

Aspects of gravitational clustering and structure formation in the Universe

Swati Gavas

*A thesis submitted for the partial fulfillment of
the degree of Doctor of Philosophy*



Department of Physical Sciences
Indian Institute of Science Education and Research Mohali
Knowledge City, Sector 81, SAS Nagar, Manauli PO, Mohali 140306, Punjab, India.

December 2024

Declaration

The work presented in this thesis has been carried out by me under the guidance of Prof. Jasjeet Singh Bagla and Dr. Nishikanta Khandai (NISER Bhubaneswar) at the Indian Institute of Science Education and Research Mohali. This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bona fide record of original work done by me and all sources listed within have been detailed in the bibliography.

Swati Gavas

In my capacity as the supervisor of the candidate's thesis work, I certify that the above statements by the candidate are true to the best of my knowledge.

Prof. Jasjeet Singh Bagla (Supervisor)

Dr. Nishikanta Khandai (Co-supervisor)

ગુરુજનાંપ્રતિ

Acknowledgements

I sincerely thank Prof. Jasjeet Singh Bagla, my primary supervisor, for his guidance and for providing a valuable learning experience. I am grateful to Dr. Nishikanta Khandai, my co-supervisor, for his crucial feedback and encouragement. I extend my gratitude to Dr. Girish Kulkarni for his suggestions and support as a collaborator. I also thank the anonymous referees for their thorough reviews of my thesis.

I thank my doctoral committee members, Dr. Harvinder Kaur Jassal, Dr. Kinjalk Lochan, Dr. Anosh Joseph, Dr. Ambresh Shivaji and Dr. Pankaj Kushwaha, for their time, constructive feedback, and smooth conduct of my PhD procedures.

I thank IISER Mohali for financial support and for providing research facilities. Thanks to NISER Bhubaneswar and their Computer Center for supporting me with a High-Performance Computing facility and related help. I am grateful to DST-ANRF and IISER Mohali for funding my international travels. Thanks to the hosts of the conferences, schools and workshops for providing an opportunity to present our work and travel support: IIT Indore, IISER Tirupati, KIAS Seoul, IAU, IIT Guwahati, IISc Bangalore, ICTP Trieste, IAP Paris, NCRA-TIFR Pune, ICTS Bangalore, Observatório Nacional, Rio.

I extend my sincere gratitude to the professors who shared their valuable feedback on our work during various events: Dr. Surhud More, Prof. Thanu Padmanabhan, Dr. Juhan Kim, Dr. Aseem Paranjape, Dr. Shadab Alam, Prof. Arif Babul, Prof. Ravi Sheth, Dr. Fabian Schmidt, Prof. Pierluigi Monaco, Prof. Stephane Colombi.

Throughout my PhD journey, I have been blessed with a wonderful and varied circle of friends and colleagues. I appreciate the Astro Journal Club members, Scholar roommates, batchmates and roommates for all the help and enriching activities we enjoyed together. Thanks are also due to the Gulati family for their kind help during the COVID-19 pandemic and beyond.

Finally, I value having a family.

Abstract

The distribution of galaxies, halo abundance, and peculiar velocities are influenced by non-linear gravitational interactions, making the study of non-linear evolution crucial for accurate cosmological predictions. Thus, it is essential to study the underlying assumptions in these processes to improve predictions like halo abundance and determination of the Hubble-Lemaître constant H_0 . We explore these aspects using N-body simulations. The halo model can be used to make various predictions and interpretation of observations, with the halo mass function as a core ingredient. The theoretical models of mass function can be formulated without referencing the cosmological model and input power spectrum. Mass functions obtained from N-body simulations show systematic deviations of 5-20% from theoretical predictions. The physical origin of deviations is complex to understand and may result from cosmology, the power spectrum, or both. To investigate the issue, we examine mass function deviations from universality for scale-free power spectra with an Einstein-de Sitter cosmology. We show that the mass function has explicit dependence on the slope of the input power spectrum. We extend our analysis to Λ CDM cosmologies and show that an effective index of the Λ CDM model can correspond to the mass function from scale free cosmologies as a first approximation. Our results indicate that an improved analytical theory is required to provide better fits to the mass function. Furthermore, structure formation has led to deviations from homogeneity and isotropy on scales up to at least 100 Mpc/h, expected to affect measurements of H_0 . Various efforts have been made to quantify this effect. In our second study, we revisit this issue of the concordance model. We find a correlation between errors in H_0 estimates and the density around the observer. Further, our mock observations reveal that deviations of up to 5% can occur in Milky Way-sized halos. While this finding alone does not fully resolve the Hubble tension, it may account for part of it. We rely on N-body simulations for these studies. Hence, it is essential to understand their limitations to avoid misinterpreting data. We show that the missing power at small scales introduces errors in the root-mean-square fluctuations and, consequently, in the simulated mass function. These errors are expected to diminish as the scale of non-linearity increases. Our analytical calculation indicates that mode coupling between small and large scales depends on resolving collapsed halos. Therefore, accurate mode coupling estimates require sufficient halos in the simulation.

Contents

| | |
|---|-------------|
| Abstract | vii |
| List of publications | xiii |
| 1 Introduction: Large Scale Structure | 1 |
| 1.1 Observing large scale structure | 1 |
| 1.2 Theory of large scale structures | 4 |
| 1.2.1 Homogeneous Universe | 4 |
| 1.2.2 Dynamics of gravitational instability | 6 |
| 1.2.3 Eulerian perturbation theory | 7 |
| 1.2.4 Lagrangian perturbation theory | 7 |
| 1.2.5 Spherical and ellipsoidal collapse | 9 |
| 1.2.6 Non-linear approximations | 10 |
| 1.3 Challenges in large scale structure studies | 11 |
| 1.3.1 Non-universality of halo mass function | 11 |
| 1.3.2 Determination of the Hubble-Lemaître constant | 12 |
| 1.3.3 Mass discreteness effect in simulations | 12 |
| 2 Structure formation toolkit | 13 |
| 2.1 Statistical description of the density field | 13 |
| 2.1.1 Average and ergodicity | 13 |
| 2.1.2 Gaussian random field (GRF) | 14 |
| 2.2 Measures of gravitational clustering | 15 |
| 2.2.1 n -point correlation functions | 15 |
| 2.2.2 Power spectrum | 16 |
| 2.2.3 Mass variance | 16 |
| 2.2.4 Halo mass function | 17 |
| 2.3 N-body simulations | 18 |
| 2.3.1 Equations of motion and Basic algorithm | 18 |
| 2.3.2 Nuances of cosmological N-body simulations | 19 |
| 2.3.3 Structure classification | 21 |

| | | |
|----------|--|-----------|
| 2.4 | Suite of simulations in this thesis | 22 |
| 2.4.1 | Scale-free cosmology | 23 |
| 2.4.2 | Scale-free simulation catalogue | 25 |
| 2.4.3 | Λ CDM simulation suite | 30 |
| 3 | Halo mass function in scale-invariant models | 33 |
| 3.1 | Methodology | 35 |
| 3.1.1 | Mass function | 35 |
| 3.1.2 | Simulations | 36 |
| 3.1.3 | Finding halos | 37 |
| 3.2 | Results | 37 |
| 3.2.1 | Mass function | 38 |
| 3.2.2 | Dependence of Sheth-Tormen parameters on power-law index and epoch | 41 |
| 3.2.3 | Tolerance to parameters | 43 |
| 3.3 | Going beyond power-law and EdS | 45 |
| 3.4 | Conclusions | 49 |
| 4 | Dispersion in the Hubble-Lemaître constant measurements from gravitational clustering | 51 |
| 4.1 | Methodology | 52 |
| 4.2 | Simulations | 53 |
| 4.3 | Results | 55 |
| 4.3.1 | The δ_H distributions | 55 |
| 4.3.2 | Statistics of distributions with distance | 56 |
| 4.3.3 | Correlation with local density | 57 |
| 4.3.4 | Milky Way-sized halos | 59 |
| 4.4 | Conclusion and Discussion | 60 |
| 5 | On the origin of transient features in cosmological N-body Simulations | 63 |
| 5.1 | Mode Coupling | 65 |
| 5.2 | Sources of errors at small scales | 69 |
| 5.3 | Discussion and Summary | 71 |
| 6 | Conclusion | 75 |
| A | Observables in the expanding Universe | 79 |
| B | Fourier transform | 81 |
| C | Mass function: supplementary analysis | 85 |

| | |
|--|----|
| D H_0 measurements: supplementary analysis | 91 |
|--|----|

| | |
|---------------------|----|
| E Simulations in §5 | 93 |
|---------------------|----|

List of publications

- Swati Gavas et al. “Halo mass function in scale invariant models”. In: *Monthly Notices of the Royal Astronomical Society* 521.4 (Mar. 2023), pp. 5960–5971. ISSN: 1365-2966
- [In communication]
Swati Gavas, J. S. Bagla, and Nishikanta Khandai. “Dispersion in the Hubble-Lemaître constant measurements from gravitational clustering”. In: (July 2024). arXiv: 2407 . 10139 [astro-ph.CO]
- [In communication]
J. S. Bagla and Swati Gavas. *On the origin of transient features in cosmological N-Body Simulations*. 2024. arXiv: 2408 . 05118 [astro-ph.CO]

List of Figures

| | | |
|-----|--|----|
| 1.1 | Historical overview of observing Large Scale Structure | 2 |
| 2.1 | Finite box size effect | 27 |
| 2.2 | Large scale structure in scale-free cosmologies | 28 |
| 2.3 | Demonstration of self-similarity in scale-free models | 29 |
| 2.4 | Large scale structure in Λ CDM cosmology | 31 |
| 3.1 | Mass function in scale-invariant models ($n = 0.0, -0.5, -1.0, -1.5$) | 39 |
| 3.2 | Mass function in scale-invariant models ($n = -1.8, -2.0, -2.2$) | 40 |
| 3.3 | Power spectrum n dependence of the Sheth-Tormen parameters | 43 |
| 3.4 | Epoch independence of the Sheth-Tormen parameters | 44 |
| 3.5 | Mass function: tolerance to parameters | 45 |
| 3.6 | Mass function for Λ CDM models | 46 |
| 3.7 | Sheth-Tormen parameters in Λ CDM cosmology | 47 |
| 3.8 | χ^2_{red} values for Λ CDM mass function fits | 47 |
| 4.1 | Local H_0 probability distribution functions | 55 |
| 4.2 | Statistics of H_0 distribution with distance | 56 |
| 4.3 | H_0 correlation with local overdensity | 58 |
| 4.4 | H_0 distributions for Milky way-sized halos | 59 |
| 5.1 | Effect of missing small scale power on root-mean-square fluctuations and mass function | 69 |
| 5.2 | Effect of missing small scale power for Λ CDM models | 70 |
| 5.3 | Errors in the simulation mass function | 72 |
| C.1 | Mass function: Effect of initial redshift | 85 |
| C.2 | Mass function: Effect of Box size | 86 |
| C.3 | Mass function: Poisson error | 87 |
| C.4 | Mass function: Jackknife error | 87 |
| C.5 | Mass function: Sample variance error | 88 |
| C.6 | Mass function: Effect of error-types | 88 |

| | | |
|-----|---|----|
| C.7 | Box-wise Λ CDM mass function | 89 |
| D.1 | H_0 PDF: Effect of substructure | 91 |
| D.2 | H_0 PDF: Convergence in the simulations | 92 |
| E.1 | Mass function: Effect of finite mass resolution | 94 |
| E.2 | Mass function: Effect of finite box size | 94 |

List of Tables

| | | |
|-----|---|----|
| 2.1 | Scale-free simulation catalogue | 26 |
| 2.2 | Λ CDM simulation suite | 32 |
| 3.1 | Halo mass function simulation setup | 36 |
| 4.1 | H_0 estimates: Simulation details | 54 |
| 4.2 | Standard deviation of PDF for Milky Way-sized halos | 60 |
| E.1 | Simulations in §5 | 93 |

1 Introduction: Large Scale Structure

The General theory of relativity describes gravity as the curvature of spacetime resulting from the presence of matter (or energy). This curvature, manifested as the force of gravity, causes matter elements to attract each other, leading to the condensation of matter and the formation of structures such as galaxies, galaxy clusters, and larger cosmic arrangements [4–10]. These structures, collectively known as the ‘Large Scale Structure (LSS) of the Universe’, encompass various features like halos, filaments, and voids. The study of LSS holds significant importance for several reasons, including understanding cosmic evolution [11–13], testing cosmological theories [14–16], gaining insights into the galactic dynamics [17–21], and cosmography [22–28]. LSS is integral to modern cosmology; its study involves theoretical modeling [9, 29–37], observational astronomy [38–46], and *simulations* [35, 47–51].

