UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Astrophysics & Cosmology

Testing Void Linearity Using The Zel'Dovich Approximation

Thesis supervisors: Candidate:

Prof: Jens Jasche   **Kai Aidan Gowcoot**

Prof: Nicola Bartolo

Thesis co-supervisors:

Stephen Stopyra

Abstract

In this study, we investigate voids, which are regions of low density within the cosmic web, which provide valuable insights into cosmological evolution and can help constrain cosmological parameters. To identify voids, we utilize simulations with both normal and inverted initial conditions and employ the Zel'dovich approximation, a first-order Lagrangian perturbation theory that simplifies the evolution of density fluctuations by assuming matter moves along straight lines. The initial conditions are generated by the `GenetIC` (Stopyra et al., 2021b) algorithm and the simulations are evolved using `GADGET2` (Springel, 2005) with dark matter only. By mapping the halos from the inverted simulation onto the normal simulation, we effectively detect voids in the normal simulation. Our analysis reveals a high degree of correspondence between Zel'dovich and full N-body voids and halos, especially when comparing power spectra and autocorrelation functions, suggesting a promising method for void cataloguing. However, some discrepancies are observed which raise questions about the suitability of the Zel'dovich model for accurately describing void evolution and, notably, testing void linearity.

Contents

Chapter 1

Introduction

Voids are regions of space with low matter density relative to the rest of the universe. They are of interest to researchers as they might offer insights into the Universe's formation through the constraining of cosmological parameters.

These under-dense regions of the universe form a pattern called the "cosmic web". We use cosmological simulations to simulate this web accurately by adjusting cosmological parameters and introducing equations of motion for particles, which form the cosmic web.

To achieve high simulation accuracy, we must increase the number of particles and use sophisticated equations. Increasing these comes at a cost of time and CPU power.

In this thesis, I introduce the concept of simulation pairs, which involves comparing the halos formed from inverted initial conditions to find voids. Moreover, I quantify the reliability of the Zel'dovich approximation, a first-order Lagrangian perturbation theory, in predicting the linear evolution of voids within the cosmic web.

The overarching objective is to assess the validity of the Zel'dovich approximation in being able to reproduce the linear evolution of voids within the cosmic web accurately. This approach may first enable us to improve our cosmological simulation efforts, as well as refine our understanding of essential cosmological parameters and provide deeper insights into the origin and development of the universe.

1.1   What are Halos and Voids?

During the very early stages of the universe, a process called inflation occurred. The hypothetical scalar field responsible for this phenomenon is called the inflaton field. Quantum fluctuations arise in the overall density of dark matter from very early on, but because the early universe is comprised of hot, ionized plasma, the overdensities formed by the fluctuations barely affect baryonic matter (Guth, 1981).

These quantum fluctuations, which are particles that pop into existence and annihilate, change the overall distribution of energy and matter in the early universe. This initially only affects dark matter as it is only subject to the gravitational influence that these fluctuations provide and is not slowed down or heated up by interactions with radiation or other particles, allowing it to collapse into denser structures more rapidly. The collapse of these density peaks arises from the primordial density field, which exhibits a density contrast $\delta(x) = (\rho(x) - \bar{\rho}_m)/\bar{\rho}_m$, where $\bar{\rho}_m$ is the mean mass density of the Universe. Peaks with positive $\delta(x)$ serve as the seeds of structure formation as dark matter aggregates and collapses onto itself here (Kravtsov and Borgani, 2012).

We can assume that $\delta(x)$ is a homogeneous and isotropic Gaussian random field since it ties in well with observations. This density contrast quantifies the deviation of the actual density of matter from the average density across the Universe. In the beginning, any fluctuations that do exist will have a δ much less than 1. That is, the relative over (or under) density will be much lower than the average (Peebles, 1980).

As the universe expands, it cools enough for protons and electrons to combine to form Hydrogen atoms - an epoch called recombination. After recombination, the overdensities formed during inflation are now much larger and begin to affect baryonic matter. These serve as seeds for the formation of larger density fields as the Universe evolves, called **halos**. These density perturbations grow and amplify, attracting baryonic matter onto them and leading to the formation of large-scale structures that we observe in the universe today. Between these overdense regions lie underdense regions; complementary regions of space that have less density than the average density of the universe and therefore tend to contain fewer galaxies and galaxy clusters than their halo counterparts, called **voids**.

The equations that describe how these overdensities and underdensities grow are complex, but they can be simplified, at least during the early stages. They can be approximated by adding together the solutions of simpler linear equations, a process called linear perturbation theory (LPT). These equations can predict the statistical properties of the resulting large-scale structure, for example, the distribution of matter and the clustering properties of galaxies. One such example of LPT is the Zel'dovich approximation, which relies only on first-order perturbation theory.

Now, the evolution of these density fluctuations or fields is not exactly linear in reality. It is quite non-linear and involves, as previously mentioned, complex equations that describe processes such as shell crossing and virialization. It is, therefore, useful to consider simplified models to better describe the linear and non-linear regimes of these density fields to simulate them adequately (Peebles, 1980).

1.2 Introduction to Cosmological Simulations

Cosmological simulations are tools used to model the evolution of the universe at different epochs. Typically, they simulate the universe's early stages up to the present day. They include several physical processes by employing physical models such as gravity, structure formation, and gas dynamics.

When studying halos and voids, cosmological simulations are very important in several ways:

1. The distribution of dark and baryonic matter at large scales can be simulated and thus researchers can study the properties of the distribution of voids.

2. The formation and evolution of voids can be simulated, revealing how gravitational collapse and expansion can form these structures.

3. The size, shape, density profile, and abundance of halos and voids are important properties that simulations can discern. Indeed, this thesis relies heavily on the computational ability to discern these properties and then compare them to observational data. This helps constrain cosmological parameters.

4. Galaxies present within voids have some interesting properties that Cosmological simulations allow researchers to investigate. The way voids can affect galaxy formation, evolution, and even clustering is an ongoing and interesting subject (Correa, 2022).

When simulating many different aspects of the universe, such as dark matter, dark energy, baryonic matter and gas dynamics, the computational stress increases drastically. The number of particles that simulations can handle increases every year following Moore's Law (Figure 1.1). While this seems promising, it poses a problem of computational power in that the time needed to simulate the increasing number of particles increases more rapidly than the readily available CPU power. Indeed, the complexity of an N-body, collisional, simulation grows as N^2, where N is the number of particles.

> **IllustrisTNG example**
>
> The IllustrisTNG simulation on galaxy formation is a good example. They run several simulations, one of which involves all hydrodynamical aspects, evolving $1820^3 = 6,028,568,000$ hydrodynamical cells, with the same number of dark matter particles and Monte Carlo tracers, in a box of size 106.5 Mpc^3 from redshift $z = 127$ to $z = 0$. This required $8,192$ computer cores, a peak memory of 25TB, and 19 million CPU hours (IllustrisTNG, 2018).

Hence, if we want to increase the number of simulated particles to better understand our Universe, we are going to need ever-growing computational power or a lot of patience. It is no surprise, therefore, that attempts at minimizing the required power
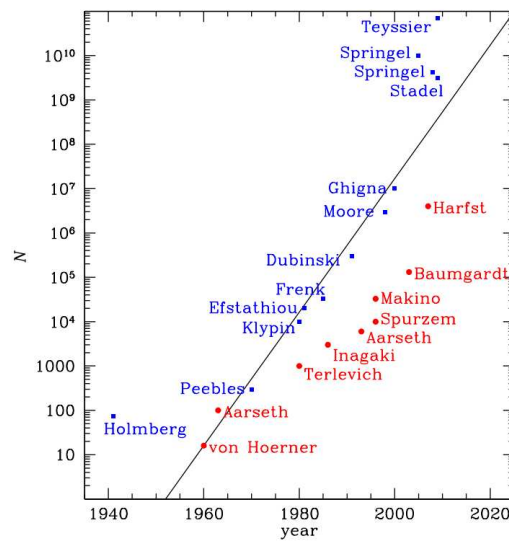
Figure 1.1: Figure showing how the number of particles in collisionless (blue) and collisional (red) simulations increases over the years, taken from Dehnen and Read (2011).

are constantly underway and several methods of simplification have been developed by the scientific community. Methods such as considering only specific types of particles, zooming techniques, softening, and tree codes, just to name a few.

Before thinking about applying a simplification, a framework within which the approximation will operate should be constructed. The objective or purpose of this experiment is to see if a useful number of voids can be simulated accurately only by considering a first-order Lagrangian approximation to the gravitational evolution of particles that make up the void. This would decrease the computational power needed significantly because if only first-order mechanics are considered, the number of equations a PC must solve decreases for each particle, exponentially decreasing the CPU power needed to simulate a universe. To assess if the desired accuracy holds, we must first be able to detect the voids and measure important quantities, such as the density field and halo mass function. To this end, we must first define what halo mass and density are.

Figure 1.1 shows, for example, that the difference in simulated particles in 2010 between a collisional (Harfst) and collisionless (Teyssier) was very significant. Indeed, the Teyssier simulation included 10^5 times more particles than the Harfst simulation.

1.3 Particle Linear Theory

Here I review particle linear theory (PLT) as developed by Marcos et al. (2006), analogous in this work to linear perturbation theory (LPT). PLT is important because it exactly describes the initial configuration of particles as presented in cosmological

simulations, and the evolution of a collection of particles that are self-gravitating in an expanding universe. It is a method that the Zel'dovich approximation attempts to simplify, due to its inherent complexity.

PLT Formalism

Garrison et al. (2016) offers a concise background on PLT which I shall briefly discuss as it is useful when considering the Zel'dovich approximation.

Consider an imaginary box of side lengths L. Inside this box, there are N particles arranged in a cubic lattice.

This box has special walls that let particles go from one side to the other, i.e. we apply periodic boundary conditions.

We can explain the variation in each particle's speed, considering that the universe around them is expanding and given that each particle exerts a gravitational pull on every other particle. This is explained with the following equation:

$$\ddot{\boldsymbol{x}}_i + 2H(t)\dot{\boldsymbol{x}}_i = -\frac{1}{a^3}\sum_{i\neq j}\frac{Gm_j(\boldsymbol{x}_i - \boldsymbol{x}_j)}{|\boldsymbol{x}_i - \boldsymbol{x}_j|^3} \tag{1.1}$$

Where H is the Hubble parameter, G is the gravitational constant and a is the scale factor. The left side of the equation describes the acceleration of each particle. The right side of the equation instead describes the gravitational force between particles, which depends on their relative positions. This is, therefore, an equation that primarily describes the balance between two forces: the expansion of the universe and the gravitational pull of particles on one another.

Now we can describe the position of any of the particles relative to the expanding universe: i.e. comoving position:

$$x_i(t) = R_i + u(R_i) \tag{1.2}$$

Where R_i is the particle's original position and therefore $u(R_i)$ is the displacement that particle i experiences. Because there are (almost) infinitely many particles in the universe, summing up their gravitational contribution exactly is very difficult. Therefore, special tools are used (such as Ewald-type Summation (Marcos et al., 2006) or an N-body Force Solver) to accurately simulate the contributions.

For the scope of this thesis, this is where depart, since the Zel'dovich approximation does not take into account the full description and certainly does not calculate the gravitational contribution of every particle in the system.

1.4 The Zel'dovich Approximation

First proposed in 1970 (Zel'dovich, 1970), this approximation is a simplification of otherwise very complex mechanics that occur between dark matter particles. What it proposes is a simple, ballistic trajectory that dark matter particles undertake. An application of such an idea is revisited years later in 1989.

Shandarin and Zeldovich (1989) originally imagined a mini-universe where all the particles are cold and moving with some initial speed $v(q)$, where q are their initial Lagrangian coordinates. The Lagrangian coordinate is used to track particles - think of them as markers that are initially attached to the particle and can be used to track the movement. Because this universe is cold, particles will keep moving in a straight line at the same speed (hence, ballistic trajectory).

They propose that, initially, particles have random velocities, rather than some pattern.

Since the gas is cold, no viscous or pressure forces are acting on the particles and hence no acceleration. This means that all particles move with constant velocity. Over time, particles will experience "trajectory intersections", where their paths will cross. This leads to forming overdensities. Particularly, the authors mention that particles with these initial properties and dynamics will go on to form "cellular structures" or "foam" regions - exactly reminiscent of the large-scale structures we see today.

Much like turbulent flow, the evolution of these particles follows a smooth, orderly initial state called *laminar flow*, which then, over time, evolves and becomes chaotic and unpredictable, characterised by irregular motion and the formation of overdensities (vortices and eddies in the case of gas dynamics).

The term "turbulence" originates from classical descriptions of incompressible fluids. In such fluids, density remains constant, leading to a divergence-free flow. Then the velocity is defined as the curl of a potential $\boldsymbol{v} = \text{curl}\mathbf{A}$. At large scales, irregular eddies of intermediate size emerge, while smaller scales are governed by viscosity. This phenomenon mirrors the complexity that the Zel'dovich approximation attempts to resolve by considering only ballistic trajectories and therefore almost entirely ignoring the small-scale viscosity effects.

If we were considering normal gas particles, these would collide, heat up and form shock waves; processes necessary for the formation of large-scale structures. However, in a medium of collisionless dark matter, we only observe *multi-stream configurations* which means that at some Eulerian point $\mathbf{x}$, particles that have arrived from different Lagrangian coordinates $\mathbf{q}_1$, $\mathbf{q}_2$, ... , $\mathbf{q}_n$ will intersect and form regions of higher density. These particles have different velocities $\mathbf{v}_1$, $\mathbf{v}_2$, ... , $\mathbf{v}_n$. The particles' position over time is therefore given by:

$$\mathbf{x}(t, \mathbf{q}) = \mathbf{q} + t \cdot \mathbf{v}(\mathbf{q}) \tag{1.3}$$

Where $\mathbf{v}(\mathbf{q})$ is a vector containing particle velocities at their Lagrangian coordinates $\mathbf{q}$.

