find for voids. Moreover, in galaxy clustering analyses, bias parameters are commonly treated as nuisance parameters. Therefore, while theory is important for understanding how voids evolve in time, in practice it is enough if we find some phenomenological description of void clustering where bias factors are marginalized over, so long as there are not too many free parameters. Finally, even if it seems that one can never predict accurately the abundance and clustering of the surveyed voids from first principles, and consequently that voids cannot be useful to precisely obtain the information we need for the description of cosmological features of the universe, our finding that the void auto power spectrum can be well described by a combination of exclusion and biasing terms, directly observable in galaxy surveys, and our considerations about void density profile and void cross power spectrum and its bias respect to the dark matter power spectrum enable us to deduce important information about the dark matter and cosmological structures distribution in the universe, even without the knowledge of the precise values of the bias parameters.

Conclusions

In this work of thesis we tried to study clustering of voids determining an analytical model of the two-point correlation function of voids in the large scale distribution of galaxies, knowing that a clear picture of the relation between void galaxies and their surroundings is only just becoming available, thanks to the large scale surveys, that now probe a sufficiently large cosmological volume containing a statistically significant number of large voids. Therefore we have to identify and study voids and void galaxies implementing our algorithms numerically, because at the moment analytic formation processes for the evolution of voids are only approximately plausible.
First of all we considered which kind of voids was useful for the evolution and clustering situations we developed during the thesis, choosing to examine in more detail the uncompensated void with a collisionless gas solution, because the approximate volume filling domains observed for most of the range of cosmological structure formation scenarios are constituted by primordial underdensities that developed in these kind of voids through shell crossing, that happens only in case of a collisionless gas solution. The evolution of these kind of voids, analyzed as isolated negative density perturbations, even if we know the importance that the interaction of voids with their surroundings has in order to understand the hierarchical picture scenario, was studied considering that voids evolve towards a spherical form during the expansion, with a distribution of matter characterized by a reverse top-hat profile. These characteristics, together with shell crossing, determine the evolution from underdensity regions to voids. Then, as we did in the last section of chapter 1, it is fundamental to calculate the thresholds establishing which density values correspond to the turnaround or the collapse of an halo and the formation of a void. These results and the theory of excursion set approach with absorbing barrier used to study the dark matter halo abundance, developed in the second chapter, are the basis for the conclusions about void sociology and hierarchy that we described in chapter 3. Exploiting a recent theory by Sheth and van de Weygaert, which studied the behaviour of Brownian random walks in void size space, knowing that for voids we had to face with the complications arising respect to the case of haloes, we introduced a second barrier, extending the excursion set approach to underdense regions, and we analyzed the phenomenon of the void in cloud, that represents the influence of haloes collapse on voids evolution. The improvement of the void in cloud process by Sheth and van de Weygaert was the argument of the third and fourth section in chapter 3, requiring that the volume fraction and shape of the abundance function is fixed during the expansion, rather than

assuming that the expansion of isolated voids preserves their total number density, or following a better model of the void in cloud problem passing to an Eulerian treatment of the spherical evolution model of voids.
The chapter 4 opened the discussion about clustering properties of dark matter haloes with a formalism that is based to the absorbing barrier problem, easily extendable in this case to multiple space correlated random walks, thereby, providing the correct framework to predict a plausible analytical void two-point correlation function, that we tried to express in the first section of chapter 5. In the final section of chapter 4 we presented a deeper analysis of the halo correlation function considering also Eulerian correlation functions of objects with combination of same or different mass and of same or different collapse redshift.
Finally, in chapter 5, given all the useful tools in the precedent chapters, we developed two different methods trying to express analytically the two-point correlation function of voids. For the first one we inspired to the conclusions about haloes in chapter 4, and for the second we modelled the void power spectrum, comparing analytical fits with the results of the numerical simulations, knowing that the distribution of voids is biased relative to the underlying dark matter distribution and that voids are very much spatially extended.
Nowadays it is still complicated to relate voids identified in the galaxy distribution with the underlying void distribution in the dark matter. Therefore, we cannot access to important cosmological information. However, it will be crucial to continue studying large scale structures of our universe through galaxy redshift surveys. This is the way which deserves further analysis in the future, with the challenge to be able to build valid theoretical models, that will have the possibility to furnish comparable predictions with a quantity of numerical data more and more rich and precise with the improvement of the next experiments.