1.1 Observing large scale structure

Notable first attempts to map galaxy distribution can be traced back to Edwin Hubble’s examination of 400 ‘extragalactic nebulae’ in 1926 [52], revealing a general uniformity on large scales. However, further analysis using the larger *Shapley-Ames catalogue* in 1932 indicated *clumpiness* on smaller scales [53]. In 1934, with a larger statistical sample Hubble found an excess of galaxies within angular scales less than $\sim 10^\circ$ compared to a random Poisson distribution [54].

Harlow Shapley and Fritz Zwicky led the foundations of constructing galaxy catalogues using photographic plates in the 1930s-50s [55, 56]. The *Lick Galaxy Catalogue* (1967), produced by Shane & Wirtanen [38, 57], marked a significant advancement, containing data on approximately a million galaxies, revealing a non-uniform distribution in the sky [58], showing foam-like patterns with possible *walls and filaments* [59]. Using Hubble’s counts of faint galaxies, Gérard de Vaucouleurs calculated that the total light from all galaxies, including those fainter than counted, would contribute 0.3%, making it undetectable [60]. Further, analysis by Peebles and collaborators [61] indicated that the clustering amplitude decreases for fainter galaxy populations, likely due to larger projection effects along the line of sight, as faint galaxies tend to lie at larger distances. During the 1980s, the emergence of the cosmic web structure was observed through the catalogue of approximately 2 million galaxies generated by the *Automated Plate Measurement (APM) Galaxy Survey* [62]. It hinted at the presence

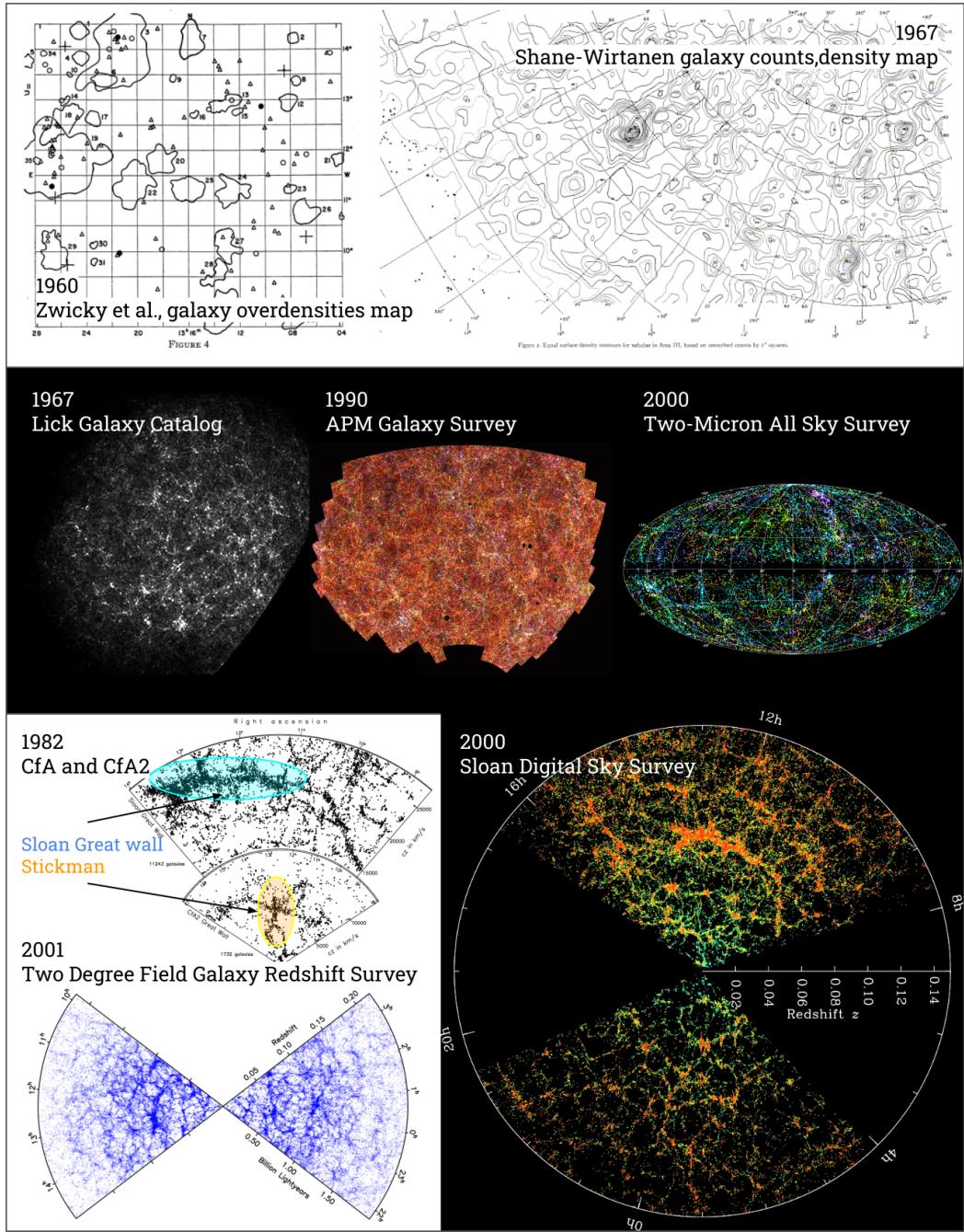


Figure 1.1: *Observing large scale structure*: During the 1930s-50s, Shapley and Zwicky manually mapped the galaxy overdensities using photographic plates. The Lick Galaxy Catalogue(\sim million galaxies) revealed the foamy structure of the cosmic web. The automation for reading photographic plates led to the APM galaxy survey (~ 2 million galaxies). The Two-Micron All Sky Survey aimed to capture galaxies in the infrared spectrum to attempt photometric redshifts. The CfA ($\sim 9,000$ galaxies, $\sim 9,000 \text{ deg}^2$) and CfA2 ($\sim 20,000$) were among the initial spectroscopic redshift surveys unveiled the famous ‘Stickman’ and ‘Sloan Great Wall’. Based on the APM data, the Two-Degree Field Galaxy Redshift Survey covered $\sim 220,000$ galaxies across $\sim 1500 \text{ deg}^2$. The largest galaxy redshift survey is the Sloan Digital Sky Survey, encompassing $\sim 9,30,000$ galaxies and covering $\sim 14,555 \text{ deg}^2$. Refer §1.1

of a cosmological constant for the first time. In the early 2000s, the *Two-Micron All Sky Survey* (2MASS) mapped the local Universe in the infrared regime [39]. It marked the initial attempt to obtain photometric redshifts of galaxies and provided clear evidence for the large scale structure.

These findings are taken further by the emergence of redshift surveys, which use optical spectra to assess redshifts and spatial distributions of galaxy samples. Pioneering work by Gregory & Thompson (1978) mapped the spatial distribution of 238 galaxies around the Coma/Abell 1367 supercluster, unveiling significant void ($> 20h^{-1}$ Mpc) in the foreground [63]. Further, Joeveer and collaborators [64, 65] utilised redshift data to map galaxies across the southern galactic hemisphere, revealing clustered formations, including chains of clusters, now identified as *filaments*. During a similar period, the *KOS* (*Kirshner, Oemler, Schechter*) survey measured redshifts focusing on 164 brighter galaxies in eight separate fields on the sky, revealing strong clustering in the velocity space [40].

Gérard de Vaucouleurs and collaborators constructed the earliest large scale galaxy redshift surveys in 1950-1970s [66, 67]. The first CfA (*Center for Astrophysics*) survey, completed in 1982, catalogued redshifts for $\sim 9,000$ galaxies across both galactic poles, covering ~ 9000 square degree, aiming to assess the three-dimensional clustering of galaxies [41]. It generated extensive three-dimensional maps of large scale structure, highlighting clusters, voids, and filamentary structures, prompting comparisons with N-body simulations. The survey revealed the famous ‘Stickman’ figure, which is the Coma cluster distorted in redshift space¹ (appendices A.2 and A.3). The second CfA redshift survey, spanning from 1985 to 1995, encompassed spectra for approximately $\sim 20,000$ galaxies, unveiling the wide ‘Sloan Great Wall’ supercluster spanning $170h^{-1}$ Mpc and highlighting widespread underdense voids with a density of about 20% of the mean density [68]. *Las Campanas Redshift Survey* (LCRS), conducted from 1991 to 1997, mapped galaxy redshifts over large area out to a redshift of $z = 0.2$ [69]. It is considered one of the first surveys where the ‘Cosmological Principle’ of isotropy could be tested. *Canada-France Redshift Survey* (CFRS), performed from 1989 to 1997, provided the first systematic study of galaxies at redshifts $z > 0.5$ [70].

In the last three decades, redshift surveys have advanced significantly due to multiobject spectrographs and larger telescopes, facilitating deep surveys of nearby and distant galaxies. In late 1990s and early 2000s, massive surveys, such as the *Two Degree Field Galaxy Redshift Survey* (2dFGRS) [42] and the *Sloan Digital Sky Survey* (SDSS) [43], provide comprehensive maps of large scale structure, uncovering a sponge-like pattern with voids of approximately 10 Mpc/h and intersecting filaments around galaxy groups and clusters. SDSS has evolved through multiple phases (I, II, III, IV, V), each focusing on different scientific goals. As we now transit to the precision cosmology era, recent Large Scale Structure (LSS) surveys like *Baryon Oscillation Spectroscopic Survey* (BOSS) [71] (it is an extension of SDSS), *Dark Energy*

¹Positions are mapped across the celestial sphere, while velocities are measured along the line of sight. Converting these velocities to distances can be distorted due to the peculiar motions of galaxies within clusters.

Spectroscopic Instrument (DESI) [44], *Large Synoptic Survey Telescope (LSST)* [45, 72] and *Euclid Space Telescope* [46, 73] are launched to focus of specific aspects of LSS. These surveys are helping us to constrain cosmological models.

Large scale structure can be explored through a variety of indirect observational probes, such as Weak Gravitational Lensing [74–76], the Sunyaev-Zel’dovich effect [77], Cosmic Microwave Background (CMB) observations [78], peculiar velocity measurements [79] and the Lyman-alpha forest [80]. Integrating direct and indirect measurements provides insights into the cosmic structure and its evolutionary history.

1.2 Theory of large scale structures

The current understanding of the theory of Large Scale Structure (LSS) starts with the end of the inflationary phase, ~ 13.8 billion years ago, where the exponential expansion of the tiny early Universe made quantum fluctuations imprinted on cosmic constituents at that time. This was followed by the nucleosynthesis era, ~ 3 minutes after the inflation, where protons and neutrons combined to form the first elements. However, the Universe remained too hot for atoms to form due to scatterings of photons and free electrons. Due to the adiabatic nature of the expansion of the Universe, radiation fluid looses energy faster compared to matter particles, leading to matter sector domination in the total energy of the Universe, $\sim 50,000$ years after the inflation. At $\sim 380,000$ years after the inflation, the Universe became cold enough to sustain atoms, called the recombination phase [81–84]. The evolution of the matter density field during these times involves intricacies like Mészáros effect [85, 86], Baryon Acoustic Oscillations (BAO) [87], Silk damping [88], integrated Sachs-Wolfe (ISW) effect [89], and change in superhorizon scale. All of these can be encapsulated within the transfer function framework² [93, 94], it is a mechanism that maps the gravitational potential perturbations field at the end of the inflationary phase to the density and velocity perturbations field for each constituent at the end of recombination. It is possible as the perturbations have a small amplitude at these stages, and the equations governing the evolution are linear. These calculations are consistent with observations to great precision [95]. Recombination is followed by the Dark ages, the formation of the first stars, reionisation, and dark energy domination [83, 84]. During these phases, the problem of structure formation transitions from a *linear, quasi-linear* to the *non-linear regime* [4, 6, 82]. This section discusses some theoretical modelling in linear and quasi-linear regimes.

1.2.1 Homogeneous Universe

A homogeneous Universe is one where the properties, such as energy density and temperature, are the same at every point in space on large scales. The spacetime geometry of such a

²Sophisticated Boltzmann codes like CMBFAST [90], CAMB [91] and CLASS [92] are used to calculate transfer function efficiently.

universe is given by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric [96–99],

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dx^2}{1 - \kappa x^2} + x^2(\sin^2 \theta d\phi^2 + d\theta^2) \right], \quad (1.1)$$

where $a(t)$ is the scale factor normalised to unity at present ($a_0 = 1$). (x, θ, ϕ) is the comoving coordinate of the fundamental observer. Due to the expansion or contraction of the Universe, working with comoving coordinates \mathbf{x} is more convenient than using physical coordinates \mathbf{r} . They are related by $\mathbf{r} = a\mathbf{x}$. κ is the curvature constant (the positive, zero and negative values of κ indicates hyperspherical, flat, and hyperbolic universe respectively). t stands for cosmic time. Alternatively, the conformal time (τ), such that $dt = a(\tau)d\tau$, can be used as a time variable. Similarly, redshift defined as $1 + z \equiv a^{-1}$, is often used as alternative to scale factor (appendix A.2).