As time goes on, these regions of higher density grow to form a web-like structure characterised by thin filaments of high density separated by large regions of low density (voids). These structures can be seen today through observations of large-scale structures.

At this point, the use of Fourier analysis to analyse the different modes that have chaotically arisen fails to work, since the different modes are not statistically independent. Hence, they propose to study the patterns formed by particles, without allowing them to interact with each other. This helps them to focus on how particles move and form patterns without getting distracted by other factors.

As the universe expands, matter behaves like cold dust due to the influence of gravity. To include the growth rate of regions of higher density which causes perturbations, as well as the expansion of the universe, a simple formula is proposed by Zel'dovich (1970):

$$\mathbf{r}(t, \mathbf{q}) = a(t)[\mathbf{q} - b(t)\mathbf{s}(\mathbf{q})] \tag{1.4}$$

Where, as in Eq 1.3, $\mathbf{r}$ and $\mathbf{q}$ are the Eulerian and Lagrangian coordinates respectively and $a(t)$ is the cosmological scale factor which accounts for the expansion of the universe over time: As the universe expands ($a(t)$ increases), distances between objects increase proportionally. $b(t)$ is the growth rate of linear density fluctuations in an expanding universe. This term represents how quickly small density fluctuations in the universe grow as the universe expands. It's related to the evolution of structures in the universe due to gravitational instability.

$\mathbf{s}(\mathbf{q})$ is the gradient of the gravitational potential $\mathbf{s}(\mathbf{q}) = \nabla\Phi(\mathbf{q})$ and represents spatial perturbations or how the gravitational force varies from its average value at each point in space. To clarify, the right-hand side of the equation combines the effects of cosmic expansion ($a(t)\mathbf{q}$) and the growth of density perturbations ($b(t)\mathbf{s}(\mathbf{q})$). The term $a(t)\mathbf{q}$ represents how the particle's Lagrangian position $\mathbf{q}$ evolves due to the expansion of the universe. As $a(t)$ increases with time, the particle's Lagrangian position is stretched along with the expansion of space.

The term $b(t)\mathbf{s}(\mathbf{q})$ describes how the particle's position is affected by the growth of density fluctuations. The gradient of the gravitational potential $\mathbf{s}(\mathbf{q})$ captures the spatial variations in density, and $b(t)$ scales the effect of these density fluctuations on the particle's motion. As density perturbations grow, they influence the particle's

trajectory, causing it to move towards regions of higher density and away from regions of lower density.

This equation is very different from Eq 1.1, as it describes the spatial distribution of matter in the universe, rather than the positions of individual particles. The Zel'dovich approximation is a first-order Lagrangian perturbation approximation because it involves perturbations. The growth function $b(t)$ depends on the mean density of the universe $\Omega = \bar{\rho}/\rho_{\text{crit}}$, but if we consider a flat matter-dominated universe ($\Omega = 1$), then the growth factor $b(t)$ and the scale factor $a(t)$ are approximately equal.

The overdensity $\Delta = \delta\rho/\rho$ is proportional to the gradient of the spatial perturbation $\mathbf{s}(\mathbf{q})$ and therefore $\Delta \propto \nabla^2\Phi(\mathbf{q})$. We can make the following transformations:

$$\mathbf{x} = \mathbf{r}/a(t), \quad \tau = b(t), \quad \mathbf{v}(\mathbf{q}) = -\mathbf{s}(\mathbf{q}), \tag{1.5}$$

to Eq 1.4 to obtain:

$$\mathbf{x} = \mathbf{q} + \tau \cdot \mathbf{v}(\mathbf{q}) \tag{1.6}$$

Which is equivalent to Eq 1.4. Eq 1.6 can describe the motion of a non-interacting medium up to the formation of caustics (regions of infinite density), after which the approximation becomes invalid.

The Zel'dovich approximation tends to remain accurate for a longer duration compared to the Eulerian linear theory, which describes how the density contrast δ evolves according to linear differential equations - derived from linearised versions of the equations of motion.

To conclude: the Zel'dovich approximation is more reliable because it uses the Lagrangian approach rather than using a Eulerian framework, where observations are made at fixed points in space. The method is also able to translate the Lagrangian positions of particles into their Eulerian positions, as can be seen in Figure 1.2

1.5   Defining a Halo

Now that we understand cosmological simulations, why they are important, how they are generated, what theoretical process is used to evaluate the evolution of constituent particles within the simulation, and why this Zel'dovich approximation can be very useful, we go on to define Halos and their properties.

To identify halos within a simulation, some constraints on their mass and density must be employed. There are a few theoretical definitions of halos, so we shall start by introducing the Spherical Top-Hat collapse model and the Friends-Of-Friends method, which are tools used by software to "detect" halos.
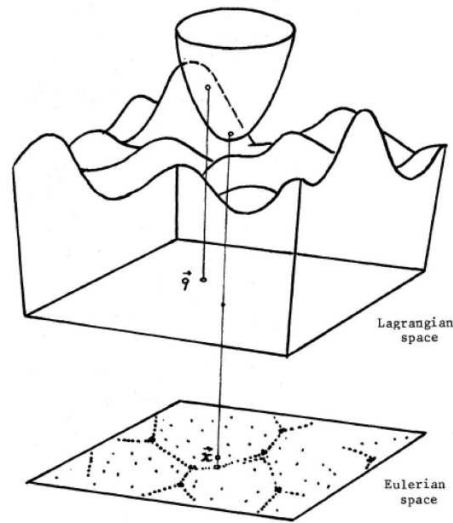
Figure 1.2: In the Zel'dovich approximation, the Lagrangian positions of particles are mapped into Eulerian space by using the initial gravitational potential field, denoted as $\Phi_0(q)$. This process involves a geometrical technique where a paraboloid, described by Equation 17 in Shandarin (2009), is adjusted until it is tangent to the surface of the initial potential at the Lagrangian point q. The Eulerian position of the particle is then determined by projecting the apex of the paraboloid onto the surface, providing the corresponding Eulerian coordinates

1.5.1   Top-Hat Spherical Collapse Model

The Top-Hat Spherical Collapse (TSC) (Gunn and Gott, 1972) model describes how a spherically symmetric density perturbation expands and collapses to form a halo in an Einstein de Sitter (EdS) Universe. When a halo of dark matter first forms, its expansion rate is the same as that of the background universe. However, since it contains more matter than its surroundings, the halo's expansion gradually slows due to self-gravitation until it reaches its turn-around radius R_{ta}, where it stops expanding. At this point, the dark matter halo begins to re-collapse onto itself due to its growing gravity. The collapse halts when the density of the halo reaches a critical density value $\Delta_c = 18\pi^2 \simeq 178$ times the background critical density. The halo is now virialized and hence in gravitational equilibrium with itself and the surroundings (Figure 1.3).

Figure 1.3 shows the evolution of a halo's density over time with the scale factor on the x-axis. Four different scenarios are shown:

1. The lowest curve shows the density evolution of the background universe, which is well-known to follow a a^{-3} scaling.

2. The next lowest dotted curve shows a standard linear Eulerian theory approach which, as we've discussed, breaks down quite early in being able to describe the density evolution of halos.
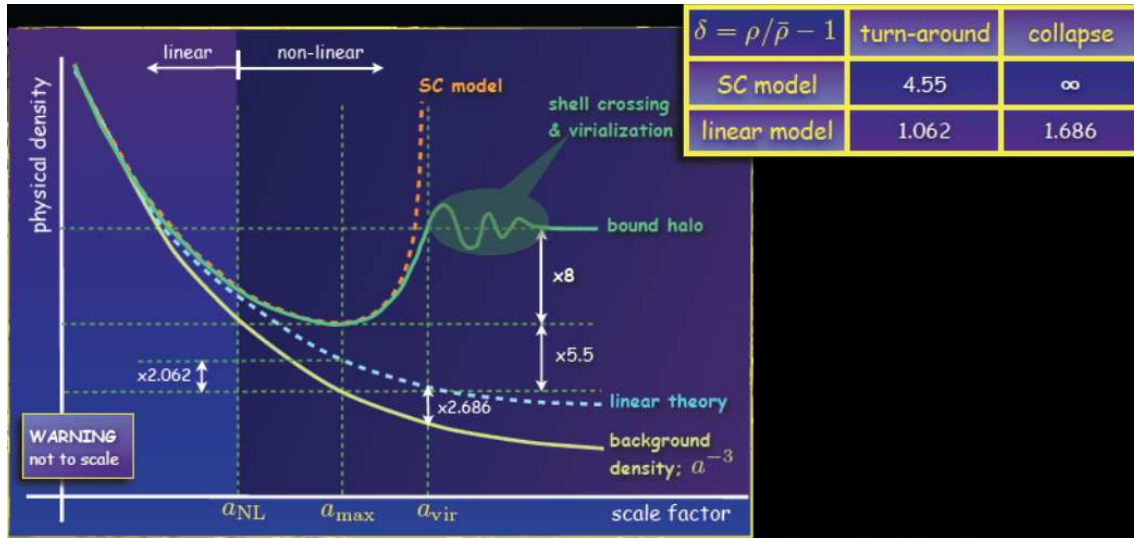
Figure 1.3: Depiction of density evolution for different models (Brooks, 2021).

3. The top-most dotted curve shows the Spherical Collapse model, which correctly predicts the turn-around of an overdensity but then fails to incorporate virialization and shell crossing.

4. The middle, the solid line shows the real path an overdensity would take in an expanding universe and therefore correctly shows shell crossing and virialization - processes necessary for the formation of halos.

Throughout this work, different definitions of halo mass are employed. When a halo is located using halo finding techniques, and a "centre" is defined, its mass, M_Δ, is calculated by considering the density contrast Δ contained within a certain radius r_Δ:

$$\int_0^{r_\Delta} r^2 dr \rho(r) = \frac{\Delta}{3} \rho_{\text{crit}} r_\Delta^3 \tag{1.7}$$

Recall that the density contrast Δ is the ratio of the actual density of the halo to the mean density of the universe at a given time. We call this the virial mass from a spherical top-hat collapse model, $M_{\text{th - vir}}$.

Alternatively, we define the virial radius as the radius at which all mass has virialised, and beyond which the mass continues to collapse onto the dark matter object. In general, this occurs at a density contrast $\Delta = 200$, but it does depend on the cosmology used. Hence, a common estimator is M_{200}. (White, 2001)

Alternatively, a different mass estimator can be employed, similar to the approach adopted by Jenkins et al. (2001). In their analysis of simulations done by the Virgo consortium, they assign a mass equal to the sum of masses of the particles that make

up the halo, as found by their halo finder. Although it is hard to interpret theoretically, this method has an advantageous trait: its mass function is independent of cosmology. This is very similar to the approach I employ when calculating the Halo Mass Function in Chapter 2.6.

1.5.2 The Friends of Friends method

One approach to distinguishing halos from a group of dark matter particles is the Friends of Friends algorithm. This specific algorithm is employed and improved upon in Amiga's Halo Finder (AHF), which we shall cover shortly.

In this method of halo finding, pairs of particles that supposedly make up a halo are "linked" to one another by selecting a linking length h_{link} and checking if these particles lie within the linking length distance. If they are, they are designated *friends*. Halos are defined as sets of particles that are connected by one or more friendship relations. This means that if Particle A is a friend of Particle B, and Particle B is a friend of Particle C, then Particles A, B, and C would all belong to the same halo because they are connected through the friendship relations.

The linking length is controlled with a certain density parameter:

$$\rho_{\min} = \frac{2\bar{m}}{\frac{4}{3}\pi h_{\text{link}}^3} \tag{1.8}$$

Where $\bar{m}$ is the single average mass of particles in the simulation. The minimum density that a sphere can have to be considered a halo is controlled by the linking length that defines a volume. While computationally simple, this method struggles when it comes to choosing the arbitrary minimum number of particles to consider, $N_{\min}$(Figure 1.4).

A minimum number of particles must be chosen to eliminate spurious halos. The problem here is that if $N_{\min}$ is too low, a larger fraction of unwanted halos will be detected. If too large, some smaller, real halos will be rejected. Furthermore, FoF has a tough time distinguishing between a large halo, and a collection of small halos nearby, not to mention halos within halos.

In both TSC and FoF, the most crucial parameter that is used to detect a halo is its density.

1.6 Amiga's Halo Finder

The Amiga Halo Finder (AHF) (Knollmann and Knebe, 2009) is the successor to the halo finder MHF (Gill et al., 2004) and therefore is based on its operational idea.

A hierarchical grid structure is generated by a process called Adaptive Mesh Refinement (AMR). Instead of using a traditional grid, which has a fixed resolution
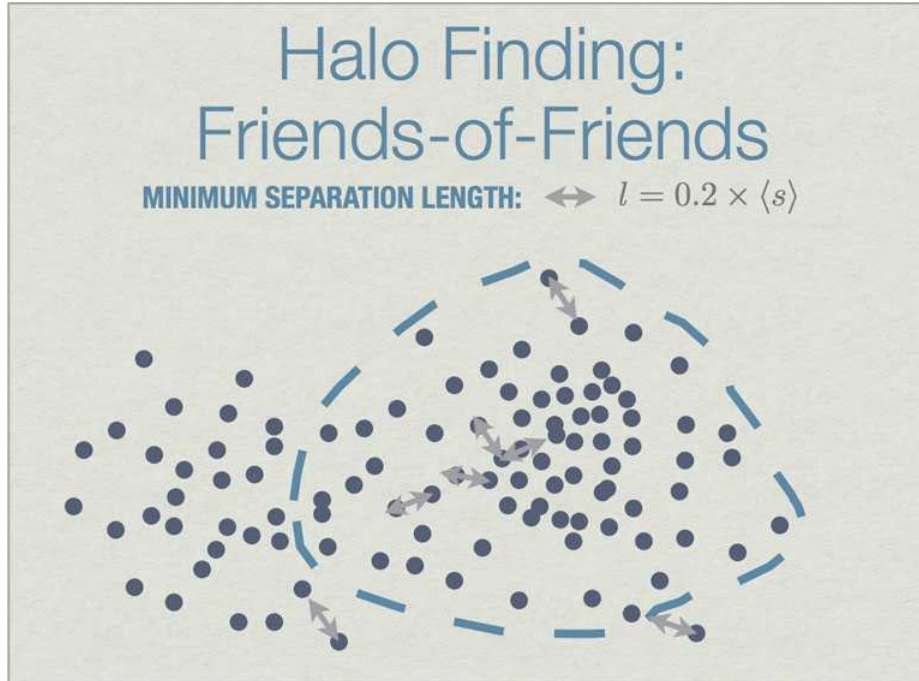
Figure 1.4: All particles closer than 0.2 of the mean inter-particle separation (s) are linked, and all links of particles are followed recursively to form the halo groups. (Skory, 2010)

throughout, which can be wasteful in regions where we don't need high precision, AMR intelligently and dynamically adjusts its resolution based on the complexity of the problem. Vast voids and smooth airflow can have a coarse grid associated with them, whereas regions of dense star formation and turbulent eddies have a refined grid associated with them.