Bibliography

[1] E. Bertschinger, *The self-similar evolution of holes in an Einstein-de Sitter Universe*, The Astrophysical Journal Supplement Series, 58, 1-37 (1985)

[2] J. R. Bond, S. Cole, G. Efstathiou, N. Kaiser, *Excursion set mass function for hierarchical Gaussian fluctuations*, The Astrophysical Journal, 379, 440-460 (1991)

[3] R. K. Sheth, R. van de Weygaert, *A hierarchy of voids: much ado about nothing*, Mon. Not. R. Astron. Soc. 350, 517-538 (2004)

[4] A. Paranjape, T. Y. Lam, R. K. Sheth, *A hierarchy of voids: more ado about nothing*, Mon. Not. R. Astron. Soc. 420, 1648-1655 (2012)

[5] C. Porciani, S. Matarrese, F. Lucchin, P. Catelan, *Excursion set approach to the clustering of dark matter haloes in Lagrangian space*, The Astrophysical Journal, Mon. Not. R. Astron. Soc. 298, 1097-1112 (1998)

[6] E. Scannapieco, R. J. Tacker, *Toward an improved analytical description of Lagrangian bias*, The Astrophysical Journal, 619, 1-11 (2005)

[7] E. Jennings, Yin Li, W. Hu, *The abundance of voids and the excursion set formalism*, Mon. Not. R. Astron. Soc. 434, 2167-2181 (2013)

[8] K. C. Chan, N. Hamaus, V. Desjacques, *Large-scale clustering of cosmic voids*, Phys. Rev. D 90, 103521 (2014), arXiv: 1409.3849 [astro-ph.CO]

[9] K. C. Chan, N. Hamaus, M. Biagetti, *The Constraint of Void Bias on Primordial non-Gaussianity*, arXiv: 1812.04024v1 [astro-ph.CO] (December 12, 2018)

[10] J. M. Bardeen, J. R. Bond, N. Kaiser, A. S. Szalay 1986, ApJ, 304, 15

[11] J. R. Bond 1987, in Cosmology and Particle Physics, ed. I. Hinchcliffe (Singapore: World Scientific), 22

[12] J. R. Bond 1988a, in the Early Universe, ed. W. G. Unruh (Dordrecht: Reidel), 283

[13] J. R. Bond 1988b, in Proc. Vatican Workshop on Large Scale Streaming Motions, ed. V. C. Rubin & G. V. Coyne (Princeton: Princeton University Press), 419

[14] J. R. Bond 1989, in Frontiers of Physics-From Colliders to Cosmology, ed. B. Campbell & F. Khanna (Singapore: World Scientific), 182

[15] J. R. Bond, A. S. Szalay, J. Silk 1988, ApJ, 324, 627

[16] S. Cole 1991, ApJ, 367, 45

[17] S. Cole, N. Kaiser 1988, MNRAS, 233, 637

[18] S. Cole, N. Kaiser 1989, MNRAS, 237, 1127

[19] G. Efstathiou, M. Rees 1988, MNRAS, 230, 5p

[20] R. Epstein 1983, MNRAS, 205, 207

[21] W. H. Press, P. Schechter 1974, ApJ, 187, 425

[22] A. J. Benson, F. Hoyle, F. Torres, M. S. Vogeley, 2003, MNRAS, 340, 160

[23] J. A. Fillmore, P. Goldreich, 1984, ApJ, 281, 9

[24] S. R. Furlanetto, T. Piran, 2006, MNRAS, 366, 467

[25] A. Cooray, R. K. Sheth, 2002, Phys. Rep., 372, 1

[26] J. Dubinski, L. N. Da Costa, D. S. Goldwirth, M. Lecar, T. Piran, 1993, ApJ, 410, 458

[27] P. Catelan, F. Lucchin, S. Matarrese, C. Porciani, 1998, MNRAS, 297, 692

[28] H. J. Mo, S. D. M. White, 1996, MNRAS, 282, 347

[29] R. K. Sheth, G. Tormen, 1999, MNRAS, 308, 119

[30] C. Lacey, S. Cole, 1993, MNRAS, 262, 627

[31] C. Lacey, S. Cole, 1994, MNRAS, 271, 676

[32] E. Scannapieco, R. Barkana, 2002, ApJ, 571, 585

[33] R. K. Sheth, H. J. Mo, G. Tormen, 2001, MNRAS, 323, 1

[34] A. Heavens, J. A. Peacock, 1988, MNRAS, 232, 339

[35] A. Heavens, J. A. Peacock, 1990, MNRAS, 243, 133