The energy density in the Universe governs the evolution of $a(t)$. In order to solve for $a(t)$, Einstein's field equations [100] for the FLRW metric need to be solved. Under the assumptions of *homogeneity* and *isotropy*, derived dynamical equations are called Friedmann equations [4, 9, 101–106].

$$\text{Friedmann equations: } \left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3}\rho - \frac{\kappa}{a^2} + \frac{\Lambda}{3}, \quad (1.2a)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3}, \quad (1.2b)$$

where the dot denotes differentiation with respect to t , ρ is the total energy density and P is the pressure and Λ is the cosmological constant [81, 83, 84]. G denotes the gravitational constant, and $c = 1$ unit is adopted.

Energy conservation in an expanding Universe leads to the evolution of energy density ρ , known as the continuity equation.

$$\dot{\rho} = -3H(\rho + P), \quad (1.3)$$

where, $H = \dot{a}/a$ is the *Hubble parameter*, its value at present ($a = a_0$) is called the *Hubble-Lemaître constant* H_0 . Equation (1.3) can be solved for perfect fluids with equation of state $P = \omega\rho$. Various contributors to the total energy include radiation ($\omega_r = 1/3$), non-relativistic matter ($\omega_m \approx 0$), and cosmological constant ($\omega_\Lambda = -1$). Let us define *critical density* $\rho_c \equiv 3H^2/8\pi G$, corresponding to the total energy density in a flat Universe. A dimensionless cosmological parameter such as $\Omega = \rho/\rho_c$, can be formulated. From this, it follows that for radiation $\Omega_r \sim a^{-4}$, matter $\Omega_m \sim a^{-3}$ and cosmological constant $\Omega_\Lambda \sim a^0$. Present values of these ratios are constrained by various observations [95, 107, 108].

1.2.2 Dynamics of gravitational instability

The Classical Boltzmann equation (CBE) dictates the fundamental description for particles in the phase space, accounting for advection, external forces, and collisions. For the scope of this thesis, the CBE for a cold dark matter particle field, described as a *collisionless* fluid without external forces, is given as

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f - \nabla \Phi \cdot \frac{\partial f}{\partial \mathbf{v}} = 0, \quad (1.4)$$

where $f(\mathbf{x}, \mathbf{v}, t)$ is the distribution function, which gives the number density of particles at a specific position \mathbf{x} , velocity \mathbf{v} , and time t in phase space. Φ is a gravitational potential field, given by the Poisson equation,

$$\nabla^2 \Phi = 4\pi G \bar{\rho}_m a^2 \delta, \quad (1.5)$$

where, $\bar{\rho}_m$ is average matter density in the Universe. $\delta \equiv \delta(\mathbf{x})$ is the matter density contrast field,

$$\delta(\mathbf{x}) \equiv \frac{\rho_m(\mathbf{x}) - \bar{\rho}_m}{\bar{\rho}_m}. \quad (1.6)$$

The complete analytic solution to the CBE is elusive, but its moment equations can be useful. The zeroth moment of eq. (1.4) yields the continuity equation, and the first moment provides the Jeans equation (the collisionless counterpart of the Euler equation in fluid dynamics). Incorporating the expansion of the Universe, they are expressed as

$$\text{Continuity equation: } \frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot [(1 + \delta) \mathbf{v}] = 0, \quad (1.7a)$$

$$\text{Jeans equation: } \frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a} \mathbf{v} + \frac{1}{a} (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{\nabla \Phi}{a} - \frac{\nabla \delta \sigma^2}{a(1 + \delta)}, \quad (1.7b)$$

where, $\sigma_{ij} \equiv [\langle v_i v_j \rangle - \langle v_i \rangle \langle v_j \rangle]^{1/2}$ is the stress tensor. After the onset of the matter dominant phase of the Universe, the growth of dark matter perturbations (δ) is governed by

$$\frac{\partial^2 \delta}{\partial t^2} + 2 \frac{\dot{a}}{a} \frac{\partial \delta}{\partial t} = \frac{1}{a^2} \nabla \cdot (1 + \delta) \nabla \Phi + \frac{1}{a^2} \frac{\partial}{\partial x_i \partial x_j} [(1 + \delta) v^i v^j]. \quad (1.8)$$

In eq. (1.8), Einstein summation conventions is assumed. It is called the *Vlasov equation*, and is derived using eq. (1.7). It is the basic equation for the evolution of density perturbations due to gravitational interaction in a cosmological setting. Detailed derivations can be found in [4, 82, 83]. Equation (1.8) can be solved analytically within the linear regime ($|\delta| \ll 1$) or by considering specific geometries for the initial perturbations.

1.2.3 Eulerian perturbation theory

It is assumed that matter is slightly perturbed from background values, i.e. $|\delta| \ll 1$ in eq. (1.6). It holds true for certain epochs in the early Universe. Also, this approximation effectively describes the behaviour of matter on large scales, although there are strong non-linear effects at small scales [4]. The Vlasov eq. (1.8) can be linearised in δ as

$$|\delta| \ll 1 \quad \text{and} \quad c_s \ll c, \quad \frac{\partial^2 \delta}{\partial t^2} + 2 \frac{\dot{a}}{a} \frac{\partial \delta}{\partial t} = 4\pi G \bar{\rho}_m \delta, \quad (1.9)$$

where c_s is the sound speed and c is speed of light. To arrive at eq. (1.9) a Poisson eq. (1.5) is used. For structure formation, a *growing mode* solution of eq. (1.9) is relevant, and it has the following form:

$$\delta(\mathbf{x}, t) = \mathcal{G}(t)\delta(\mathbf{x}, 0), \quad \text{where} \quad \mathcal{G}(\mathbf{x}) \propto \frac{H}{H_0} \int_{a(t)}^{\infty} \frac{da}{a} \left[\frac{H_0}{H(a)} \right]^3. \quad (1.10)$$

It shows that the linear matter density field can be scaled in time using the growth solution $\mathcal{G}(t)$. Note, $\mathcal{G} \propto a$ as $\Omega_m \rightarrow 1$.

At higher order, the Eulerian perturbation theory can be implemented by expanding the density and velocity fields,

$$\delta(\mathbf{x}, t) = \delta^{(1)}(\mathbf{x}, t) + \delta^{(2)}(\mathbf{x}, t) + \dots, \quad (1.11a)$$

$$v(\mathbf{x}, t) = v^{(1)}(\mathbf{x}, t) + v^{(2)}(\mathbf{x}, t) + \dots, \quad (1.11b)$$

where, $\delta^{(1)}(\mathbf{x}, t)$ is first order δ same as in eq. (1.10). $\delta^{(2)}(\mathbf{x}, t)$ describes to leading order the non-local evolution of the density field. It is proportional to the *second order growth factor* \mathcal{G}_2 [109]. Depending on the cosmological parameters, different expressions for \mathcal{G}_2 can be found in the literature [9, 109].

A detailed non-linear Eulerian perturbation theory involves components like kernels, propagators, and vertices. For a comprehensive discussion, refer to [9].

1.2.4 Lagrangian perturbation theory

In §1.2.2, fluid approximation for cold dark matter (CDM) is demonstrated. Eulerian perspective, which focuses on specific spatial locations to describe fluid fields, is discussed in §1.2.3. Alternatively, the field can be described by following the trajectories of particles or fluid elements. It is called the Lagrangian description, which forms the basis of the Lagrangian perturbation theory. It is used to compute the velocities of the particles to generate initial conditions for the simulations.

In this description, the object of interest is the displacement field $\Psi(\mathbf{q})$ which maps the initial comoving particle position \mathbf{q} to its final comoving Eulerian position \mathbf{x} [9, 109, 110].

$$\mathbf{x}(\mathbf{q}, \tau) \equiv \mathbf{q} + \boldsymbol{\Psi}(\mathbf{q}, \tau). \quad (1.12)$$

Let $J(\mathbf{q}, \tau)$ be the Jacobian of the transformation between \mathbf{q} and \mathbf{x} ,

$$J(\mathbf{q}, \tau) \equiv \left| \frac{\partial \mathbf{x}}{\partial \mathbf{q}} \right| = \left| \det \left(\mathcal{I} + \frac{\partial \boldsymbol{\Psi}}{\partial \mathbf{q}} \right) \right|. \quad (1.13)$$

The Jacobian can have the following form if the conservation of mass element is implemented,

$$J(\mathbf{q}, \tau) = \frac{1}{1 + \delta(\mathbf{x}, \tau)} \quad \text{or} \quad \delta(\mathbf{x}, \tau) = J^{-1}(\mathbf{q}, \tau) - 1. \quad (1.14)$$

When trajectories first cross, fluid elements from different initial positions \mathbf{q} converge to the same Eulerian position \mathbf{x} via mapping in eq. (1.12). This is called *shell-crossing*, upto it eq. (1.14) is valid. The Jacobian becomes zero, indicating a singularity and an expected collapse to infinite density. Beyond this point, the dynamics described by the mapping is not valid.

Equation of motion

The Vlasov eq. (1.8) can be modified to obtain the equation of motion of fluid element with the conformal time as,

$$J(\mathbf{q}, \tau) \nabla_{\mathbf{x}} \cdot \left[\frac{\partial^2 \boldsymbol{\Psi}}{\partial \tau^2} + \mathcal{H}(\tau) \frac{\partial \boldsymbol{\Psi}}{\partial \tau} \right] = \frac{3}{2} \Omega_M(\tau) \mathcal{H}^2(\tau) [J^{-1}(\mathbf{q}, \tau) - 1], \quad (1.15)$$

where $\mathcal{H} = aH$ and $\nabla_{\mathbf{x}}$ is gradient in the Eulerian coordinates. The main challenge of the Lagrangian approach is evident in this equation; the gradient operator needs to be defined with respect to the Eulerian variable \mathbf{x} , which is dependent on \mathbf{q} . Expressing eq. (1.15) in Lagrangian coordinates introduces the displacement shears. The resulting non-linear differential equation can then be solved perturbatively, expanding around its linear solution.

The Zel'dovich approximation

The Zel'dovich approximation [5, 29] is first order Lagrangian perturbation theory; it takes the linear solution of eq. (1.15) for the displacement field while keeping the general equation with the Jacobian eq. (1.14). Using linear order relations, $J(\mathbf{q}, \tau) \nabla_{\mathbf{x}} \approx \nabla_{\mathbf{q}}$ and $J(\mathbf{q}, \tau) \approx 1 + \nabla_{\mathbf{q}} \cdot \boldsymbol{\Psi}$, eq. (1.15) takes the form

$$\nabla_{\mathbf{q}} \cdot \boldsymbol{\Psi}^{(1)}(\mathbf{q}, \tau) = -\mathcal{G}(\tau) \delta^{(1)}(\mathbf{q}), \quad (1.16)$$

where $\mathcal{G}(\tau)$ denotes the linear growth factor as in eq. (1.10). Note that the evolution of fluid elements in linear order is local. In linear order, the displacement field is entirely determined

by its divergence, i.e., no vorticity. This holds true as initial vorticity diminishes due to the expansion of the Universe, as only the divergence of the velocity field appears in the linear-order fluid equations. Therefore, particles go straight (in comoving coordinates) in the direction set by their initial velocity.

Shell-crossing in the Zel'dovich approximation

Let $\phi(\mathbf{q})$ be Zel'dovich potential such that $\Psi^{(1)}(\mathbf{q}, \tau) = -\mathcal{G}(\tau)\nabla_{\mathbf{q}}\phi$. If $\lambda_1(\mathbf{q}) \leq \lambda_2(\mathbf{q}) \leq \lambda_3(\mathbf{q})$ are the local eigenvalues of the Hessian of the Zel'dovich potential, then using eq. (1.14) the density contrast δ can be written as [9, 109],

$$1 + \delta(\mathbf{x}, \tau) = \frac{1}{[1 - \lambda_1(\mathbf{q})\mathcal{G}(\tau)][1 - \lambda_2(\mathbf{q})\mathcal{G}(\tau)][1 - \lambda_3(\mathbf{q})\mathcal{G}(\tau)]}. \quad (1.17)$$

This equation provides insights into shell-crossing events within the Zel'dovich Approximation (ZA). The perturbation is linear as long as $\mathcal{G}(\tau)\lambda_i \ll 1$. When $\mathcal{G}(\tau) = 1/\lambda_i$ shell crossing happens along the i^{th} direction of the eigenvector. Therefore, if $\lambda_3 > 0$, and larger than λ_2 and λ_1 , the ZA predicts a planar collapse to infinite density along this axis, forming a two-dimensional *cosmic pancake*. In cases where at least two eigenvalues are positive, $\lambda_3, \lambda_2 > 0$ and larger than λ_1 , a collapse into a filament occurs. For $\lambda_1, \lambda_2, \lambda_3 > 0$, gravitational collapse happens in all directions, with the largest eigenvalue direction collapsing first, leading to triaxial collapse. For spherical collapse, $\lambda_1 \approx \lambda_2 \approx \lambda_3$ is required near the centre of the spherical symmetric region where the density distribution is isotropic; away from the centre, $\lambda_1 > \lambda_2, \lambda_3$ and $\lambda_2 \approx \lambda_3$, reflecting an axisymmetric collapse. If all eigenvalues are negative, it indicates an evolving underdense region eventually reaching $\delta = -1$. This concept leads to a cosmic web classification algorithm that categorises regions as voids, sheets, filaments, and halos [111, 112].