This refinement of the grid is done dynamically, as the simulation progresses. It allows Mhf to then organize, hierarchically, individual prospective halo centres by simply identifying regions of localized peaks in density. This is what we mean by a hierarchical grid structure.

Within Ahf, the AMR technique sections off grids that are effectively isodensity contours that encompass particles. Isodensity contours represent surfaces in three-dimensional space where the density of dark matter particles remains constant.

As long as the contours are not ambiguous, Ahf will collect them. If they are ambiguous, misidentification of halos can occur. This could happen if the density variations in the region of interest are too subtle or irregular.

Once Ahf has traced the isodensity contours and identified potential halo regions, it analyses the density peaks and pinpoints a prospective halo centre.

Ahf assumes spherical symmetry ($\rho = \rho(r)$) in the distribution of dark matter within halos. It implies that the density ρ of dark matter particles depends only on
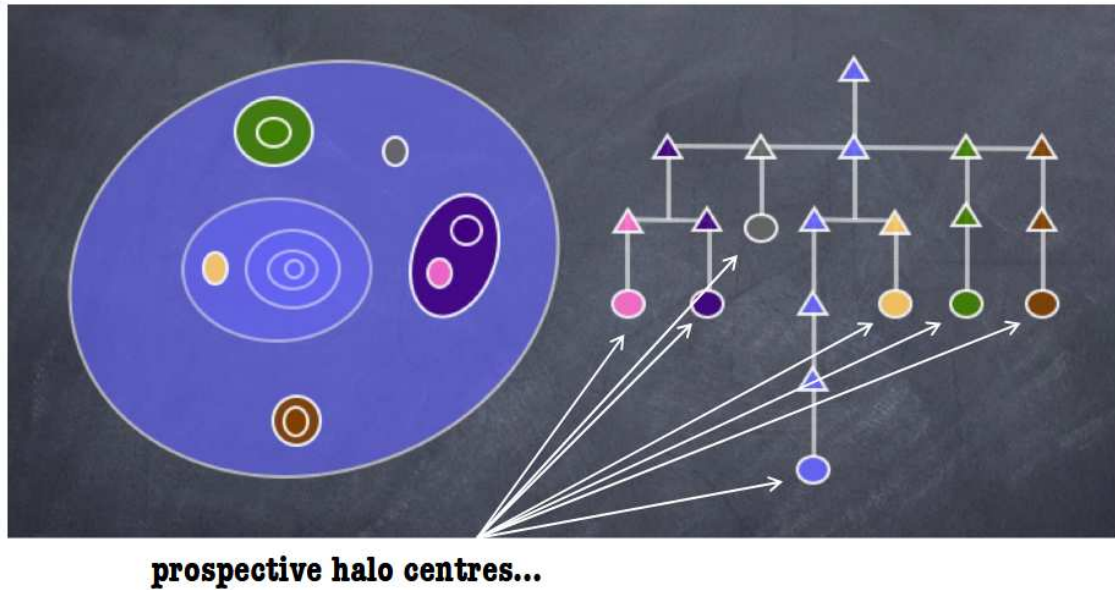
prospective halo centres...

Figure 1.5: Hierarchical grid that contains several prospective halo centres. Adopted from Knollmann and Knebe (2009)

the radial distance r from the halo centre[1].

Both Poisson's equation ($\Delta\phi = 4\pi G\rho$) and Newton's Force equation $\left(\frac{d\phi}{dr} = \frac{GM(<r)}{r^2}\right)$ are solved. The former determines the gravitational potential within halos by relating it to the density distribution of matter using the Laplacian operator Δ, while the latter is used to calculate the gravitational force acting on the dark matter particles inside the prospective halo.

Once the equations are solved, AHF iteratively examines the distribution of density within the region to attempt to identify a halo edge. This is because not all particles may be gravitational bound. Unbound particles are removed iteratively until a fully bound halo edge is deduced.

This was the process employed on all simulations that were run. AHF outputs files that can be analysed. Given the halos, we can now turn to the Halo Mass Function.

1.7 Halo Mass Function

In cosmology, the halo mass function (HMF) is a mass distribution of dark matter halos. Specifically, it gives the number density of dark matter halos per mass interval. Tinker et al. (2008) chose to portray the HMF as follows:

---

[1]This can be quite a powerful assumption as, realistically, halos do not form completely spherically but, rather, follow a more complex pattern of flattening in one direction and then collapsing in the other two dimensions, much more like a pancake (Achitouv and Corasaniti, 2011).

$$\frac{dn}{dM} = f(\sigma)\frac{\rho_m}{M}\frac{d\ln\sigma^{-1}}{dM} \tag{1.9}$$

Where $f(\sigma)$ is parameterised as:

$$f(\sigma) = A\left[\left(\frac{\sigma}{b}\right)^{-a} + 1\right]e^{-c/\sigma^2} \tag{1.10}$$

expressed as a combination of constants (A, a, b, c) to be calibrated.

Then, $\sigma(M, z)$ is the variance of the linear matter density field smoothed on a certain scale R, defined as:

$$\sigma = \int P(k)\hat{W}(kR)k^2 dk \tag{1.11}$$

and $P(k)$ is the linear, matter power spectrum as a function of wave-number k, and $\hat{W}$ is the Fourier transform of the real-space top-hat window function of radius R ($\hat{W}(kR)$). A top-hat window function just means that it assigns a value of 1 within a certain radius R and 0 outside that radius. Its Fourier transform is the frequency of the window function. All in all, this describes the spatial distribution of density fluctuations.

Indeed the number of halos we expect ($N_{\exp}$) is just the integral of the halo mass function in Equation 1.9 (Stopyra et al., 2021a):

$$N_{\exp} = \frac{\rho_m}{M}\int \frac{dn}{dM}dM \tag{1.12}$$

In this work, we utilize the `pynbody` routine `hmf` to compare the expected halo mass function of a simulation with the real halo mass function derived from Eq 1.12. This comparison is crucial to understanding the behaviour of a simulation.

Analyzing the halo mass function is important when it comes to constraining cosmological parameters because it is sensitive to these parameters. For instance, the Navarro–Frenk–White (NFW) profile (Navarro et al., 1996) characterizes halos in terms of their spatial mass distribution and is sensitive to the critical density of the universe, denoted as ρ_c:

$$\rho_c = \frac{3H_0^2}{8\pi G} \tag{1.13}$$

This explains the importance of the analysis of the HMF; if dark matter halos can accurately be simulated, important cosmological parameters can be constrained.

1.8   Generating Initial Conditions

When initializing simulations to study how galaxies and structures form in the universe, initial conditions are required. These conditions include the position of particles and how they move at a very early time in the universe's history when deviations from homogeneity are still linear - meaning just before the clumping of matter begins.

To this end, a random pattern is formed that matches our current understanding of the distribution of matter, a Gaussian Random Field (GRF).

A GRF, pictured in Figure 1.6, is what we believe the universe may have looked like soon after the Big Bang and is also the distribution of matter in the initial conditions file that will be fed to `Gadget`, the simulation algorithm.
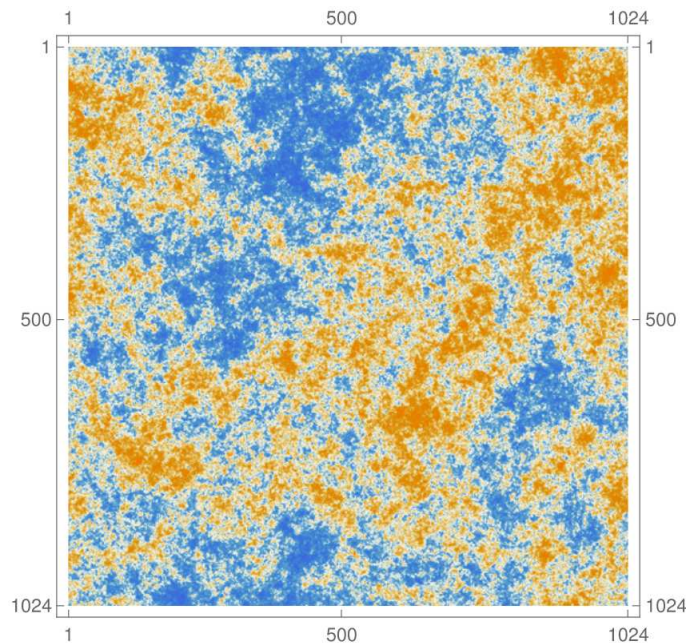

Figure 1.6: Shown here is a simulated depiction of a two-dimensional, Gaussian random field exhibiting a scale-invariant power spectrum $P(k) \propto k^{-2}$. This realization encapsulates statistical properties fully characterized by the two-point function $\langle \phi(x)\phi(y) \rangle$, emphasizing the Gaussian nature of the field. Adopted from Goon (2021).

To allow the computer to work with such an abstract concept, convolutions (a way to combine two sets of data into a third so that a CPU can understand them) are involved. `GenetIC` generates the starting conditions for the simulation, and then performs small changes to these conditions to create "genetic modifications."

By producing a cosmological covariance matrix (i.e. a matrix representing the relationship between different density fluctuations), we can set the scene for density

contrasts to be input. A vector δ is formed which contains discrete values: the average density of specific cells or regions of the simulation. Then, each vector element is translated into particles with positions and velocities using Lagrangian perturbation theory, the lowest-order terms of which can be approximated by the Zel'dovich approach.

With the initial conditions generated, we are ready to run the simulation.

1.9   Running the Simulation

`Gadget2` (Springel, 2005) is the software used to simulate the generated initial conditions in this thesis. It combines the hierarchical tree method of gravitational interaction and the smoothed particle hydrodynamics (SPH) method for gases and liquids to accurately simulate many aspects of the Universe. Since this thesis mainly dealt with dark matter, SPH was not implemented in any of the simulations. Rather, the calculation and evolution of dark matter solely used gravitational interaction.

This hierarchical tree method of gravitational interaction involves two steps. One involves computing the gravitational forces using a hierarchical multipole expansion (HME) method, which works on various scales.

The other step is the TreePM algorithm, which combines the HME method with Fourier techniques for more advanced computation, able to simulate short-range and long-range gravitational interactions.

More specifically, it operates by solving the collisionless Boltzmann equation coupled with the Poisson equation within an expanding background Universe, typically represented by the Friedman-Lemaitre model. These equations describe the evolution of dark matter particles in a vast cosmic web of structure formation.

`Gadget` takes initial conditions, including positions and velocities of particles, and can employ various methods to generate the final z=0 snapshot. This thesis uses two methods: the full N-body approach and the Zel'dovich approach. The latter has already been discussed. In the former method, phase-space density is sampled using a finite number of N tracer particles. By iteratively solving the equations governing gravitational interactions and particle dynamics, Gadget generates simulations that capture the emergence and evolution of cosmic structures, such as galaxies, clusters, and filaments. An example of a final product of simulated particles is shown in Figure 1.7.

1.10   Power Spectrum

It's important to understand how our knowledge of the density field has evolved. Scientists have studied the Cold Dark Matter (CDM) theory in detail, which helps us understand how density varies across the universe.
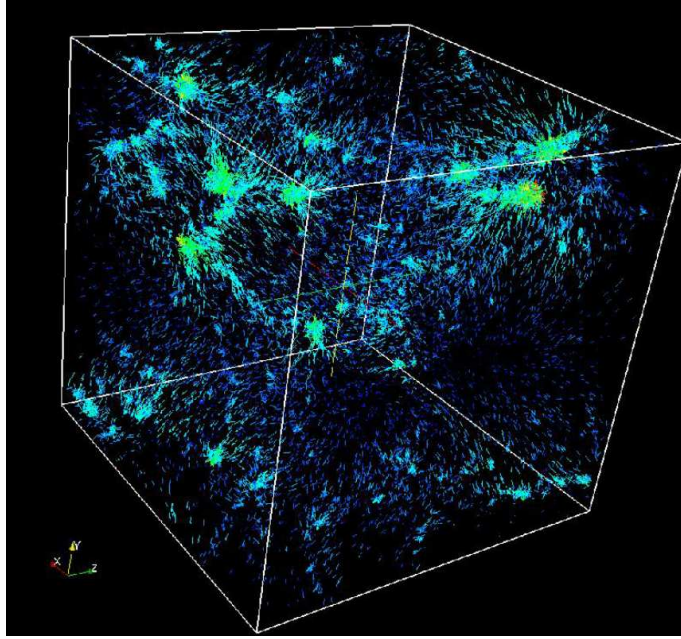
Figure 1.7: A subset of the 20,000 particles at z = 0 from the GADGET-2 simulation. The particles are shown with vector arrow glyphs which are sized and coloured by their velocity magnitude (blue: slowest, red: fastest) (Heitmann et al., 2007).

When the universe was dominated by radiation, density fluctuations could only grow significantly if they were bigger than the horizon scale. Fluctuations smaller than this faced limitations due to interactions with radiation and pressure. As the universe shifted to being matter-dominated, gravity started affecting all fluctuations equally, leading to similar growth rates for all.