The solution of eq. (1.15) up to second order in $\phi(\mathbf{q})$ takes into account the fact that gravitational instability is non-local, i.e. it includes the correction to the ZA displacement due to gravitational tidal effects [113]. It is referred to as *second-order Lagrangian perturbation theory*, offering a significant enhancement over the Zel'dovich Approximation.

1.2.5 Spherical and ellipsoidal collapse

The spherical collapse model offers a simplified yet insightful depiction of how galaxies or clusters form from initial spherical symmetric perturbations. Such perturbations can be treated as a separate Universe, where the gravitational interaction of matter within the perturbation governs its evolution independently of the surrounding Universe [30, 31, 114]. Although the evolution of overdense/underdense regions and background densities ideally requires Friedmann eqs. (1.2a) and (1.2b), often Newtonian mechanics is sufficient to address this problem.

Let R denote the comoving size of the spherical region, then its evolution is given as

$$\frac{d^2R}{dt^2} = -\frac{4\pi}{3}G\bar{\rho}_m R. \quad (1.18)$$

In general, solving equation eq. (1.18) reveals the existence of a deterministic relationship between the initial comoving Lagrangian size R_0 , the linear density of an object δ_{lin} , and its Eulerian size R at any subsequent time, particularly in an Einstein-de Sitter (EdS) universe,

$$\frac{R}{R_0} = \frac{1 - \cos \theta}{2\bar{\delta}_0}, \quad \frac{t}{t_0} = \frac{3}{4\bar{\delta}_0^{3/2}} [\theta - \sin \theta], \quad \delta_{\text{lin}} = \frac{5}{3} \left(\frac{3}{4}\right)^{2/3} [\theta - \sin \theta]^{2/3}. \quad (1.19)$$

Initially, overdense region collapse when $\theta = 0$, it *turnaround* at $\theta = \pi$, and collapse completely when $\theta = 2\pi$. Overdense regions collapse in finite time when their extrapolated linear overdensity reaches a threshold of $1.686 \equiv \delta_{sc}$.

Similarly, for evolution of uniform ellipsoid [115],

$$\frac{d^2R_i}{dt^2} = -\frac{4\pi}{3}G\bar{\rho}_m R_i \left[1 + \frac{3}{2}\alpha_i \bar{\delta}\right], \quad (1.20)$$

where, i denotes coordinate axes and $\alpha_i \equiv \alpha_i(a_1, a_2, a_3)$, a_i denoting the half-lengths of the principal axes of the ellipsoid. Unlike spherical collapse where threshold for collapse is constant, ellipsoidal collapse lead to the threshold collapse overdensity $\delta_{ec} \equiv \delta_{ec}(R, t, a_i)$ [116–118].

1.2.6 Non-linear approximations

In §1.2.3 and §1.2.4, we delve into the theoretical modelling of structure formation, primarily applicable within the linear and quasi-linear regimes. Analysing the non-linear regime analytically proves to be considerably complex. However, several non-linear approximations to the equations of motion have been proposed in the literature to extend analytical calculations into the non-linear regime. These approximations involve substituting one of the dynamic equations (continuity eq. (1.7a), Jeans eq. (1.7b), and Poisson eq. (1.5)) with a different assumption.

The non-linear extension of the Zel'dovich approximation suggests that the relationship between the velocity field and gravitational potential holds beyond just the first order in Ψ [109, 119]. It is also possible to formulate local approximations suited for cylindrical or spherical collapse [37, 120]. Another approach is the ‘frozen flow’ approximation, where the velocity field remains linear while the density field can venture into the non-linear regime [121, 122]. Similarly, if a gravitational potential is assumed to be static, it gives a ‘frozen potential’ approximation [35]. In the ‘adhesion’ approximation, a viscosity term is introduced to the single-stream Euler equation [32, 123–125].

Although all such approximations offer valuable insights, computational methods, particularly N-body simulations are essential for studying gravitational instability for generic density fields.

1.3 Challenges in large scale structure studies

In the past several decades, progress in cosmology has been remarkable. From complex galaxy distribution patterns to minute variations in the cosmic microwave background (CMB), discoveries have shaped our understanding of the Universe. Using powerful computational tools for simulating structure formation has further paddled cosmology forward, bringing us closer to solving how complex structures in the Universe have emerged from an initially uniform, hot, and dense state. However, these developments have also exposed new questions that need to be addressed, including the validity of the cosmological principles and general relativity, early Universe processes, roles of the dark matter and dark energy, and accurate modelling of astrophysical phenomena such as galaxy formation [126–130]. On the other hand, it is an era of precision cosmology, where various cosmological tensions have arisen from inconsistencies between different observational data sets or between observations and theoretical predictions [131, 132]. These discrepancies challenge our existing understanding of the Universe. During the last two decades, numerical simulations have become essential in large scale structure studies. Understanding these tools in detail, including how approximations in simulation algorithms impact structure formation, is crucial [133–135].

The observed distribution of galaxies, abundance of halos/clusters lensing effects, and peculiar velocities (deviations from the smooth Hubble flow due to local gravitational forces) are influenced by non-linear gravitational interactions. Therefore, Studying the non-linear evolution of large scale structure is essential for accurate cosmological predictions. In the present phase, testing the underlying assumptions and investigating the role of non-linear evolution while making predictions, such as the abundance of halos and the determination of the Hubble-Lemaître constant, is necessary. In this thesis, we discuss the case of the abundance of halos and the determination of the Hubble-Lemaître constant. We rely on N-body simulations for our studies; therefore, we also discuss the simulation approximations affecting the accuracy of our calculations.

1.3.1 Non-universality of halo mass function

In §3, we explore the halo mass function in scale-invariant models [1]. The halo mass function represents the comoving number density of the bound objects, or halos, of mass m at redshift z . Theoretical models of the halo mass function are grounded in the assumption that these halos form from regions in the initial density field that eventually collapse due to their sufficient overdensity. The theoretical formulation of mass function can be written in a form that is *universal*, i.e., independent of the cosmology model and initial condition. However, cosmological simulations have shown that this universality is approximate. We investigate the initial power spectrum³ dependence of halo mass function through N-body simulations

³Power spectrum is $\langle \delta(\mathbf{x})\delta(\mathbf{x}') \rangle$ in Fourier space. Refer §2.2.2

of power-law power spectrum initial conditions in an Einstein-de Sitter cosmology (Flat universe with only matter component. Refer §2.4.1). This specific choice of cosmology and initial condition ensures the self-similar evolution of the dark matter distribution, allowing us to pinpoint the influence of the power spectrum on the mass function. Our findings reveal a *clear departure from the assumed universality of the mass function*.

1.3.2 Determination of the Hubble-Lemaître constant

In §4, we discuss local measurements of the Hubble-Lemaître H_0 constant in cosmological N-body simulations. Hubble constant denotes the rate of expansion of the Universe and shows major differences between early and late time measurements. This discrepancy is called ‘Hubble tension’. Various potential solutions have been proposed, from modifications to the cosmological model to addressing observational systematics [136]. The peculiar velocities of galaxies, resulting from gravitational clustering, add complexity to H_0 measurement. It introduces a Doppler component to the redshift, affecting the inferred recessional velocity (appendix A.3). The effect is significant on scales less than about 100 Mpc/h. In this work, we study the effect of peculiar velocities on the measurement of H_0 in the local Universe. *We show a correlation between gravitational clustering and local H_0* through simulations. Such correlation can affect the calibration in the local observations, may justify the Hubble tension.

1.3.3 Mass discreteness effect in simulations

In §5, we study the effect of gravitational clustering at small scales on larger scale by studying mode coupling between virialised halos. The motivation of this study is to understand the impact of finite mass resolution of cosmological N-body simulations on the evolution of perturbations at early times. This difference between the expected evolution and the evolution obtained in cosmological N-body simulations can be quantified using such an estimate. We also explore the impact of a finite shortest scale up to which the desired power spectrum is realised in simulations. Several simulation studies have shown that this effect is small in comparison with the effect of perturbations at large scales on smaller scales [137, 138]. It is nevertheless important to study this effects and develop a general approach for estimating their magnitude. We provide basic estimates of the magnitude of finite mass resolution effects and their power spectrum dependence.

In this chapter, we trace the history of observing large scale structure (LSS) using galaxy surveys. We explore theoretical models of LSS after establishing the prerequisites. Lastly, we outline general challenges for further exploration in the field and list some of the problems addressed in this thesis. Before delving into these problems, the next chapter introduces some essential tools required for their analysis.

2 Structure formation toolkit

To interpret the vast datasets from the galaxy surveys and the matter distribution derived from N-body simulations, a mathematical description of their statistical properties is essential. These properties include n -point correlation functions, power spectra, cross-correlations, halo mass functions, fractal structure analysis, and Minkowski functionals [4, 139–141]. Further, N-body simulations serve as pivotal computational tool for investigating the formation and evolution of cosmic structures by modelling gravitational interactions among a large number of particles, representing dark matter, gas, and stars [35, 48–51]. Thus, structure formation is a multi-faceted field that draws from various disciplines, such as physics, mathematics, and computational science. In this chapter, we review the fundamentals of these tools and key definitions used in the subsequent thesis chapters. We run a suite of simulations required for our studies, described in detail in section §2.4.

2.1 Statistical description of the density field

This section describes the statistical properties of the density contrast field $\delta(\mathbf{x})$ (eq. (1.6)), considering it as a random field. This description is valid for any other random field, such as gravitational potential and velocity divergence field. As discussed in §1.2 the process of generating seed perturbations in δ is stochastic. For an arbitrary number n of spatial positions \mathbf{x}_i , one can define the joint multivariate *Probability Distribution Function* (PDF) \mathcal{P} to have $\delta(\mathbf{x}_1)$ between δ_1 and $\delta_1 + d\delta_1$, $\delta(\mathbf{x}_2)$ between δ_2 and $\delta_2 + d\delta_2$, etc. Such a probability is expressed as

$$\mathcal{P}(\delta_1, \delta_2, \dots, \delta_n) d\delta_1 d\delta_2 \dots d\delta_n. \quad (2.1)$$

2.1.1 Average and ergodicity

The average of a quantity can be defined in two ways. Firstly, one can average by considering numerous realisations drawn from the distribution, all generated in the same manner (e.g. N-body simulations). This is termed the *ensemble average*, for any quantity $X(\delta_1, \delta_2, \dots, \delta_n)$, it is defined as

$$\langle X \rangle \equiv \int X(\delta_1, \delta_2, \dots, \delta_n) \mathcal{P}(\delta_1, \delta_2, \dots, \delta_n) d\delta_1 d\delta_2 \dots d\delta_n. \quad (2.2)$$

Also, one can compute the average by considering the quantity of interest at various locations within a single realisation of the distribution. This is referred to as the *sample average*. For a volume V in the Universe, the sample average of a quantity X can be expressed as

$$\bar{X} \equiv \frac{1}{V} \int_V X(\mathbf{x}) d^3\mathbf{x}. \quad (2.3)$$

If the ensemble average of the quantity can be estimated as the sample average, then the system is said to be *ergodic*. In cosmology, ergodicity is frequently assumed if the volume under consideration is sufficiently large. When ergodicity holds in cosmology, the system is considered a *fair sample* of the Universe.

2.1.2 Gaussian random field (GRF)

Gaussian random fields (GRFs) has an extensive application in cosmostatistics [142, 143]. Evidence suggests that the initial fluctuation field closely resembled a GRF [144–146]. Perturbation theory based descriptions are used to explain the gravitational clustering originating from Gaussian initial fluctuations [9]. In the late-time structure formation, the large scale distribution of galaxies can be approximately represented as a GRF, especially on scales where gravitational evolution remains well described by linear perturbation theory.

A Gaussian random field is characterised by multivariate Gaussian probability distributions for various m -point distributions, such as eq. (2.1). The joint Gaussian probability distribution for random variables δ_i can be expressed as

$$\mathcal{P}(\delta_1, \delta_2, \dots, \delta_n) d\delta_1 d\delta_2 \dots d\delta_n = \frac{e^{-\mathcal{Q}}}{[(2\pi)^n \text{Det}(M)]^{1/2}} d\delta_1 d\delta_2 \dots d\delta_n, \quad \text{where} \quad (2.4a)$$

$$\mathcal{Q} \equiv \frac{1}{2} \sum \Delta\delta_i (M^{-1})_{i,j} \Delta\delta_j, \quad \text{and} \quad (2.4b)$$

$$\Delta\delta_i \equiv \delta_i - \langle \delta_i \rangle, \quad M_{ij} \equiv \langle \Delta\delta_i \Delta\delta_j \rangle. \quad (2.4c)$$

Therefore, the averages of the random variables $\langle \delta_i \rangle$ and covariance matrix M suffice to define the distribution.