When we track the fluctuation spectrum over cosmic history, we notice a change in growth rate at a specific scale, marking the transition between these two cosmic eras (check out Figure 1.8). If the initial density fluctuations follow a certain pattern, we see a gradual change in density power over time. This shift is significant, spanning a wide range of sizes, from larger to smaller. We can describe this change with an equation that shows how density evolves:

$$|\delta_k|^2 \propto= Ak/(1 + \alpha k + \beta k^{3/2} + \gamma k^2)^2 \qquad (1.14)$$

Here, A is a normalization constant (Davis et al., 1985). k can be taken with different units, but if it is taken to have present length units, then the parameters in the above equation take the values $\alpha = 1.7l$, $\beta = 9.0l^{3/2}$ and $\gamma = 1.0l^2$ where l is the cosmological density parameter at the present time.

Understanding the initial conditions of the universe is crucial, particularly in this thesis which plays with these conditions. Therefore, changing the Hubble parameter, for example, would alter the power spectrum overall. The power spectrum of
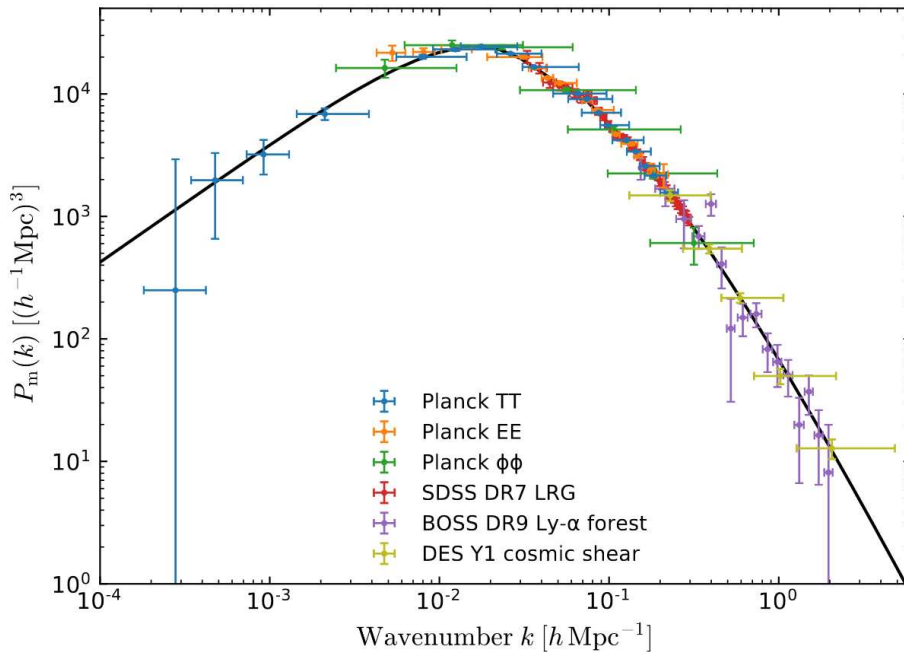
Figure 1.8: Matter power spectrum inferred from various cosmological probes. (Collaboration, 2018)

matter, especially at large scales, provides insights into the typical sizes and distribution of voids, which is why it is so important to our understanding of cosmological structures.

1.11 Autocorrelation Function

Often denoted as $\xi(r)$, the autocorrelation function is a measure of the degree of spatial correlation between density fluctuations at different points in the universe as a function of their separation distance, r, which can have any units. In this thesis, r is chosen to have simple units of Mpc as we are analysing voids.

More plainly put, this function can show how matter, at different scales, tends to cluster and offers the probability of finding two halos separated by a distance r. It can reveal patterns across different scales; patterns that only exist based on the cosmological parameters (Peebles, 1980). The relationship between the power spectrum $P(k)$ and the autocorrelation function $\xi(r)$ is given by the forward Fourier transform:

$$\xi(r) = \frac{1}{(2\pi)^3} \int e^{i\mathbf{k}\cdot\mathbf{r}} P(k) d^3 r \qquad (1.15)$$

It is also defined as the average of the product of fluctuations in density at two points, normalized by the average density:

$$\xi(|x_1 - x_2|) = \langle \delta(x_1)\delta(x_2) \rangle \tag{1.16}$$

where $\delta(x) = (\rho(x) - \bar{\rho})/\bar{\rho}$. In this thesis, the autocorrelation function is used as a tool to quantify analytically the usefulness of the Zel'dovich approximation in showcasing the clustering properties at small and large scales when compared to a full N-body simulation of dark matter - which tends to exhibit a more accurate result.

In conclusion, this function is a viable tool and should be used to evaluate how accurately voids are distributed in a simulation solely using dark matter and the Zel'dovich approximation to model particle dynamics. The more accurate the distribution of voids is when compared to full N-body simulations and real observations, the more promising the conclusions will be and therefore the more likely this approximation can be used for new science.

1.12   Void Density Profile

In the vast regions known as cosmic voids, the distribution of matter is still far from uniform. Spanning tens of millions of light-years across, these voids represent significant components of the cosmic web which play an important role in the shaping of the large-scale structures of the universe.

The study of voids and especially the defining of their density profiles is of great importance to modern cosmology and is thus explored thoroughly in this thesis. The profiles provide unique insights into the processes that control the formation and evolution of cosmic structures. By analysing the distribution of matter density, valuable information on void evolution, cosmic expansion, dark matter and the nature of dark energy can be gathered.

Hence, voids serve as valuable laboratories for testing cosmological theories since their relatively simple geometry and low-density contrast with the outside universe make them ideal candidates for theoretical modelling and numerical simulations. By comparing the N-body simulation void density profile to the Zel'dovich void profiles, we can analyse whether the approximation can predict and/or constrain important cosmological parameters, such as the density of dark matter and dark energy, and the amplitude of primordial density fluctuations (Beygu et al., 2016).

Moreover, the properties of void density profiles, such as their shape, have very important implications for galaxy formation and evolution. This is not explored in this thesis but if we can constrain cosmological parameters and therefore the shape of the density profile of voids, then the evolution of galaxies within can be further explored. For example, Courtois et al. (2023) delve into the relationship between luminous and dark matter in cosmic voids, highlighting discrepancies in void emptiness assessments and identifying the Hercules void as a potential "pristine

volume" with minimal dark matter content, which has very important consequences for both galaxy formation and cosmological parameter constraining.

This thesis hoped to compare the density profiles of voids found in all simulations with a more theoretical approach, but it was limited due to time. If time was given, then the calculated density profiles would have been compared to a theoretical profile, such as the one in Hamaus et al. (2014), where the void density profile is defined as the spherical average of the deviation in mass density around a void centre from the mean density across the Universe, expressed as $\rho_v(r)/\bar{\rho} - 1$. To then estimate the density within a shell of thickness $2\delta r$ at a distance r from the void centre located at the origin, tracer particles are used. This density is expressed as follows:

$$\rho_v(r) = \frac{3}{4\pi} \sum_i \frac{m_i(\mathbf{r}_i)\Theta(r_i)}{(r + \delta r)^3 - (r - \delta r)^3} \tag{1.17}$$

where m_i is the mass of any singular particle i, r_i is its coordinate and $\Theta(r_i)$ is a sort of "radial bin selector" and helps with counting the number of particles that fall within a specified shell thickness, much like a filter. This theoretical process is extremely similar to the process outlined in methodology Section 2.8

Overall, the density profile of voids is a fundamentally important concept which should be studied in more detail and the Zel'dovich approximation for large-scale structure evolution could offer new insights into the overall profiles, since with a simpler approximation to PLT, one may generate much larger cosmological simulations and be able to stack a greater number of voids, thus possibly reducing uncertainties and constraining important cosmological parameters, shedding new light on the origin and evolution of the Universe.

Chapter 2

Methodology

In this section, the steps taken for the work done in this thesis are explored. Below is a brief outline, with more specific descriptions to follow.

First, initial conditions are generated using `GenetIC`. This is done for both the normal simulation pair and the Zel'dovich simulation pair. A pair of simulations consists of a normal and reverse initial condition simulation.

Then, the full N-body simulations are evolved using `Gadget2`, wherein only dark matter particles are considered. When the simulations reach $z = 0$ from $z = 50$, their power spectrum is generated to see the distribution of matter, clustering properties, cosmological parameters, etc. The Zel'dovich simulations do not require the same evolution from $z = 50$ and can be modelled at $z = 0$ since the effects of cosmological expansion are ignored.

Then, AHF is employed to find the halos in the pairs of simulations using a hierarchical grid structure and adaptive mesh refinement. The halo mass function is calculated and plotted which shows the halo number density that falls within specified mass bins. After the HMF, a plot of the position of halos is done that shows promising matching between Zel'dovich-generated halos and full N-body halos.

Particles that belong to specific halos have IDs associated with them. They are matched with the ID location of the normal simulation to detect the voids. The same is done to the Zel'dovich pair.

Then, the density profile of voids is calculated in the following way: First, the density for each particle is retrieved, and then its volume. The volumes are summed over all particles in the void to get the total void volume. The radius is that of a sphere with the same volume. The centre of mass is then computed and the density profile is generated in radial shells around each centre of mass. The profile shows density as a function of radius. The density profiles of Zel'dovich voids are compared to full N-body voids.

From the density profile, we can retrieve important conclusions, such as void evolution and possibly constraining cosmological parameters.

2.1 Generating the Initial Conditions

The plot on the left in Figure 2.1 shows the positions of particles based on the initial conditions, which have been generated by `GenetIC` (Stopyra et al., 2021b). When generating the initial conditions, only very few parameters are tweaked to explore different simulation outcomes. To test whether the Zel'dovich approximation is suitable for linear void evolution, the cosmological parameters for all simulations are kept the same and can be seen in Table 2.1.

Table 2.1: Cosmological Parameters

Parameter	Value
Ω_b	0.0
Ω_m	0.3111
Ω_Λ	0.6889
s_8	0.8288
n_s	0.9665
Hubble	0.6766
z_{in}	50.0

where:

1. Ω_b is the fraction of the critical density of the universe contributed by baryonic matter (ordinary matter composed of protons and neutrons). In this case, it's set to 0.0 because we are only interested in dark matter.

2. Ω_m is the fraction of the critical density of the universe contributed by total matter, including both dark matter and baryonic matter.

3. Ω_Λ is the fraction of the critical density of the universe contributed by the cosmological constant or dark energy. It represents the energy density associated with the vacuum of space, causing the universe's expansion to accelerate.

4. s_8 is the amplitude of matter fluctuations on scales of $8 \, \text{h}^{-}1 \, \text{Mpc}$. It's a measure of the overall level of clustering of matter in the universe.

5. n_s is the spectral index of the primordial power spectrum of density fluctuations. It describes the shape of the spectrum and how the amplitude of fluctuations varies with scale. A value of 0.9665 indicates a slightly red-tilted spectrum, where smaller scales have slightly less power compared to larger scales.

6. "Hubble" is the present-day value of the Hubble parameter, which describes the rate of expansion of the universe, expressed in units of km/s/Mpc.
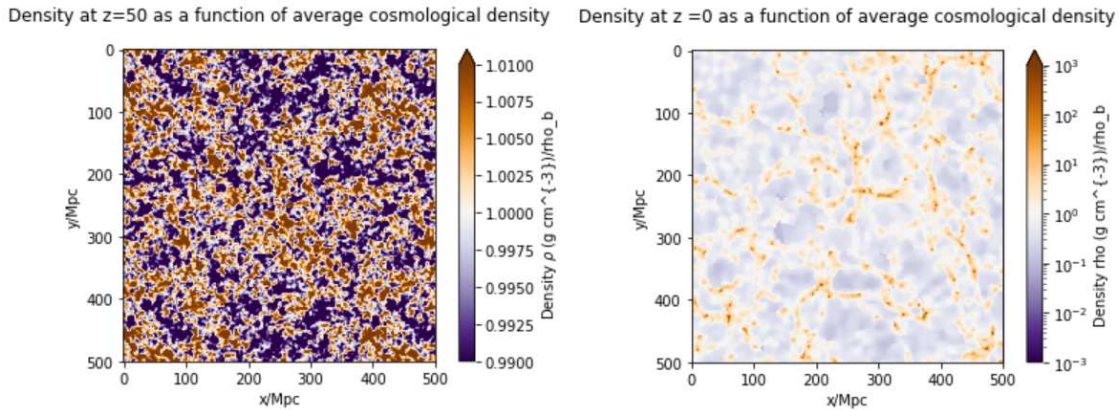
Figure 2.1: Comparison between matter density plots at z = 50 (right) and z = 0 (left) for the full N-body simulation.

7. z_{in} is The initial redshift or starting redshift of the simulation. It represents the redshift at which the simulation begins its evolution.

2.2 Running the Simulations

Initial conditions are generated using `GenetIC` at z = 50 for the full N-body model and z = 0 for the Zel'dovich model. While simulating the universe from an earlier time, say z = 99, would give more detail, z = 50 was chosen to maximize accuracy while minimizing the computational power needed, thereby also maximizing efficiency. A pair of simulations for both models are generated. One has its initial conditions reversed compared to the other meaning that overdensities are considered underdensities and vice versa so that $\delta_A(\mathbf{x}, t_{\text{initial}}) = -\delta_B(\mathbf{x}, t_{\text{initial}})$, a process that was suggested by Pontzen et al. (2015). Once the initial conditions are prepared on both simulations, they are evolved using the software `Gadget2` which, for the case of the Zel'dovich simulation, can be configured to implement the relevant approximation. Therefore, this simulation is computationally less expensive compared to full gravitational simulation. An example of the initial conditions for the case of the full N-body simulation can be seen in Figure 2.1.