The renowned BBKS (Bardeen, Bond, Kaiser, Szalay) paper provides a comprehensive analytical overview of statistics of peaks in Gaussian random fields [116]. Note that the assumption of Gaussianity for $\delta(\mathbf{x})$ breaks down in the nonlinear regime of structure formation. $\delta_{\text{lin}}(\mathbf{x})$ (denoted as $\delta^{(1)}(\mathbf{x})$ in eq. (1.16)), is essentially the divergence of the velocity field, represents a Gaussian random field since the initial velocities are Gaussian. The nonlinear density contrast $\delta(\mathbf{x})$ (eq. (1.17)) is constrained by $1 + \delta(\mathbf{x}) \geq 0$, implying $\delta(\mathbf{x}) \geq -1$. This

lower bound and the growth of nonlinear corrections skew the distribution, violating Gaussian symmetry. $\delta_{\text{lin}}(\mathbf{x})$ resembles $\delta(\mathbf{x})$ only when $(|\delta(\mathbf{x})| \ll 1)$; thus, the Gaussian assumption holds only in the linear regime, where higher-order contributions are negligible.

2.2 Measures of gravitational clustering

An enormous variety of clustering statistics is used for large scale structure study. These include correlation functions and their Fourier counterparts, mass functions, clustering/halo models, angular correlations, scaling relations, neighbour statistics, and the Mandelbrot prescription, which are few among all [4]. This section only focuses on the clustering statistics used for this thesis work.

2.2.1 n -point correlation functions

n -point correlation functions provide a systematic way to measure the clustering properties of a dataset by examining the spatial relationships between groups of points. These functions quantify the excess probability of finding n -points within certain distances of each other, compared to what would be expected in a random distribution. For example, the two-point correlation function measures the excess probability of finding pairs of points at a given distance compared to a random distribution. Higher-order correlation functions, such as the three-point, four-point, or higher, provide information about the clustering beyond pairs of points, capturing more complex spatial arrangements.

The n -point correlation function ξ_n of density fluctuations $\delta(\mathbf{x})$ is defined by

$$\langle \delta(\mathbf{x}_1) \dots \delta(\mathbf{x}_n) \rangle_c \equiv \xi_n(\mathbf{x}_1 \dots \mathbf{x}_n). \quad (2.5)$$

The subscript c stands for connected moments/cumulants¹. If all the \mathbf{x}_i are the same, then

$$\langle \delta \rangle_c = \langle \delta \rangle, \quad (2.6a)$$

$$\langle \delta^2 \rangle_c = \langle \delta^2 \rangle - \langle \delta \rangle_c^2 \equiv \sigma^2, \quad (2.6b)$$

$$\langle \delta^3 \rangle_c = \langle \delta^3 \rangle - 3\langle \delta^2 \rangle_c \langle \delta \rangle_c - \langle \delta \rangle_c^3, \quad (2.6c)$$

$$\langle \delta^4 \rangle_c = \langle \delta^4 \rangle - 4\langle \delta^3 \rangle_c \langle \delta \rangle_c - 3\langle \delta^2 \rangle_c^2 - 6\langle \delta^2 \rangle_c \langle \delta \rangle_c^2 - \langle \delta \rangle_c^4. \quad (2.6d)$$

For the density perturbation field eq. (1.6), $\langle \delta \rangle = 0$. Thus, $\langle \delta^n \rangle_c = \langle \delta^n \rangle$ upto $n = 3$. For a Gaussian random field, all higher order moments can be expressed using the two-point correlation function. For fair sample, the two-point correlation function ξ is defined as

$$\xi(\mathbf{r}) = \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \rangle = \frac{1}{V} \int d^3 \mathbf{x} \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}). \quad (2.7)$$

¹Connected moments/cumulants are a set of quantities that provide an alternative to the moments of the distribution.

Note that the subscript n on ξ is dropped. For discrete N points distribution over volume V , the two-point correlation function can be defined in a way such that,

$$\delta\mathcal{P} = \bar{n}\delta V [1 + \xi(\mathbf{r})], \quad (2.8)$$

where, $\delta\mathcal{P}$ is probability of finding neighbour at distance $r = |\mathbf{r}|$ in small volume δV with respect to randomly chosen point (Refer section §31 of [4]). Similarly, higher order correlation functions can be formulated. The volume average of the correlation function $\bar{\xi}(r)$ is interpreted as the excess number of neighbours within a distance r of a given point

$$\bar{\xi}(r) = \frac{3}{r^3} \int_0^r y^2 \xi(y) dy. \quad (2.9)$$

2.2.2 Power spectrum

Many of the calculations involved in structure formation studies get simplified considerably in the *Fourier space*. Here, we discuss the power spectrum, which is the Fourier transform of the correlation function.

Let $\delta_{\mathbf{k}}$ be the Fourier transform of $\delta(\mathbf{r})$ (appendix B.1). Thus,

$$\xi(\mathbf{r}) = \int \frac{d^3k}{(2\pi)^3} |\delta_{\mathbf{k}}|^2 e^{i\mathbf{k}\cdot\mathbf{r}} = \int \frac{d^3k}{(2\pi)^3} P(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}}, \quad (2.10)$$

where, $P(\mathbf{k}) \equiv \langle |\delta_{\mathbf{k}}|^2 \rangle$ is called power spectrum. In an isotropic Universe, the perturbations cannot have a preferred direction; therefore, $\xi(\mathbf{r}) = \xi(r)$ and $P(\mathbf{k}) = P(k)$. This means ξ and P are statistical quantities rather than local quantities like δ . Similarly, higher order quantities like bispectrum and trispectrum can be defined (appendix B.3). Working with a dimensionless power spectrum is often more convenient. It is defined as

$$\Delta^2(k) \equiv \frac{k^3 P(k)}{2\pi^2}. \quad (2.11)$$

2.2.3 Mass variance

Mass variance is root-mean-square (rms) fluctuations in mass at a given length scale r . A commonly used parameter σ_8 is the filtered mass variance in spheres of radius $r = 8h^{-1}$ Mpc, roughly corresponding to the scale of massive galaxy clusters. Its more generic form for arbitrary r is given by

$$\sigma^2(r) = \frac{\langle [M - \langle M \rangle]^2 \rangle}{\langle M \rangle^2} = \int \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} |W(kr)|^2, \quad (2.12)$$

where $W(kr)$ is some window function² (appendix B.4) used to smooth the density field so that $M = f_W r^3 \rho_m$, for a realistic case, we need $f_W = 4\pi/3$, which is satisfied by spherical top-hat window function, however, there are other window functions which in use, some of them and their Fourier transforms are discussed in appendix B.4.

2.2.4 Halo mass function

The halo mass function is a statistical distribution that describes the abundance of dark matter halos of different masses at different times. Formally, it is the comoving number density of halos of mass m at redshift z ,

$$\frac{m^2 n(m, z) dm}{m} = \nu f(\nu) \frac{d\nu}{\nu}, \quad \nu \equiv \frac{\delta_{sc}^2(z)}{\mathcal{G}^2(z) \sigma_{lin}^2(m)}. \quad (2.13)$$

where δ_{sc} is the critical overdensity threshold for spherical collapse, it generally has a weak dependence on cosmology. \mathcal{G} is growth factor. σ_{lin} variance in the initial density perturbation field extrapolated to the present time using linear theory and smoothed with a top hat filter ($W(k, m)$) of scale $r = (3m/4\pi\bar{\rho})^{1/3}$. It is calculated as shown in eq. (2.12). There are two fundamental models for $\nu f(\nu)$ based on spherical and ellipsoidal collapse (discussed in §1.2.5), namely Press-Schechter and Sheth-Tormen mass function,

$$\text{Press-Schechter: } \nu f(\nu) = \sqrt{\frac{\nu}{2\pi}} e^{-\nu/2}, \quad (2.14a)$$

$$\text{Sheth-Tormen: } \nu f(\nu) = A(p)[1 + (q\nu)^{-p}] \sqrt{\frac{q\nu}{2\pi}} e^{-q\nu/2}. \quad (2.14b)$$

where p and q are termed as the Sheth-Tormen parameters, we revisit them in §3. When $p = 1/2$ and $q = 1$ in eq. (2.14b), the eq. (2.14a) is recovered. For $\nu \ll 1$, the mass function scales as $\nu^{0.5-p}$. While the exponential cutoff at $\nu \gg 1$ depends on the value of q . The characteristic mass scale denoted as m_* , is defined at $\nu = 1$; note that halos more massive than m_* are rare. Both Press-Schechter and Sheth-Tormen formalism assume that all the mass in the Universe is present in the form of halos of some mass, i.e., $\int_0^\infty f(\nu) d\nu / \nu = 1$. This provides the normalisation for the mass functions.

$$A(p) = \left[1 + \frac{2^{-p} \Gamma(0.5 - p)}{\sqrt{\pi}} \right]^{-1}. \quad (2.15)$$

The derivations of eq. (2.14a) in [147–149] demonstrate its relationship to a model wherein halos form from spherical collapse. When extended to the ellipsoidal collapse model described by [150], the same ideas yield equation eq. (2.14b) [117, 118], which has been found to offer a satisfactory description of the mass function in numerical simulations. However, improved

²A window function is convolved with density fields for smoothing or filtering to reduce noise and emphasise larger scale features (appendix B.4).

simulation precision has revealed that the Sheth-Tormen mass function predicts an excess of halos near m_* and underpredicts at both tails. Further, both eqs. (2.14a) and (2.14b) exhibit *universality*, implying they are independent of the choice of cosmological model and initial conditions. However, §3 demonstrates this does not hold true.

2.3 N-body simulations

When density contrast approaches unity, perturbations enter a non-linear regime, interacting significantly across scales. Solving Vlasov eq. (1.8) for perturbations with generic initial conditions becomes challenging. Hence, N-body simulations are indispensable for studying this regime. Additionally, solutions provided by quasi-linear approximations [9, 29, 32–37] or scaling relations [151–157] must be validated against simulations before application. Emerging fields like Machine Learning (ML) and Artificial Intelligence (AI) offer an alternative approach to traditional simulations [158–163]. They are used for emulating, predicting, generating, and detecting patterns and anomalies in large scale structures. However, a substantial amount of training data generated by full N-body simulations is required for these methods.

Over the last three decades, significant progress has been made in cosmological simulations driven by refinement in techniques and increased computing power. State-of-the-art simulations, such as MillenniumTNG [164] and EUCLID Flagship Simulation [165], involve more than 10^{12} particles. Developing new numerical methods, optimisation algorithms, and parallel programming has been crucial in such advancements. These developments have enabled simulations to provide in-depth insights into the structure formation process in the Universe, which is critical for precision cosmology. In this section, we explore fundamental algorithms for N-body simulations, essential concepts in cosmological simulations, and algorithms for structures classification.

2.3.1 Equations of motion and Basic algorithm

In the N-body approach, the continuous matter density field is sampled through a set of discrete identical point-like tracers called ‘particles’, which carry a given mass and only interact via gravity. This is called the *particle approximation* of the Vlasov eq. (1.8). The dynamics of particles is treated in the non-relativistic medium in an expanding cosmological background. It can be studied in the Newtonian limit at much smaller scales than the Hubble radius $d_H = cH_0^{-1}$ (appendix A.1). For a set of N particles, the system of equations that describe their evolution of trajectories is given by

$$\ddot{\mathbf{r}}_i(t) = \ddot{a}(t)\mathbf{x}_i(t) + 2\dot{a}(t)\dot{\mathbf{x}}_i(t) + a(t)\ddot{\mathbf{x}}_i = -\nabla[\phi_b + \phi(\mathbf{x}_i)], \quad (2.16a)$$

$$\phi(\mathbf{x}_i) = -\frac{G}{a^2} \sum_{i \neq j} m_j \frac{1}{|\mathbf{x}_i - \mathbf{x}_j|}, \quad (2.16b)$$

where $\phi_b = -\ddot{a}r^2/2a$ is contribution from the background Universe. \mathbf{r}_i and \mathbf{x}_i are physical and comoving position and m_j is the mass of the i^{th} particle. These are a set of coupled differential equations that are solved in a N-body simulation.

The basic algortim to solve equations eq. (2.16) is given in four steps [166, 167],

1. *Setting up the initial conditions (ICs)*: Initial positions for the particles are distributed according to a Gaussian field using the desired power spectrum for the model of interest [168–172]. This involves sampling positions from a Gaussian distribution with mean zero and variance determined by the power spectrum. Lagrangian perturbation theory (§1.2.4) is used to compute the initial velocities of the particles. It enables the generation of velocities that that are consistent with the selected initial density field for adiabatic perturbations in the growing mode.
2. *Force calculations*: Forces acting on the particles are calculated using eq. (2.16b). Various standard techniques like particle-particle (PM), particle mesh (PM), fast multipole method (FMM), and the Barnes-Hut tree method are used to calculate the forces [48, 166, 173–176]. Advanced codes use hybrid methods such as Particle-Particle Particle-Mesh (P^3M) and Tree Particle-Mesh (TreePM) for force calculations [48, 167, 177].
3. *Time integration*: Positions and velocities of particles are then updated using a numerical integration scheme such as the leapfrog integration to advance the simulation in time [178].
4. *Repeat*: Steps 2 and 3 are repeated multiple times to simulate the evolution of the system over time.

2.3.2 Nuances of cosmological N-body simulations

The algorithm discussed above is very naive. Working simulation codes implement various techniques for optimisations and modifications because the system is cosmological. We list some of them below.

Periodic boundary condition (PBC)

Periodic boundary conditions are used in cosmological simulations to mimic an infinite space by imposing a periodic repetition of the simulation volume. This means that when a particle moves out of one side of the simulation box, it re-enters from the opposite side. The mapping of coordinates (denoted by ℓ) with PBC can be expressed as follows,

$$\mathbf{x}' = \mathbf{x} - \left\lfloor \frac{\mathbf{x}}{L} \right\rfloor L, \quad (2.17)$$

where, $\lfloor \cdot \rfloor$ is the floor function and L is the length of periodic box.

Short-range and long-range force

Most of the modern simulations, including those utilised for this thesis work, use a hybrid TreePM algorithm for force calculations [48, 167]. This algorithm divides the gravitational force into long-range and short-range components by splitting unity in the Poisson eq. (1.5).

$$\phi_k^l = -\frac{4\pi G \rho_k}{k^2} \left[e^{-k^2 r_s^2} \right], \quad (2.18a)$$

$$\phi_k^s = -\frac{4\pi G \rho_k}{k^2} \left[1 - e^{-k^2 r_s^2} \right], \quad (2.18b)$$

where ϕ_k^s and ϕ_k^l are the short-range and long-range potentials in the Fourier space $\phi_k = \phi_k^s + \phi_k^l$. And r_s is a *force split scale*. ϕ_k^l is computed in the Fourier space with the particle mesh method and involves solving the Poisson equation on the *grid*. ϕ_k^s is solved in real space using the Tree method or Fast Multipole Method (FMM).

Short-range forces, dominant at small scales, are calculated with higher computational resolution, ensuring accuracy in capturing local interactions between nearby particles. Long-range forces, significant at larger scales, are efficiently computed using approximation methods such as the particle mesh, enabling simulations to scale to larger computational domains and particle counts while maintaining computational efficiency. This approach optimises computational resources, allowing for accurate and scalable gravitational dynamics simulations across various scales.

Force softening

The gravitational potential generated using Poisson eq. (2.16b) approaches infinite values as the distance between particles approaches zero. This creates a problem in simulating collisionless particles affected by the force of gravity. To fix the issue, eq. (2.16b) can be modified to soften the potential at small scales by introducing a small term ϵ , called the *softening length*, in the denominator to regularise the divergences at zero separation,

$$\phi(\mathbf{x}_i) = -\frac{G}{a^2} \sum_{i \neq j} m_j \frac{1}{[(\mathbf{x}_i - \mathbf{x}_j)^2 + \epsilon^2(\epsilon_0, r)]^{1/2}}. \quad (2.19)$$

The function $\epsilon(r)$ describes a gravitational softening law. Although there are many possible forms for ϵ , in this thesis

$$\epsilon(\epsilon_0, r) = -\frac{2.8\epsilon_0}{W_2(r/2.8\epsilon_0)} - r \quad (2.20)$$

is used. The function $W_2(u)$ is given in [179] (eq. 71).

If ϵ is excessively small, it cannot effectively prevent infinite forces, resulting in noisy, unstable evolution and longer integration times for the simulation. Conversely, large values diminish spatial resolution, hindering precise calculations on small scale [180, 181].

Mass resolution

Mass resolution in N-body simulations refers to the smallest mass each simulated particle represents (m_{part}). Higher mass resolution implies that each particle represents a smaller mass, allowing for better resolution of small scale structures and gravitational interactions. Similarly, lower mass resolution means that each particle represents a larger mass, producing a coarser representation of the simulated system. Achieving higher mass resolution requires simulating more particles, increasing computational costs [182–185]. If V is the simulation volume, then

$$m_{\text{part}} = \frac{\bar{\rho}_m V}{\text{Number of particles in the simulation}}. \quad (2.21)$$

Spatial resolution

Spatial resolution is related to the size of simulation particles that determines the smallest structures that can be resolved in the simulated Universe [184, 186]. It can be defined using the grid³ spacing.

The power spectrum used to generate initial conditions for simulations is restricted at $k_{\max} = \pi/L_{\text{grid}}$, where L_{grid} is grid spacing. Therefore, the initial positions of particles generated in simulations lack information of the omitted modes ($k > k_{\max} \sim L_{\text{grid}}$). Hence, L_{grid} can be used to define spatial resolution, below which simulation results are unreliable, especially at early times. The errors due to missing modes propagate further during the evolution of the simulations. This leads to uncertainties in computed quantities at scales comparable to the grid length. We analyse this effect on mass variance and mass function in §5.

Finite box size effect

The lower end of the input power spectrum is truncated at $k_{\min} = 2\pi/L_{\text{box}}$, where L_{box} represents the length of the simulation box. This truncation introduces notable errors, as modes larger than the box size are disregarded during initial condition generation and subsequent evolution [134, 135, 187]. These errors undermine the reliability of certain quantities within a specific range. Therefore, careful consideration of the box size for simulation is crucial, and the guidelines outlined in [134] offer a prescription for this determination.

2.3.3 Structure classification

Structure classification algorithms aim to categorise and characterise the various components of the large scale structure, such as halos/galaxy clusters, filaments, and voids. These algorithms can be used on observational data from galaxy surveys and simulations. Some stan-

³The grid subdivides the simulation volume into discrete cells, facilitating the calculation of quantities at grid points within the volume.

dard algorithms for classification are Friends-of-Friends (FoF) [188, 189], density-based algorithms [150, 190–193], voronoi tessellation [194], watershed algorithm [195, 196] and topological methods [197, 198].

Halos, gravitationally bound clumps of particles, play a crucial role in understanding the structure formation. While they are not the true observables, galaxies are presumed to form within the potential wells of such halos [18–20]. Therefore, understanding the large scale evolution of galaxies relies on accurately identifying these collapsed structures within the dark matter density field by halo-finder algorithms. Various techniques exist for identifying halos in simulations, each with its advantages and disadvantages, depending on the desired information and the accuracy level required versus the simplicity of the algorithm. Some methods connect neighbouring particles in configuration space, while others incorporate information from the velocity field or identify overdensities in the matter field [150, 188, 190, 191, 199–203].

A plethora of halo-finding algorithms are available in the literature [201, 204–207]. Here, we discuss the FoF-SUBFIND code, a widely used tool for constructing halo catalogues in cosmological simulations [201, 208]. It operates in two main stages. It first identifies Friends-of-Friends (FoF) groups in a given particle distribution. Particles within a certain *linking length* are considered ‘friends’ and grouped into FoF groups based on spatial proximity. Once the FoF groups are identified, the code employs the excursion set algorithm to decompose each FoF object into smaller substructures. This step helps identify subhalos or smaller clumps within larger halos, providing a more detailed characterisation of the halo population.

The FoF-SUBFIND employs a spherical overdensity criterion to define halo masses. This implies that halos are identified based on regions of space where the local density exceeds a certain threshold above the background density (e.g. 200 times overdense is m_{200b} , $18\pi^2$ times overdense is m_{vir}). Consequently, the halo masses determined using different overdensity definitions and algorithms may differ.

Note that various factors can affect the accuracy of halo properties obtained from halo finders [209, 210]. For instance, the finite size of the simulation box can lead to an underestimation of halo number density at the high-mass end. In contrast, discreteness noise due to the finite number of simulation particles can affect halo number density at the low-mass end. Careful consideration and correction for these effects are essential to ensure the reliability of halo properties measurements and the validity of subsequent analyses based on these measurements.

2.4 Suite of simulations in this thesis

In this thesis, we use two types of simulations, namely scale-free and Λ CDM cosmologies. In this section, we highlight their advantages and key features.

2.4.1 Scale-free cosmology

Scale-free cosmological models are characterised by the Einstein-de Sitter (EdS) background cosmologies featuring a power-law initial spectrum of fluctuations, resulting in a self-similar clustering evolution.

In the EdS universe, the spatial curvature is flat, with no cosmological constant, i.e. $\Omega = \Omega_m = 1$. The Friedmann eq. (1.2) describing the evolution of the scale factor is

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2 \Omega_m a^{-3}. \quad (2.22)$$

The solution of the Friedman eq. (2.22) has the form

$$a \propto t^{2/3}. \quad (2.23)$$

Therefore, the scale factor follows a power-law scaling with the coordinate time, and there are no other features in the time scales. When EdS universe is combined with power-law power spectrum as an initial condition of the form

$$P(k) \propto k^n, \quad (2.24)$$

implies that there are no featured scales in the initial power spectrum either. Note that $4 > n \geq -3$; this restriction on the values of n comes from linear theory. $n < -3$ gives divergent density contrast, such as $\xi(r) \rightarrow \infty$ as $r \rightarrow \infty$ at initial times. $n < 4$ assures that non-linear processes at small scales do not affect the behaviour of perturbations at larger scales [137].

Self-similarity

These models are recognised to have self-similar solutions to the Vlasov eq. (1.8), where any dimensionless statistical quantity X derived from the phase space density can be expressed as

$$X(x, t) = X(x/t^\alpha), \quad \text{where } \alpha \in \mathbb{R}. \quad (2.25)$$

For instance, the linear power spectrum is given by

$$P_{\text{lin}}(k, a) = a^2 A k^n, \quad (2.26)$$

where A is the normalisation factor. It gives a dimensionless power spectrum as in eq. (2.11)

$$\Delta_{\text{lin}}^2(k, a) = a^2 \frac{k^3}{2\pi^2} A k^n \propto a^2 k^{3+n}. \quad (2.27)$$

We define the self-similar rescaling by the *scale of non-linearity* r_{nl} and corresponding wavenumber k_{nl} such that $\Delta_{\text{lin}}^2(k_{nl}, a) \equiv 1$. k_{nl} separates the linear regime ($\Delta^2(k) \ll 1$) from the non-linear regime ($\Delta^2(k) \gg 1$). This definition of non-linearity gives

$$k_{nl} \propto a^{-2/(3+n)} \propto t^{-4/3(3+n)}. \quad (2.28)$$

This scaling implies that Δ_{lin}^2 and consequently any dimensionless clustering statistic, is no longer an arbitrary function of two independent variables length r and scale factor a , but solely of the rescaled variable k/k_{nl} . Scale-free cosmologies can thus be represented by a single scale, the scale of non-linearity, containing all the information about temporal evolution. If the gravitational clustering evolution does not rely on other scales, it exhibits self-similarity in the rescaled coordinate. A more general version of self-similar rescaling can be given by extending eq. (2.25)

$$X(x_i, a) = X(x_i/x_{nl,i}(a)), \quad (2.29)$$

where $x_{nl,i}(a)$ is the non-linearity scale with same dimensions as x_i . In scale-free simulations, deviations from self-similarity are solely attributable to the shortcomings of N-body description. These departures originate from limitations due to the introduction of unphysical parameters in discretising the density field and the constraints of the simulation volume. Thus, these models simplify the modelling of large scale structures, offering valuable theoretical insights.

Scaling relations

As demonstrated above, rescaling of dimensionless power spectrum gives,

$$k_{nl} \propto a^{-2/(3+n)}, \quad (2.30)$$

and corresponding r_{nl} scales as

$$r_{nl} \propto k_{nl}^{-1} \propto a^{2/(3+n)}, \quad (2.31)$$

which leads to the volume averaged 2-point linearly extrapolated correlation as

$$\bar{\xi}(r) \approx \left[\frac{r}{r_{nl}} \right]^{-(n+3)}. \quad (2.32)$$

Likewise, *mass of non-linearity* defined as the mass contained within a sphere of radius r_{nl} scales as

$$m_{nl} \propto r_{nl}^3 \propto a^{6/(3+n)}, \quad (2.33)$$

leading to linearly extrapolated mass variance of the form

$$\sigma(m) = \left[\frac{m}{m_{nl}} \right]^{-\frac{n+3}{6}}. \quad (2.34)$$

When extrapolating scale-free results to Λ CDM (cosmological constant Λ and Cold Dark Matter) cosmologies (described in §2.4.3), the transition between scales that emerge during the radiation-dominated era and those that arise during matter domination occurs gradually. Consequently, the change in the slope of the matter power spectrum is slow. Therefore, the Einstein-de Sitter scenario reasonably approximates Λ CDM cosmologies during the matter domination era. Thus, Λ CDM epochs can be approximated by interpolating scale-free cosmologies with different spectral indices, effectively capturing the behaviour of the variance at a standard scale. We define the effective spectral index n_{eff} based on the top-hat variance of density fluctuations as

$$\frac{d \log \sigma(m, a)}{d \log m} \Big|_{\sigma=1} = -\frac{n_{\text{eff}} + 3}{6}. \quad (2.35)$$

2.4.2 Scale-free simulation catalogue

We conduct a series of dark-matter-only simulations using eight power-law power spectrum initial conditions. The power-law indices n of the power spectrum are listed in column 1 of table 2.1. For simulation runs, we use GADGET4 [208], configured in TreePM mode, with initial conditions generated via second-order Lagrangian perturbation theory. As discussed in §2.3.2, we set the force split radius to 2 in grid units. The softening length is set to $\epsilon_0 = 0.05$ grid units across all runs. We employ adaptive time steps with a maximum step size of 0.005. These parameter selections ensure that the force calculation errors remain well below 1%. Halos are identified using FoF-SUBFIND code.