Figure 2.1 shows the initial matter density in a 500Mpc^2 slice at z=50 (left) and z=0 (right). The density is measured in gcm^{-3}/ρ_b, meaning it has been normalised by the average cosmological density, ρ_b. What we are seeing is the initial almost completely homogeneous distribution of matter (notice the colour bar) on the left transforms, after the right mechanical equations are computed, to the very different distribution of matter on the right. The right shows a more modern-day visual of large-scale structures: nodes and filaments of high density (halos), surrounding almost empty regions (voids).

2.2.1   Snapshot Loading and Density Distribution

An analysis of the density distribution of both the Zel'dovich and full N-body simulation is carried out, firstly by loading the snapshot data using the Python package `pynbody`. Then, slices in space are formed using specific parameters, chosen to ensure consistency. The lower and upper bounds for the z-axis are a = 5 Mpc and b = 10 Mpc and the width is set to 1000 Mpc. The mean density is calculated for each slice and converted to gcm^{-3}. Then, the density distribution is generated with the `sph.image` function and is re-scaled against the cosmological mean density.

`sph.image` can visualise the distribution of a specified array, in this case density, within the simulation volume provided.

Figure 2.2 displays the notable contrast in density distributions, showing that the Zeldovich approximation (left) tends to yield a more uniform distribution with an average density of approximately 0.2, whereas the simulation incorporating higher-order mechanics (right) displays greater variability, with an average density of 0.01.
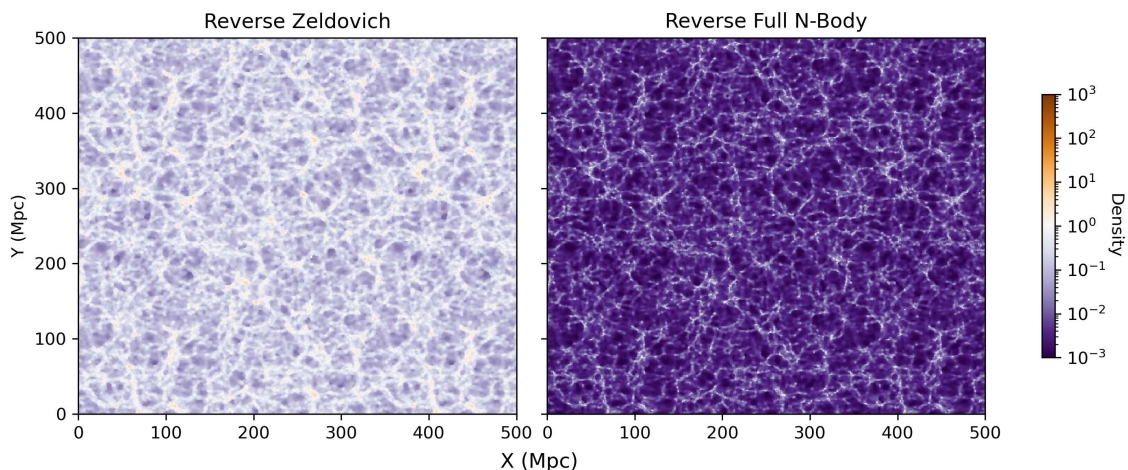

Figure 2.2: Comparison between density plots at z=0 generated from just varying the initial conditions (left) and full N-body simulation (right). The density is plotted in gcm^{-3}/ρ_b and hence is normalised by the background average density and is therefore unitless.

2.3   Power spectrum

To investigate the efficacy of the Zel'dovich approximation in simulating voids, the power spectra of the forward and reverse simulations were compared. The power spectrum serves as a valuable tool for quantifying the distribution of matter density fluctuations on different spatial scales.

First, using `nbodykit`, a CSV catalogue is formed that contains the position of every particle in the simulation as well as their masses.

Then, this catalogue is converted into a mesh. This conversion is achieved using Triangular Shaped Cloud (TSC) interpolation on a 256^3 mesh grid. This process discretizes the continuous distribution of objects in the catalogue into a discrete grid, with the following routine and parameters:

```
mesh = f.to_mesh(resampler = 'tsc', compensated = True,
interlaced = True, Nmesh=256, position='Position',
BoxSize=s.properties['boxsize'].ratio("Mpc a h**-1"))
```

Where s is the snapshot to be considered. "Compensated" being true means that correction factors to the density field in Fourier space are applied for the TSC to work. "Interlaced" means that the effects of aliasing are reduced. Aliasing occurs when we are Fourier transforming the density field onto the mesh - it involves interpolating objects onto two separate meshes distinguished by half of the cell size. When these two are combined in Fourier space, aliasing effects are reduced.

Next, the power spectrum of the density field is computed using Fast Fourier Transform (FFT). It calculates the power spectrum in 1D mode, meaning it assesses the distribution of fluctuations along the line-of-sight direction. Parameters can be varied to control the spacing and minimum value of the wavenumber k.

In Figure 2.3, we compare the power spectra of the Zel'dovich simulation with that of the full simulation.
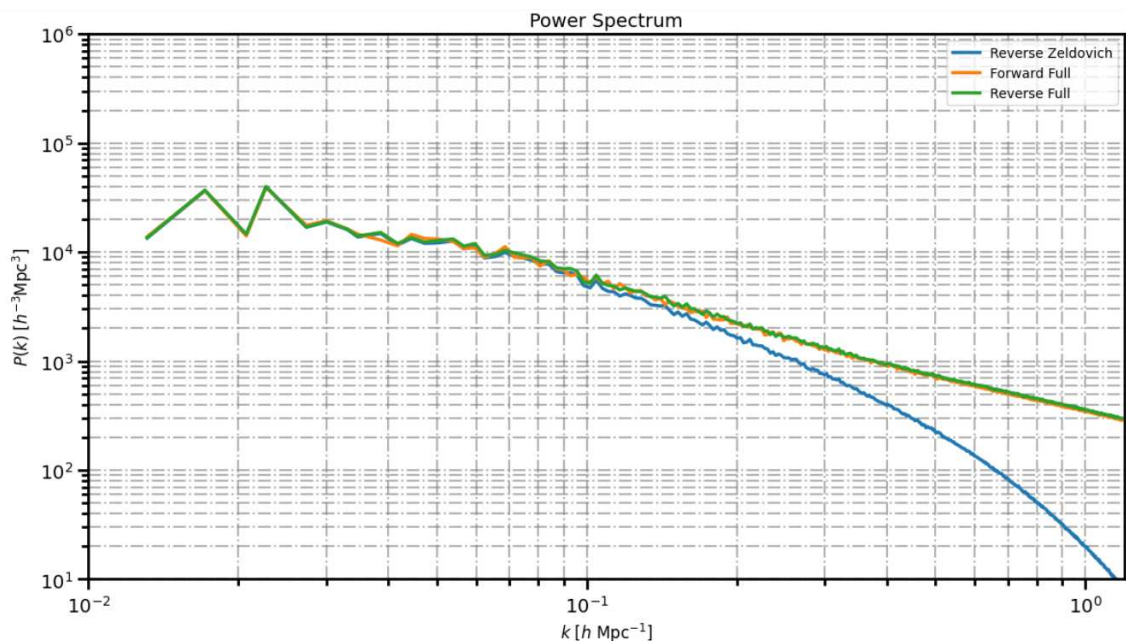

Figure 2.3: Comparison of power spectra between Zel'dovich and full simulations. The Zel'dovich power spectrum deviates noticeably from the full simulation at high wavenumbers (k), indicating its limitations in capturing small-scale structures.

2.4   Autocorrelation

The autocorrelation function is calculated for the power spectra generated in the previous chapter. We initialise a range of separation distances r for which we will compute the autocorrelation function $\xi(r)$ and, for each value of r, perform the following calculations:

- Exclude any nan numbers from the power spectrum.

- Utilize the inverse Fourier transform of the power spectrum to obtain the integrand, which means multiplying the exponential term e^{ikr} by the power spectrum $P(k)$.

- The integral is calculated using the trapezoidal rule over the specified range r following Eq 1.15.

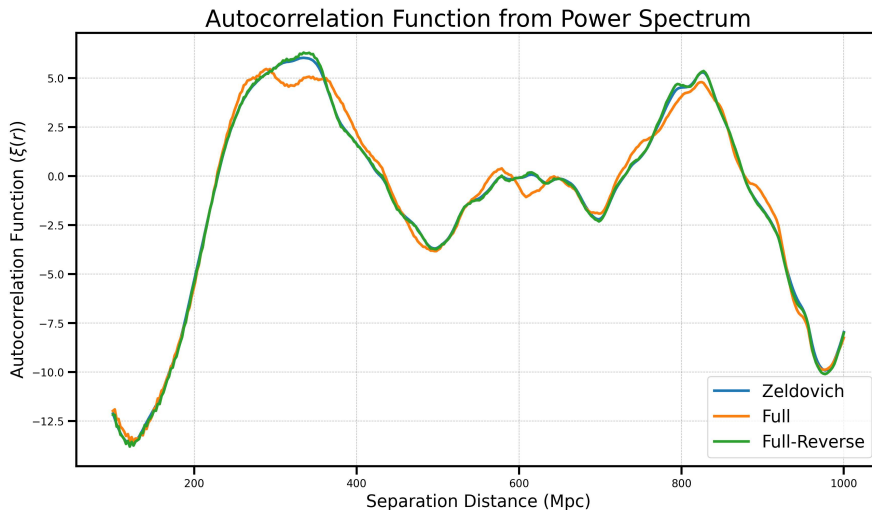Finally, it is plotted in Figure 2.4. Two other separation distance ranges are explored in the conclusion.

Figure 2.4: Comparison between different autocorrelation functions calculated using the Fourier transform of the previous power spectra and focusing on separation distances that correlate with large-scale structures ranging from 150 - 1000 Mpc, which the Zel'dovich approximation can follow very accurately.

2.5   Finding Halos using AHF

Amiga's Halo finder was employed to find halos in the full and Zel'dovich simulation using the data provided in Table 2.2.

The criteria in the table are mostly kept by default, except $N_{\mathrm{minPerHalo}}$, which controls the number of particles needed to define a halo, ρ_{vir}, controlling the virial

Table 2.2: AHF Parameters

Parameter	Full Simulation	Zel'dovich Simulation
$L_{\text{gridDomain}}$	64	64
L_{gridMax}	16777216	16777216
$N_{\text{perDomCell}}$	2.0	2.0
$N_{\text{perRefCell}}$	2.5	2.5
V_{escTune}	1.5	1.5
$N_{\text{minPerHalo}}$	25	10
ρ_{vir}	0	1
D_{vir}	200	30
$R_{\text{maxGatherRad}}$	3.0	3.0
$L_{\text{levelDomainDecomp}}$	6	6
$N_{\text{cpuReading}}$	1	1

density, and D_{vir}, which controls the virial overdensity criterion. The rest are as follows:

1. LgridDomain: Size of the domain grid in 1D.

2. LgridMax: Size of finest refinement level to be generated, which in short controls the spatial resolution.

3. NperDomCell: Number of particles triggering a refinement on LgridDom.

4. NperRefCell: Number of particles triggering a refinement on refinement grids.

5. VescTune: If during the process particles have speeds that exceed: $v > V_{\text{escTune}}*$ v_{esc}, then they are considered unbound.

6. RhoVir: The halo edge is defined with the following equation:

$$\frac{M(< R_{\text{halo}})}{\frac{4\pi}{3} R_{\text{halo}}^3} = \Delta \rho_{\text{ref}} \tag{2.1}$$

   Then this integer parameter defines what to use as ρ_{ref}:

   • RhoVir = 0 use $\rho_{\text{vir}} = \rho_{\text{crit}}(z)$

   • RhoVir = 1 use $\rho_{\text{vir}} = \rho_{\text{back}}(z)$

   For the Zel'dovich simulations, RhoVir was chosen to be 1 as opposed to 0 due to technical difficulties with finding halos when using the critical density.

7. Dvir: based on Eq 2.1, Dvir defines what to use as Δ.

8. MaxGatherRad: limits the distance within which particles are collected.

9. LevelDomainDecomp: sets the grid to be used to do the domain decomposition.

10. NcpuReading: number of CPUs reading the data.

AHF then outputs a catalogue of halos, which contains a host of important and useful information. These catalogues are then accessed using `pynbody` so that analysis can be carried out.

2.6 Halo Mass Function

The HMF was computed by loading the snapshot and extracting the masses. Then the masses of each halo were calculated by counting the number of particles present, n, and multiplying by the mass of one particle, m, since all particles share the same mass. Hence the halo mass was calculated:

$$M_{\text{halo}}(\text{M}_\odot) = nm * \frac{10^{10}}{h} \tag{2.2}$$

where the mass is normalised to retain solar mass units.

Then, halos are distributed in logarithmic bins according to their mass, ranging from the smallest halo to the largest into 15 equally spaced bins. The halo mass function is calculated by taking this distribution and dividing it by the volume of the snapshot. This yields the $dn/dlogM$. Figure 2.5 shows the difference in HMFs between theoretical (line), full and three different Zel'dovich simulations, each with varying definitions for virial density.

The noise associated with each mass bin is also calculated and is Poissonian. For each count, the lower and upper bounds of the confidence interval are calculated using the inverse chi-squared distribution function `chi2.ppf` at the 5% and 95% levels, based on the assumption that the counts follow a Poisson distribution. Then the lower and upper bounds are normalized by dividing by the volume.

This process quantifies the uncertainty associated with the observed counts of halos in each mass bin, as found by AHF.

2.7 Loading and comparing the halos

The halos identified by Amiga's Halo Finder (AHF) are imported into a Jupyter Notebook for analysis. These files contain extensive data, including identification numbers or particle IDs corresponding to the halo's centre position, allowing for comparisons with the standard forward simulation.

First, given the halo catalogue, the centre of mass positions are calculated by taking the mean position of all particles making up a halo. Then, a loop checks each centre of mass and looks for matches between the Zel'dovich and full simulations, with some
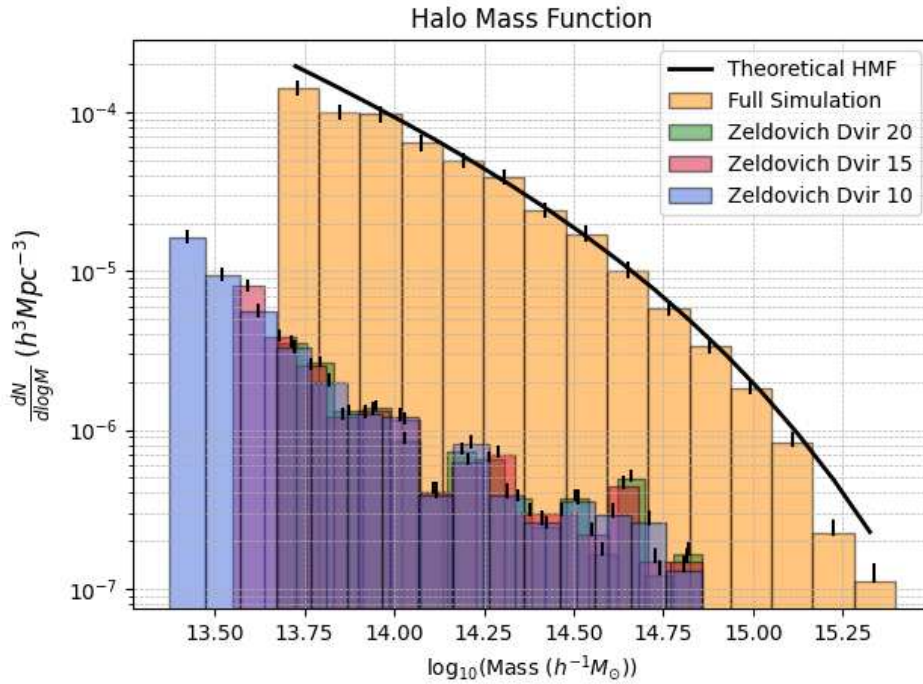
Figure 2.5: Comparison of different HMFs by varying the virial overdensity $D_{\mathrm{vir}} = \Delta_{\mathrm{vir}} \times \rho_{\mathrm{crit}}$ from 20 to 15 to 10, with associated Poissonian noise.

tolerance both in the centre of mass position and the actual mass. The tolerance is as follows:

```
if (comf[j, :] < com[i, :]+10).all() and
   (comf[j, :] > com[i, :]-10).all() and
   math.isclose(massZ, massF, abs_tol= 3* massZ):
```

Where `com` is the centre of mass of the full or Zel'dovich halos, `massZ` are the masses of each halo and `massF` are the masses of each halo in the full simulation. The tolerance therefore allows for ± 10 Mpc in each direction (x, y, z) between halos in the full simulation and Zel'dovich-generated halos as well as a total of three times the mass of the Zel'dovich halos. This is to allow for the fact that the halos in the Zel'dovich simulation have, on average, 3 to 3.5 times less mass than the Full N-body simulation (as seen in the HMF plot in Figure 2.6). As to why this occurs is explored in the conclusions.

Halo matches are collected and plotted on a single subplot (Figure 2.6). The x-axis represents spatial coordinates in megaparsecs (Mpc), while the y-axis represents similar spatial coordinates, creating a "2D slice". The number of matches is then recorded while discrepancies are noted.

Furthermore, we can analyse the similarity between Zel'dovich and full N-body halos to view the distribution of particle IDs within the halos themselves. To this end, we
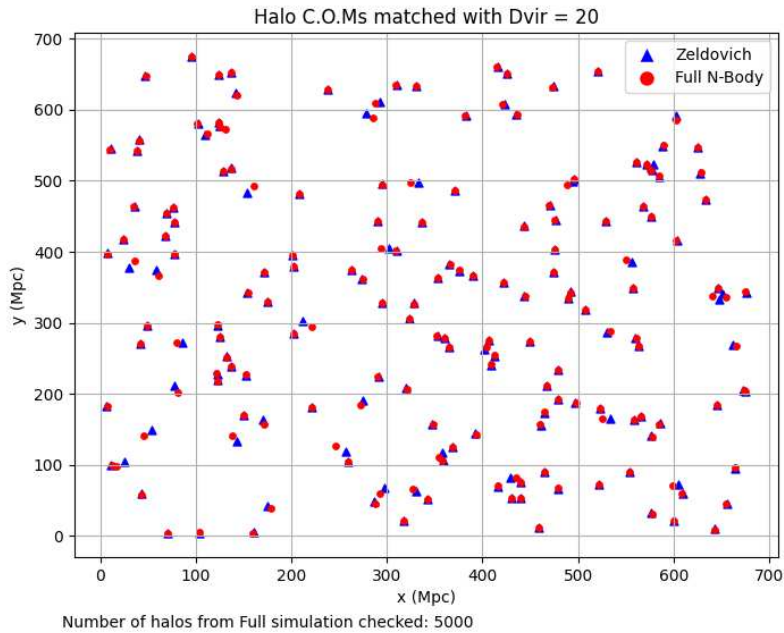
Figure 2.6: Positions of the centre of masses of the Zel'dovich (blue) simulation and the full N-body (red) simulation. **177** halos are found.

iterate over each pair of corresponding halos, retrieve the particle IDs for each and count the overlapping IDs. Then, we filter out halos with no overlapping IDs.

The ratio of overlapping particle counts is calculated as well as the total mass so that it can be normalised. The normalised mass is then plotted against the number of matching particle IDs between the halos. Figure 2.7 shows this.

2.7.1   Finding Voids

Once the halos are found a process of bridging, stacking and density profiling begins:

1. Using `pynbody`'s routine *bridge*, we create a link between the particles in the reverse simulation with those in the forward simulation, purely through the particles' ID.

2. The halos detected are bridged and the position of constituent particles are found as well as their centre of mass calculated. This "finds" the void because the location of the void in the normal simulation is approximately the location of the centre of mass of the halo in the inverted simulation.

3. For each position, we check the absolute value between particle positions and the centre of mass. If this exceeds a chosen threshold, the void is considered to be "cut" and a new position/centre of mass must be calculated.

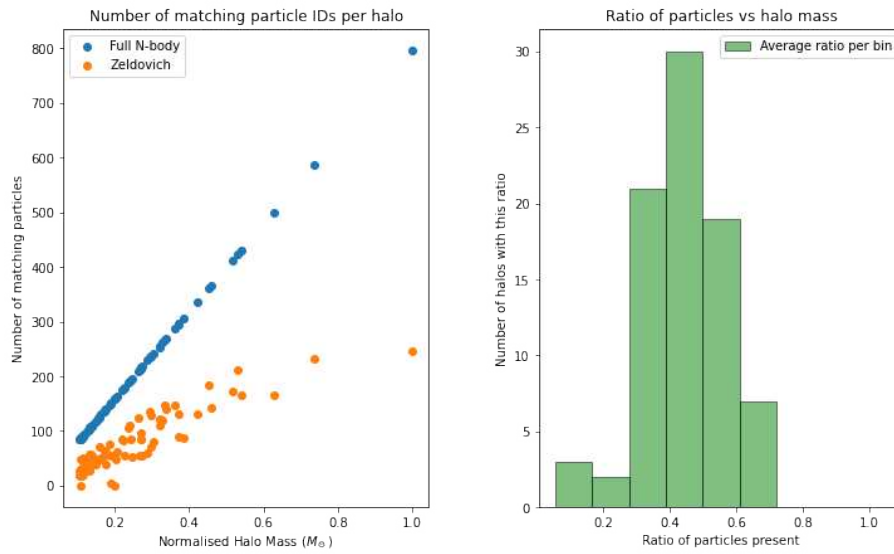4. The voids are then plotted and can be seen in Figure 2.8.

Figure 2.7: Relationship between halo mass and the ratio of particles present between Zeldovich and full N-body (left) and the number of halos with the shown ratio of particles present (right). On average, Zel'dovich halos contained only $\tilde{5}0\%$ of particles that the matching full N-body halo was made up of.
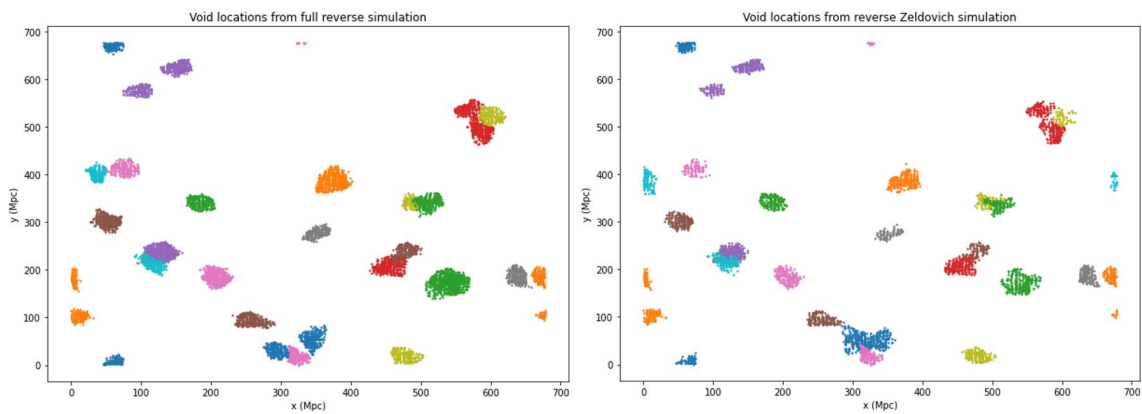

Figure 2.8: Comparison between the voids found by mapping the halos from the full reverse simulation (left) and the Zel'dovich reverse simulation (right).

2.8 Void Density Profile

The void density profile, as mentioned before, is a very important tool to be able to discern the capability of the Zel'dovich approximation in being able to accurately simulate the evolution and growth of a void. The density profile of voids is done by stacking them on top of one another - so it is a comprehensive analysis of the simulation.

First, we iterated over each halo selected that was matched between the Zel'dovich and full N-body simulations.

For each halo, the constituent particles are extracted, their IDs converted and matched onto the same particles in the corresponding reverse simulation, making use of the `bridge` function once again.

Then, the corresponding centre of mass of the *voids* is calculated by simply taking the mean of the position of the constituent particles, similarly to how the centre of mass was calculated for halos previously.

From the centre of mass, the volume and radii of the voids are calculated by measuring the total mass and density of the particles within the halo. Here, particles are removed from the void if they exceed a chosen tolerance.

Following the calculation of void properties, it can then be divided into several bins (31) to sample the density profile at different radial distances from the void centre.

For each bin, a function sums the masses of constituent particles within a spherical region of that radius centred around the void centre. Naturally, the volume of each spherical region and the density of particles within can now be calculated.

The average density is finally calculated for each bin across all selected voids and is plotted and the result is shown in Figure 2.9.
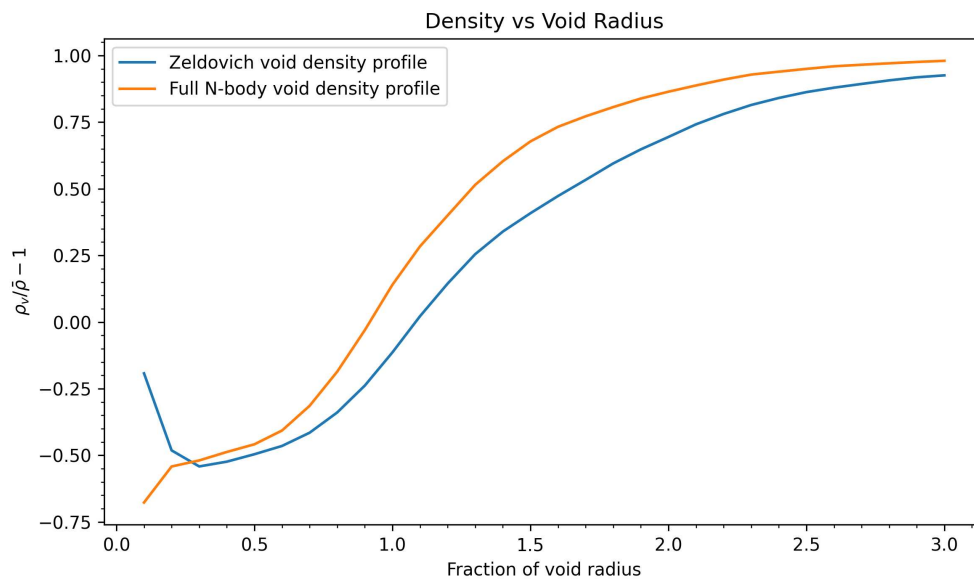
Figure 2.9: In this figure, 70 of the largest voids out of the 123 initially found voids are plotted. The filter took away any voids that originated from mismatched halos in the inverted simulation. The density profile is smooth as it has been averaged over all 70 halos.

Chapter 3

Results and Discussion

In this section, we discuss the results of the graphs, plots and other statistics that were presented in the methodology, as well as discuss the reasons why such results were obtained. Scientific insights and future improvements are briefly covered in Chapter 4.

3.1 Matter Density Plot Results

Given that we aim to evaluate the fidelity of the Zel'dovich approximation in predicting void behaviour on large scales, particularly assessing the linearity of void evolution compared to theoretical examples, our analysis of matter density plots yields promising, albeit coarse, insights. As depicted in Figure 2.2, the density distribution at z = 0 in the Zel'dovich simulation appears more uniform, with an average dark matter density of approximately 0.2g cm^{-3} compared to the full simulation, which has an average dark matter density of 0.01g cm^{-3}. This means that nodes of high density are much denser and voids are much emptier in the full simulation compared to the Zel'dovich simulation. This is to be expected, of course, because the Zel'dovich approximation struggles to fully capture the gravitational interactions of particles.

However, most large-scale structures such as massive halos and voids are still evident. While the inability of the Zel'dovich simulation to accurately capture small-scale structures may pose limitations when studying the evolution of singular structures, it does not significantly impact our current analysis when it comes to studying the linear evolution of voids, which is one of the main objectives of this thesis.