These models necessitate the adoption of Einstein-de Sitter (EdS) cosmology as the background cosmology, where the growth factor of the growing mode scales with the scale factor. As previously mentioned, the only scale involved in these models is the scale of non-linearity, r_{nl} , induced by gravity. It is defined as the scale at which the linearly extrapolated value of the mass variance at a given epoch, $\sigma_{\text{lin}}(a, r_{nl})$, equals unity. Consequently, we can denote epochs in these models in terms of $r_{nl} \propto a^{2/(n+3)}$, where the normalisation of the power spectrum determines the proportionality constant. All power spectra are normalised such that $\sigma_{\text{lin}}(a = 1, r_{nl} = 8) = 1$.

The initial redshift z_{start} for the simulation is determined by the epoch at which the mass variance at grid length reaches approximately 10^{-2} , as listed in column 2 of table 2.1. This ensures a common starting point with respect to the perturbation amplitude for all eight models used in this thesis. However, it is essential to acknowledge that systematic errors in simulations can arise from initial conditions, as indicated by [171]. Hence, we have confirmed that the results presented using these simulations remain unaffected by the chosen starting redshift.

At the beginning of the simulation, the average inter-particle separation is set at a unit grid size, wherein the number of particles in the simulations, denoted as N_{part} , is equal to

| n | z_{start} | L_{box} | $\sqrt[3]{N_{\text{part}}}$ | r_{nl} (grid unit) | r_{nl}^{\max} | CPU hours |
|------|--------------------|------------------------|-----------------------------|--|-----------------|-----------|
| 0.0 | 453 | 512 [$\times 10$] | 512 | [13] 1.26, 1.63, 2.12, 2.76, 3.58, 4.66, 6.06, 7.88, 10.24, 13.31, 17.30, 22.49, 29.24 | 58.7 | 3,21,020 |
| | 100 | 512 | 512 | [13] 1.26, 1.63, 2.12, 2.76, 3.58, 4.66, 6.06, 7.88, 10.24, 13.31, 17.30, 22.49, 29.24 | 58.7 | 3,20,92 |
| -0.5 | 270 | 512 [$\times 10$] | 512 | [14] 1.26, 1.63, 2.12, 2.76, 3.58, 4.66, 6.06, 7.88, 10.24, 13.31, 17.30, 22.49, 29.24, 38.01 | 42.3 | 3,98,608 |
| -1.0 | 160 | 512 [$\times 10$] | 512 | [12] 1.26, 1.63, 2.12, 2.76, 3.58, 4.66, 6.06, 7.88, 10.24, 13.31, 17.30, 22.49 | 27.4 | 2,30,320 |
| | 160 | 768 | 768 | [14] 1.26, 1.63, 2.12, 2.76, 3.58, 4.66, 6.06, 7.88, 10.24, 13.31, 17.30, 22.49, 29.24, 38.01 | 41.1 | 91,776 |
| | 160 | 1024 | 1024 | [15] 1.26, 1.63, 2.12, 2.76, 3.58, 4.66, 6.06, 7.88, 10.24, 13.31, 17.30, 22.49, 29.24, 38.01, 49.41 | 54.8 | 192,482 |
| -1.5 | 96 | 1024 | 1024 | [12] 1.26, 1.63, 2.12, 2.76, 3.58, 4.66, 6.06, 7.88, 10.24, 13.31, 17.30, 22.49 | 27.8 | 1,18,800 |
| -1.8 | 70 | 1024 | 1024 | [10] 1.26, 1.63, 2.12, 2.76, 3.58, 4.66, 6.06, 7.88, 10.24, 13.31 | 14.5 | 88,998 |
| -2.0 | 57 | 1536 | 1536 | [16] 0.2, 0.26, 0.34, 0.44, 0.57, 0.74, 0.97, 1.26, 1.63, 2.12, 2.76, 3.58, 4.66, 6.06, 7.88, 10.24 | 11.4 | 1,30,743 |
| -2.2 | 46 | 1536 | 1536 | [12] 0.2, 0.26, 0.34, 0.44, 0.57, 0.74, 0.97, 1.26, 1.63, 2.12, 2.76, 3.58 | 4.4 | 1,53,520 |
| -2.5 | 34 | 1536 | 1536 | [3] 0.2, 0.26, 0.34 | 0.4 | 1,24,058 |

Table 2.1: *Scale-free simulation catalogue*: Column 1: Power-law power spectrum index for the model, Column 2: Initial redshift used to start the simulation, Column 3: Side length of the cubical simulation box, Column 4: Cube root of the total number of particles put in the simulation, Column 5: Range of the scale of non-linearity(r_{nl}) covered in the simulation, Column 6: Maximum limit on r_{nl} considering the finite box size effect, Column 7: Number of CPU hours taken by the simulation.

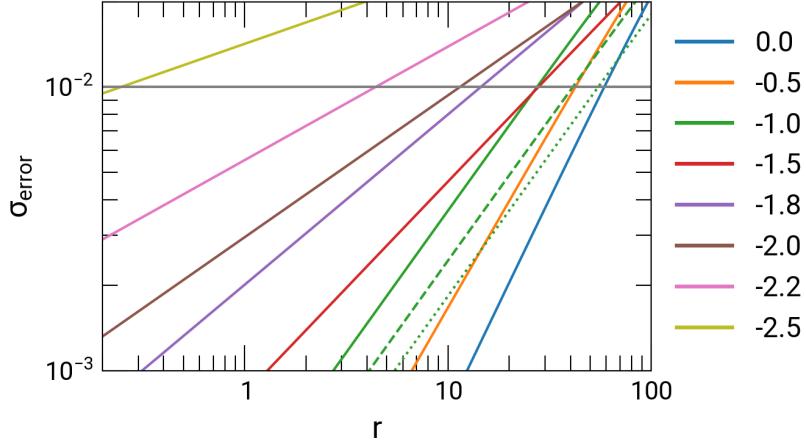


Figure 2.1: *Finite box size effect*: The plot shows the fractional error in the mass variance due to the finite box size used in models listed in table 2.1 as a function of length scale in the grid units. Different coloured solid lines represent the models with different power-law power spectrum index n . For $n = -1$, three variants of L_{box} are shown, solid 512, dashed 768 and dotted 1024. The grey line represents the 1% mark.

the volume of the simulation in grid units. Column 3 of table 2.1 displays the size of the simulation box, denoted as L_{box} , which is also equal to the cube root of the number of particles, as displayed in column 4. We conduct ten realisations of smaller boxes to account for sample variance for power-law indices $n = 0.0, -0.5$, and -1.0 . A single realisation of larger boxes is employed for other indices to cover a wider range of halo masses.

We output particle positions and velocities at different r_{nl} values listed in column 5 of table 2.1. Values inside square brackets in column 5 denote number of r_{nl} s for particular run. The epochs are chosen such that increments in $\log(r_{nl})$ are constant. We configure the spacing to the minimum necessary to ensure that outputs are considered independent.

We account for finite box size effects while determining the appropriate box size and the range of r_{nl} where results can be considered reliable. This is discussed in section §2.3.2. We set a tolerance limit of 1% error in the mass variance at the scale of non-linearity. Figure 2.1 illustrates the fractional error in the linearly extrapolated mass variance as a function of length scale in grid units, computed using the aforementioned criterion proposed by [133, 134]. The various coloured solid lines represent the models listed in table 2.1. Larger box sizes are required for models with more negative indices to maintain errors below 1%. We have verified the stability of our results against variations in the box size. Column 6 of table 2.1 provides the maximum limit on the scale of non-linearity within our criteria.

Column 7 lists the CPU time in hours taken by simulations to run. In total, the simulations have taken over 2 million CPU hours.

Figure 2.2 displays a slice of 512×96 grid units, spanning 10 grids in thickness, from N-body simulations representing scale-free models with varying power-law power spectrum

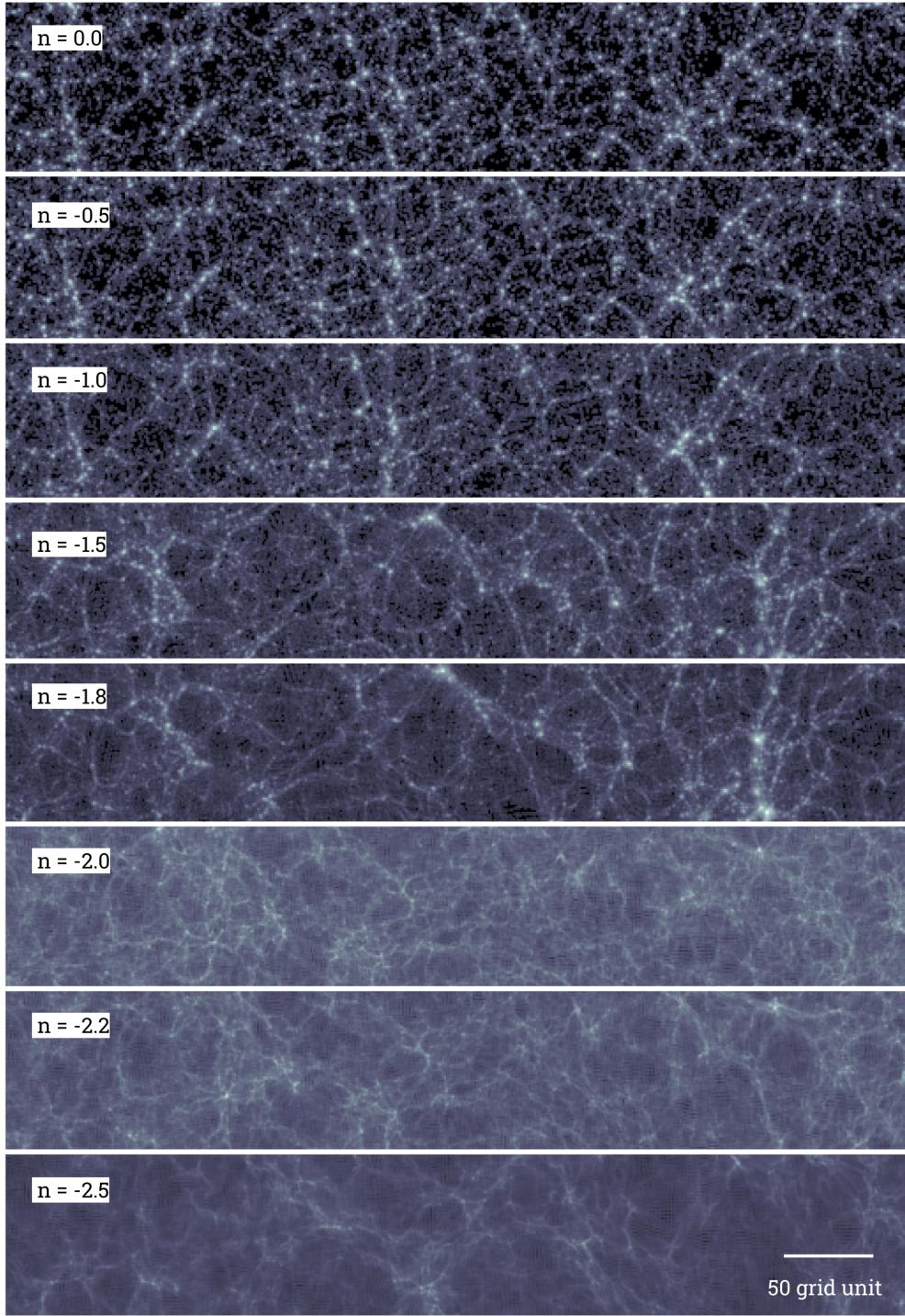


Figure 2.2: *Large scale structure in scale-free cosmologies*: A 512×96 grid units slices of simulation boxes listed table 2.1. Each slice spans ten grid units in thickness. All slices uses $r_{nl} = 10.24$. Observable differences in clustering are evident across different models and scales. Further quantification of these variations can be achieved through clustering statistics. Refer §2.4.2