3.2 Power Spectrum Analysis

The examination of the power spectrum offers valuable insights into the structure formation processes in the universe. In this analysis, the efficacy of the Zel'dovich

approximation is briefly explored in reproducing the power spectrum, particularly concerning its inability to capture smaller-scale structures such as local voids.

Our investigation reveals that while the Zel'dovich approximation provides a useful framework for understanding the overall dynamics of linear structure formation, it exhibits limitations in accurately representing smaller-scale structures. These limitations stem from the simplifications inherent in the Zel'dovich approximation, which neglects complex gravitational interactions among particles, especially over large spatial scales. Consequently, discrepancies arise between the power spectrum predicted by the Zel'dovich approximation and that derived from more detailed simulations (see Figure 2.3).

In the above figure, it is clear that the Zel'dovich approximation ceases to be accurate at around $k = 0.1 h Mpc^{-1}$, which corresponds to a scale of approximately $20 h^{-1} Mpc$ from the relation $\lambda = 2\pi/k$, corresponding to relatively large galaxy clusters, but still well within the scale of larger voids (at least $40 h^{-1} Mpc$). Hence, while the approximation may be unsuitable in reconstructing our local group, it can still provide insights into the larger picture of void structures - especially up until the end of the linear regime.

3.3   Autocorrelation Function

The autocorrelation function, as mentioned in Section 1.11, is a measure of the clustering of matter across different spatial scales. Figures 2.4 and 3.1 show an expected result: The Zel'dovich approximation can reproduce the expected clustering of matter at large scales (>100 Mpc) but it fails to reproduce the accuracy of smaller-scale structures (<20 Mpc).

Figure 3.1 shows that the Zel'dovich approximation's clustering of matter at small scales is relatively linear with varying separation distance, as expected. This is due to the simple ballistic trajectories particles undertake when simulated only using first-order perturbation theory.

This linearisation of the autocorrelation function appears at all separation distances. When zooming in to a separation distance r of $100 < r < 150$ (Figure 3.2) we see that the structure of clustering of matter within this range is relatively well-approximated by the Zel'dovich approximation. While the approximation is losing the finer details, it is very capable of following the average clustering properties of the large-scale universe.

The validity of whether the Zel'dovich approximation can be used for large-scale simulations in determining the linear evolution of voids and therefore be able to extract cosmological parameters to better understand our universe hinges on its ability to reproduce observable phenomena, such as the autocorrelation function. Comparing these results to full N-body simulations is of critical importance and
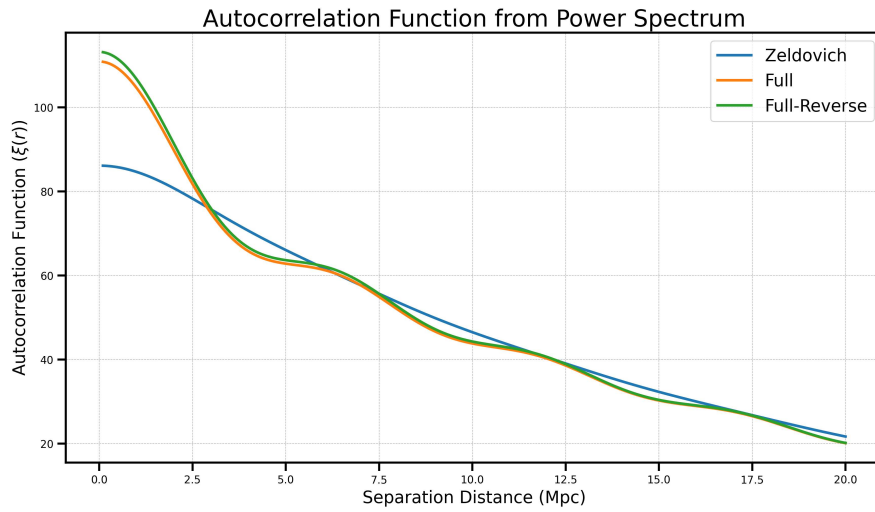
Figure 3.1: Comparison between different autocorrelation functions calculated using the Fourier transform of the previous power spectra and focusing on separation distances that correlate with large-scale structures ranging from 0 - 20 Mpc.
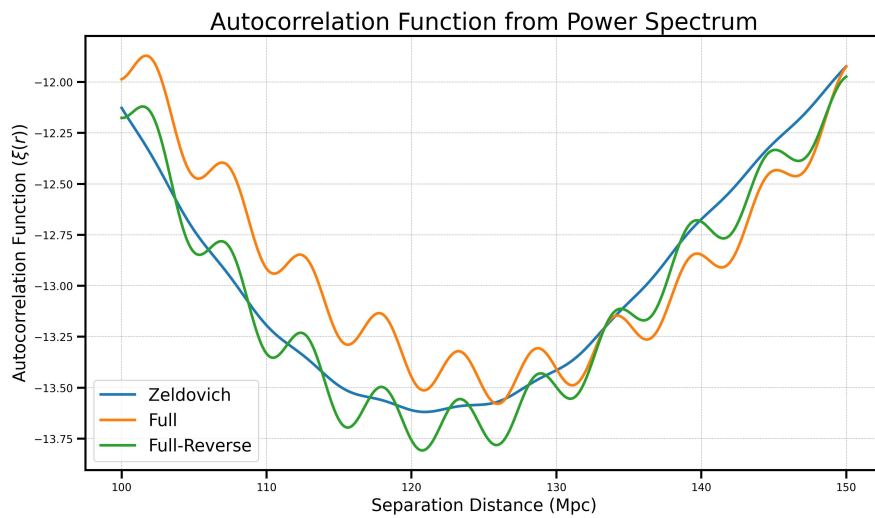

Figure 3.2: Zoom of the autocorrelation function showcasing Zel'dovich linearisation.

this should also be compared to physical data - which this work had very little access to, unfortunately.

3.4 Halo Mass Function Analysis

Three reverse Zel'dovich simulations with varying virial overdensity (D_{vir}) were analyzed and their halo mass functions were calculated alongside a fourth, full-forward simulation. Figure 2.5 shows the results and some interesting conclusions can be drawn:

1. The HMF shape changes as D_{vir} is varied. Lowering it to 10 results in a higher $\frac{dN}{dlogM}$ (HMF) compared to $D_{\text{vir}} = 20$ for a broader range of halo masses. This indicates that the choice of D_{vir} has a significant effect on the abundance of halos across different mass scales.

2. AHF with $D_{\text{vir}} = 10$ detects more halos across a wider range of masses compared to $D_{\text{vir}} = 20$. This suggests that lowering D_{vir} increases the sensitivity of the halo finder, allowing it to detect smaller halos that may have been excluded when using a higher threshold.

3. Despite the differences in the abundance of low-mass halos, the HMFs for $D_{\text{vir}} = 10$ and $D_{\text{vir}} = 20$ show similar trends for larger halo masses. This suggests that the impact of D_{vir} variations is more pronounced at smaller mass scales.

4. The emergence of smaller halos (less mass) in the Zel'dovich simulations compared to the normal simulation could be a consequence of the simplified nature of the approximation and its inability to capture all the relevant physics involved in halo formation.

Figure 2.7 shows an interesting conclusion and opens up a discussion for future work, which will be covered in the next section. None of the matching halos that were identified in the halo matching section contained above 80% of the intended particles. By intended particles we mean the particles that make up a full N-body halo. The Zel'dovich halos were all at least 20% emptier when compared to the full simulation counterparts, suggesting that many of the particles destined to populate a halo were not sufficiently affected by the growing density perturbations calculated within the Zel'dovich approximation, but instead continued along their trajectories without being captured by halos and possibly polluting other nearby halos. This is a severe limitation of the Zel'dovich approximation and hints towards its breaking point, particularly when caustics are involved. Furthermore, it was observed that smaller halo masses exhibited fewer missing particles, indicating a correlation between halo size and the extent of particle loss. Larger halos exhibit stronger gravitational pull towards new prospective particles, but this phenomenon is not present in the Zel'dovich approximation simulation. Hence, smaller halos are more closely simulated by just assuming ballistic trajectories.

It is interesting to see from Figure 2.6 that not all halos are completely matched, and make use of the tolerances discussed previously. This may be due to the applied tolerances themselves, the physical variability of halos (since they could exhibit a wide range of physical properties such as mass, shape and behaviour), simulation artefacts and possible sampling effects. This highlights the need for careful consideration of simulation parameters and methodologies in the future.

3.5 Void Profile Conclusions

Based on Figure 2.9, the Zel'dovich approximation struggles to simulate the very centre of the voids. By analysing the centre of halos and voids generated by the approximation, we were able to view an interesting conclusion: particles seem to remain in a grid pattern around the centre (at a radius of less than 0.3 $r/\bar{r}_v$, where $\bar{r}_v$ is the average void radius) rather than the expected behaviour of dispersing. It is presently unknown as to why there is an increase in density towards the centre, although it is assumed it is either noise or some other spurious behaviour.

The Figure also shows a relatively promising conclusion: the Zel'dovich approximation is fairly reliable when it comes to constructing the void linearly, by simply considering the ballistic trajectories of particles combined with the growing effect of the density fluctuations surrounding them. It is promising to see that the profiles are not very different, but it should be noted that any cosmological parameter constraint should be avoided when using the Zel'dovich approximation, as it lacks fine detail. Void evolution is now understood non-linearly, so the approximation should be used for linear studies only.

We also see that both simulations correctly show a negative relative density within the void radius, and then converge to the average radius outside the void. When compared to Figure 4.1, we can see some similarities in both shape and pattern, but this examination is left for the conclusion and improvements.

One interesting remark however is that, in the work cited above, the relative density of the void never exceeds 0.6, while the density profile in this thesis tends to converge up to 1.0. This may be due to some erroneous calculations.

Chapter 4

Conclusions

In this section, I briefly discuss some important improvements that could be made to this work given more time and resources as well as provide an overall conclusion to the work.

4.1 Future work and improvements

4.1.1 Initial Conditions

The initial conditions of a simulation are important ingredients that shape the overall physics and statistics of the work that is done on it. While theoretical models for the evolution of particles in a simulation may seem more important, the initial conditions of said simulation are as consequential to the final picture. Indeed, the statistical distribution of primordial fluctuations alone completely governs the overarching final distribution of large-scale structures of our universe. These primordial fluctuations, in turn, are completely determined by the cosmological parameters provided by the user. Given more time, multiple simulations would have been run that were able to simulate varying hypothetical models of the universe - especially those whose large-scale structures can be easily analysed. It would be interesting to see what difference the initial redshift (Table 2.1) would cause if it were to be varied.

Furthermore, since this work introduces the concept of simulation pairs wherein the initial conditions are simply reversed, it may be scientifically interesting to apply other geometric differences to the initial conditions and simulate three or 4 different versions of the same simulation to better understand the consequences of varying particle positions, velocities, masses, etc.

4.1.2 Matter Density Plots

Improvements in the matter density analysis include the use of contour levels to highlight regions of varying density levels and changing the resolutions to analyse

in more detail the smaller-scale structures. This would aid us in understanding the limitations of the approximation. A three-dimensional plot may also be visually stimulating.

4.1.3 Power Spectrum and Auotocorrelation

To bolster the argument that the Zel'dovich approximation can be used for large-scale analysis, further analytical scrutiny can be applied. This may involve conducting statistical analyses, such as comparing the correlation functions or clustering statistics derived from the Zel'dovich approximation to those obtained from full simulations. Additionally, investigating the behaviour of specific large-scale structures, such as galaxy clusters or cosmic voids, and comparing their properties between the Zel'dovich approximation and full simulations can provide valuable insights into the limitations of the approximation. Such analyses can strengthen the validity of our findings and deepen our understanding of the Zel'dovich approximation's capabilities and shortcomings in modelling large-scale structure formation.

Based on findings in Figures 2.4, 3.1 and 3.2, I propose that there could be a few significant improvements to the quality of the Zel'dovich approximation. If one would be able to use a similar technique to adaptive mesh refinement used in AHF, whereby the Zel'dovich approximation is used throughout a simulation, but higher-order perturbation theory is applied to regions of interest, then this *hybrid* approach may significantly increase the statistical significance of such a simulation while retaining high degrees of efficiency.

4.1.4 Halo Finding Problems

The process of finding halos in the simulations was carried out using the algorithm Amiga's Halo Finder (AHF), which as mentioned before uses a combination of algorithms to accurately discern whether a collection of particles (dark matter only in this case) constitute a halo or not. Several parameters can be altered to improve or change the selection criteria, thereby altering the final catalogue of halos that AHF returns to the user. "Exploring various parameters, including the virial overdensity D_{vir}, has allowed for an assessment of the Zel'dovich approximation's capability in generating voids. However, considering other parameters in both the Zel'dovich approximation and the Full N-body simulation, when applying AHF, could offer further insights. How would adjusting these parameters affect the outcomes, and how would they compare between the two methods?

4.1.5 Halo Mass Function Improvements

A few improvements could be made to the plot shown in Figure 1.9. By varying parameters other than D_{vir}, some useful conclusions could be made. It may be possible to explore the parameter space using numerical simulations to constrain the

discrepancy between the Zel'dovich approximation and the full N-body simulation, for example.

By increasing the volume of the simulation, I believe the Halo Mass Function for the simulations using the Zel'dovich approximation may more closely resemble the overall shape and statistics of the Halo Mass Function of the full N-body simulation.

It would be interesting to repeat the experiment using a hybrid approach, as mentioned in section 4.1.3, to observe the potential of combining Zel'dovich with an adaptive higher-order perturbation theory.

Furthermore, it may be interesting to inspect how poorly the Zel'dovich approximation would perform when calculating a theoretical non-spherical halo mass function that is non-gaussian in evolution, as explored in Achitouv and Corasaniti (2011). Indeed, I suspect that the approximation will do poorly in simulating such halos and therefore not be a suitable candidate for constraining any sort of important parameter.

Finally, it would be interesting to incorporate baryonic effects, such as gas cooling, star formation, and feedback processes, in simulations that primarily employ the Zel'dovich approximation. While smoothed particle hydrodynamics would be required for such simulations, at least the evolution of dark matter particles could be interpreted with the Zel'dovich approximation, reducing the required CPU hours to achieve a z=0 simulation from high redshift.

4.1.6 Void Finding and Density Profile Improvements

Detecting the voids from particle IDs has proven to be an effective method, meaning that running pairs of simulations with inverted conditions is an effective technique for finding voids. While voids have no real theoretical definition, this technique proves to be quite successful and the Zel'dovich approximation successfully follows the full N-body simulation when it comes to mapping the halos that were generated onto the forward simulation to locate voids. Figure 2.8 shows a promising correlation between the two, but some improvements can be made:

- The threshold values used to define void boundaries can be refined and re-evaluated.

- The accuracy of volume calculations of voids can be improved by implementing more sophisticated methods, such as using the volume-weighted centre rather than the centre of mass of a void, as it is less susceptible to fluctuations

- The implementation of KD-tree data structures should be considered, as it is a more rapid neighbour search tool instead of iterating through all particles within each radius, which can drastically reduce computation time and improve performance.
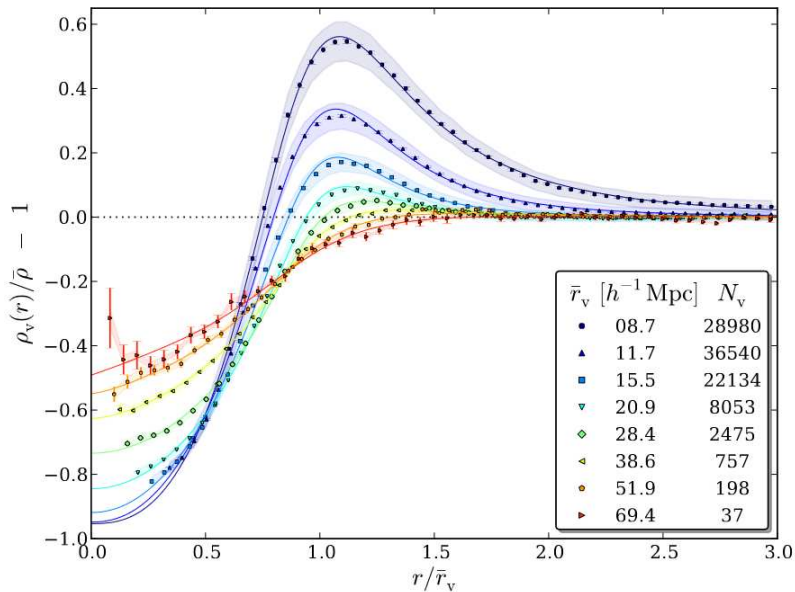
Figure 4.1: Adopted from Hamaus et al. (2014), showing the density profiles for voids in eight different radial bins. It includes mean values, standard deviations and standard errors on the mean profile $\sigma/\sqrt{N_v}$.

- The density profile calculations could be refined by applying smoothing techniques or adjusting the binning strategy to capture finer details.

- Error calculation in the void profile should be included - Poissonian error should first be considered if the counting of particles is employed.

- Finally, validation against observable data is highly needed.

The density profile of voids presented in this work shows interesting results, albeit lacking some standard important points. Statistics such as error bars and standard deviations would be included in the plots if more time was allocated, and comparison to other works could be explored. For example, Figure 2.8 adopted from Hamaus et al. (2014) shows the stacked void densities for varying radii. Interestingly, their largest halos also exhibit an increase in density towards the very centre of the void. Their y-axis is different to the work presented in this thesis, but it can be shown that the shapes are very similar and that, therefore, the Zel'dovich approximation can indeed reconstruct at least the basic shape of a void, especially larger ones and for a larger sample

4.2 Conclusion

In conclusion, our investigation into the comparative analysis of the Zel'dovich approximation (ZA) and full N-body simulations has provided significant insights into the capabilities and limitations of the ZA in modelling the linear structure of voids.

While the ZA demonstrates remarkable accuracy in reproducing the overall distribution of filaments, halos, and voids, it falls short in displaying small-scale matter densities, as seen by the obvious differences observed in matter density plots.

The analysis of the power spectrum and autocorrelation function also emphasizes the strengths and weaknesses of the ZA, especially in its ability to follow the distribution of matter density across varying spatial scales. Notably, the smoothness observed in the autocorrelation function of the ZA, compared to the undulating nature of the full N-body simulations, highlights ZA's simplicity, linearity and the diminishing of its accuracy when it comes to smaller scales, where non-linear gravitational interactions are dominant.

Our examination of the halo mass function revealed intriguing discrepancies between the ZA and full simulation, which shows the challenges in reproducing halo masses and distributions accurately. Despite these difficulties, the fundamental shape of the halo mass function is preserved, suggesting potential solutions for further refinement.

Additionally, our investigation into the void density profile unveiled interesting differences in the ZA's representation of void centres. The inaccuracies at the centre of voids are evident, but overall, the ZA still manages to follow the full N-body void density profile.
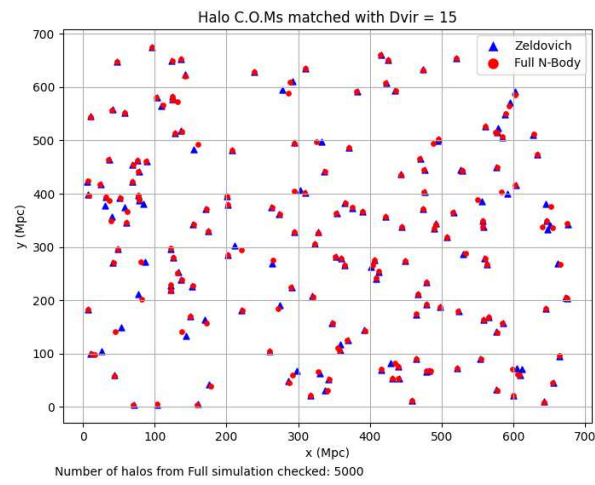
.1 Extra Plots

Figure 2: Positions of halos of the Zeldovich (blue) simulation using $D_{\mathrm{vir}} = 15$ and the full N-body (red) simulation. **187** halos are found.
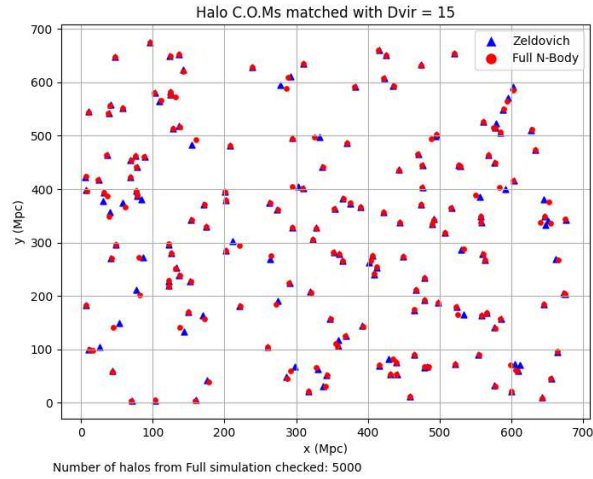
Figure 3: Positions of halos of the Zeldovich (blue) simulation using $D_{\mathrm{vir}} = 10$ and the full N-body (red) simulation. **220** halos are found.
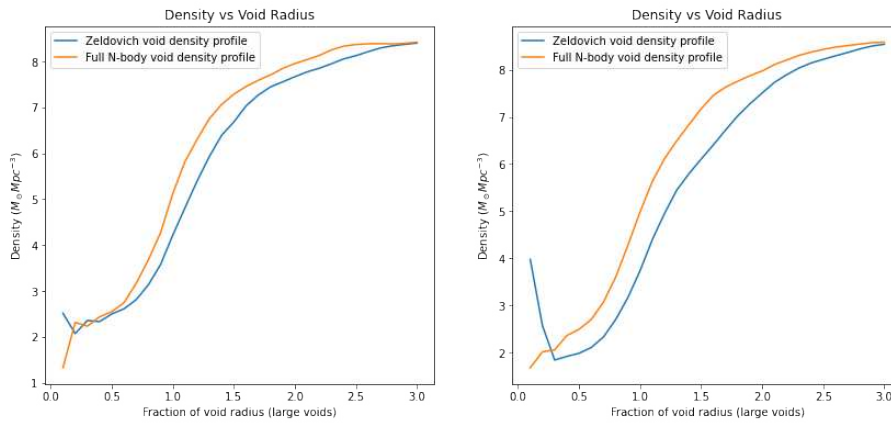

Figure 4: Here we are comparing the 40 largest halos (left) vs the remaining 83 halos (right).

Bibliography

Achitouv, I. and Corasaniti, P. S. (2011). Non-gaussian halo mass function and non-spherical halo collapse: Theory vs. simulations. *JCAP 02 (2012) 002*, 2012(02):002–002.

Beygu, B., Kreckel, K., van der Hulst, J. M., Jarrett, T. H., Peletier, R., van de Weygaert, R., van Gorkom, J. H., and Aragon-Calvo, M. A. (2016). The void galaxy survey: Star formation properties. *Monthly Notices of the Royal Astronomical Society*, 458(1):394–409.

Brooks, A. ([2021]). [from lss to a halo mass function]. PowerPoint presentation.

Collaboration, P. (2018). Planck 2018 results. i. overview and the cosmological legacy of planck. *A&A 641, A1 (2020)*, 641:A1.

Correa, C. M. (2022). Cosmic voids as cosmological laboratories.

Courtois, H. M., van de Weygaert, R., Aubert, M., Pomarède, D., Guinet, D., Domínguez-Gómez, J., Florido, E., Galbany, L., García-Benito, R., van der Hulst, J. M., Kreckel, K., Miura, R. E., Pérez, I., Planelles, S., Quilis, V., Román, J., and Sánchez-Portal, M. (2023). Sociology and hierarchy of voids: A study of seven nearby cavity galaxy voids and their dynamical cosmicflows-3 environment. *Astronomy and Astrophysics*, 673:A38.

Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. (1985). The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, 292:371.

Dehnen, W. and Read, J. (2011). N-body simulations of gravitational dynamics. *2011, Eur. Phys. J. Plus 126, 55*.

Garrison, L. H., Eisenstein, D. J., Ferrer, D., Metchnik, M. V., and Pinto, P. A. (2016). Improving initial conditions for cosmological n-body simulations. *MNRAS (October 01, 2016) 461 (4): 4125-4145*.

Gill, S. P. D., Knebe, A., and Gibson, B. K. (2004). The evolution of substructure — i. a new identification method. *Monthly Notices of the Royal Astronomical Society*, 351(2):399–409.

Goon, G. (2021). Gaussian fields.

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

Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356.

Hamaus, N., Sutter, P. M., and Wandelt, B. D. (2014). Universal density profile for cosmic voids. *Phys. Rev. Lett. 112, 251302 (2014)*, 112(25):251302.

Heitmann, K., Lukic, Z., Fasel, P., Habib, S., Warren, M. S., White, M., Ahrens, J., Ankeny, L., Armstrong, R., O'Shea, B., Ricker, P. M., Springel, V., Stadel, J., and Trac, H. (2007). The cosmic code comparison project. *Comput.Sci.Dis.1:015003,2008*, 1(1):015003.

IllustrisTNG (2018). Simulation details.

Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., and Yoshida, N. (2001). The mass function of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 321(2):372–384.

Knollmann, S. R. and Knebe, A. (2009). AHF: AMIGA's HALO FINDER. *The Astrophysical Journal Supplement Series*, 182(2):608–624.

Kravtsov, A. and Borgani, S. (2012). Formation of galaxy clusters. *Annual Review of Astronomy and Astrophysics*, 50(1):353–409.

Marcos, B., Baertschiger, T., Joyce, M., Gabrielli, A., and Labini, F. S. (2006). Linear perturbative theory of the discrete cosmological. *Physical Review D*, 73(10):103507.

Navarro, J. F., Frenk, C. S., and White, S. D. M. (1996). A universal density profile from hierarchical clustering. *Astrophys.J.490:493-508,1997*, 490(2):493–508.

Peebles, P. J. E. (1980). *The Large-Scale Structure of the Universe*. Princeton University Press.

Pontzen, A., Slosar, A., Roth, N., and Peiris, H. V. (2015). Inverted initial conditions: exploring the growth of cosmic structure and voids. *Phys. Rev. D 93, 103519 (2016)*, 93(10):103519.

Shandarin, S. F. (2009). Tessellating the universe: the zel'dovich and adhesion tiling of space. *Astrophys.J.490:493-508,1997*.

Shandarin, S. F. and Zeldovich, Y. B. (1989). The large-scale structure of the universe: Turbulence, intermittency, structures in a self-gravitating medium. *Rev. Mod. Phys.*, 61:185–220.

Skory, S. (2010). Halos and halo finding in yt.

Springel, V. (2005). The cosmological simulation code gadget-2. *Mon.Not.Roy.Astron.Soc. 364 (2005) 1105-1134*, 364(4):1105–1134.

Stopyra, S., Peiris, H. V., Pontzen, A., Jasche, J., and Natarajan, P. (2021a). Quantifying the rarity of the local super-volume. *The Astrophysical Journal Supplement Series.*

Stopyra, S., Pontzen, A., Peiris, H., Roth, N., and Rey, M. P. (2021b). GenetIC—a new initial conditions generator to support genetically modified zoom simulations. *The Astrophysical Journal Supplement Series*, 252(2):28.

Tinker, J. L., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M. S., Yepes, G., Gottlober, S., and Holz, D. E. (2008). Toward a halo mass function for precision cosmology: the limits of universality. *Astrophys.J.688:709-728,2008.*

White, M. (2001). The mass of a halo. *Astronomy & Astrophysics*, 367(1):27–32.

Zel'dovich, Y. B. (1970). Gravitational instability: An approximate theory for large density perturbations. *aap*, 5:84–89.