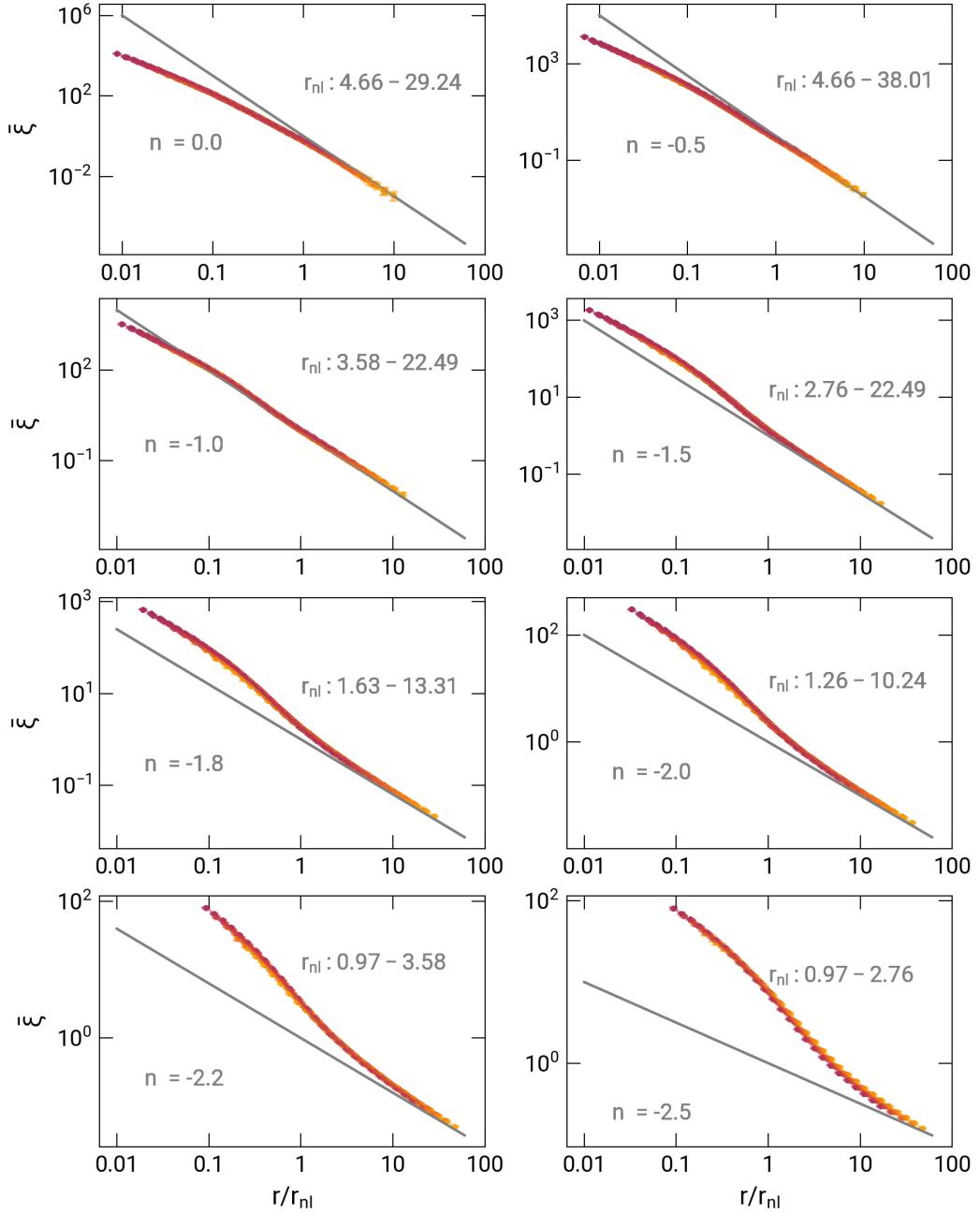


Figure 2.3: *Self-similarity of $\bar{\xi}$* : Figure shows the volume averaged two-point correlation function $\bar{\xi}$ scaled with non-linearity scale (r_{nl}) for all models listed in table 2.1. Power-law power spectrum index n is displayed on each panel to identify the models. Different coloured solid lines represent $\bar{\xi}$ at different epochs, and r_{nl} range is displayed on the panels. The grey line represents the linear theory prediction for $\bar{\xi}$. The self-similar nature of these models is evident through this figure, as scaled $\bar{\xi}$ from different epochs converge into a single curve. Further, it shows a characteristic linear, quasi-linear, and non-linear regime. Refer §2.4.2

indices. $r_{nl} = 10.24$ is chosen for this visualisation⁴. Observable differences in clustering are evident across different models and scales. Further quantification of these variations can be achieved through clustering statistics. We show the average correlation function, $\bar{\xi}$, in fig. 2.3.

When mode coupling becomes important, resulting in power transfer from large to small scales. This effect becomes significant as the spectral index approaches -3 , where all modes possess equal power. To mitigate errors involved in such effects, we employ self-similar models which allow us to validate the simulations [211]. Thus, validating self-similar clustering evolution across epochs in these simulations is essential for the analysis presented in this thesis work. For this, we use the volume-averaged two-point correlation function, $\bar{\xi}$. We compute $\bar{\xi}$ by randomly sampling particles from the simulation output. Utilising the Python module `Corrfunc` [212], we calculate the correlation function and show the results in the fig. 2.3. Different coloured solid lines represent $\bar{\xi}$ scaled with r_{nl} , while the solid grey line denotes the linear theory prediction. In these plots, we exclude epochs where $r_{nl} > r_{nl}^{\max}$ due to the significant errors arising from finite box size effects, generating unreliable data. We omit a few initial epochs to account for the initial transient evolution in simulations. The r_{nl} range corresponding to the plots is specified in the panels.

The self-similar nature of these models is evident as $\bar{\xi}$ plotted for various epochs converge into a single curve as a function of r/r_{nl} . The $\bar{\xi}$ curves exhibit three main regimes, linear ($\bar{\xi} \gg 1$), quasi-linear ($\bar{\xi} \approx 1$), and non-linear ($\bar{\xi} \ll 1$). During the evolution of the matter density field, small-scale collapses initially enter the quasi-linear regime before transitioning into the non-linear regime. The growth rate in the quasi-linear regime closely mirrors the linear rate for $n = -1$, faster for $n < -1$, and slower for $n > -1$. The slope of $\bar{\xi}$ ultimately approaches a -2 in the non-linear regime for all indices, the same as the linear slope for $n = -1$ [138].

2.4.3 Λ CDM simulation suite

The Λ CDM (Lambda Cold Dark Matter) cosmological model is the concordant model of cosmology, taking into account the presence of dark energy (represented by Λ) and cold dark matter (CDM). In this model, the late time expansion of the Universe is primarily driven by dark energy, while cold dark matter provides the gravitational skeleton necessary for the formation of large scale structures such as galaxies and galaxy clusters. A wide range of observational data supports this model, including Cosmic Microwave Background Radiation (CMBR), large scale structure surveys, and observations of distant supernovae [44, 71, 95, 107, 108].

We run five simulations based on the Planck Cosmology parameters ($\Omega_{m0} = 0.3111$, $\Omega_{\Lambda 0} = 0.6889$, $\Omega_{b0} = 0.0490$, $H_0 = 67.66$, $\sigma_8 = 0.8102$, $n_s = 0.965$, where the subscript 0 represents values at $a = a_0 = 1$) [95]. The simulations feature different box sizes of 150, 300,

⁴Note that $r_{nl} = 10.24$ is greater than the r_{nl}^{\max} for $n = -2.2, -2.5$.

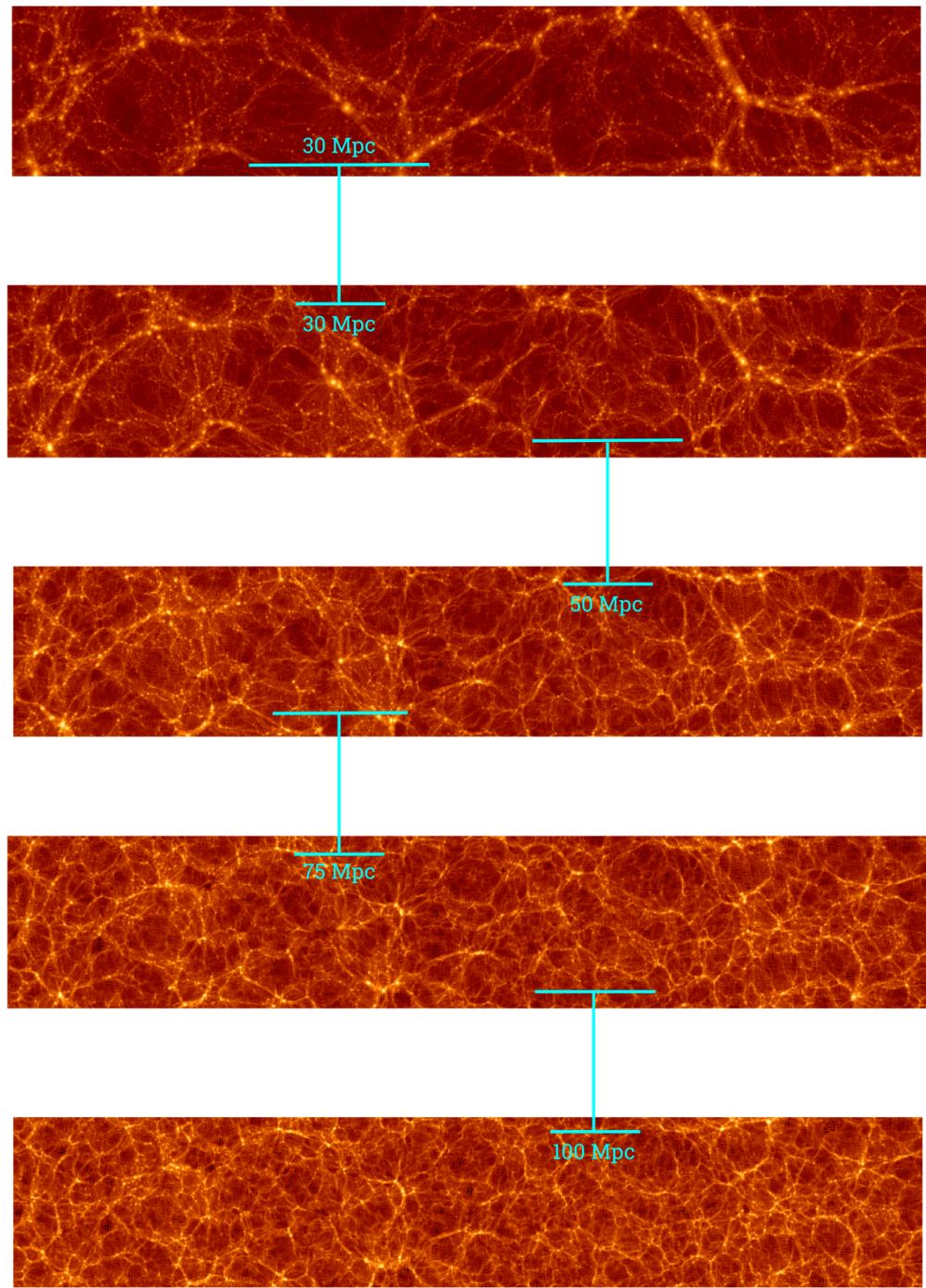


Figure 2.4: *Large scale structure in Λ CDM cosmology:* Each slice shows a 10 Mpc/h slice of different simulation boxes listed in table 2.2. The cyan markers denote different length scales in various simulation boxes. This illustration shows the hierarchical structure formation and clustering across different scales in the Λ CDM cosmology.

| L_{box} (Mpc/h) | $\sqrt[3]{N_{\text{part}}}$ | ϵ_0 (Kpc/h) | m_{part} (M_{\odot}/h) | CPU hours |
|-----------------------------|-----------------------------|-------------------------|---|-----------|
| 150 | 1024 | 5 | 2.71×10^8 | 67,500 |
| 300 | 1024 | 10 | 2.17×10^9 | 63,870 |
| 500 | 1024 | 16.6 | 1.01×10^{10} | 42,306 |
| 750 | 1024 | 24.4 | 3.39×10^{10} | 36,426 |
| 1000 | 1024 | 32.2 | 8.04×10^{10} | 25,578 |

Table 2.2: Λ CDM simulation suite: Column 1: Side length of the cubical simulation box, Column 2: Cube root of the total number of particles put in the simulation, Column 3: Softening law parameter as in eq. (2.20), Column 4: Mass of a single N-body particle in the simulation, Column 5: Number of CPU hours taken by the simulation.

500, 750, and 1000 Mpc/h, each containing 1024^3 particles, as listed in table 2.2. The smaller box simulations offer a better mass resolution, which is crucial for capturing low-mass halos and probing early epochs where massive halos are rare. Conversely, larger box simulations allow for the exploration of high-mass halos. An intermediate-sized box bridges the two ends of the halo mass spectrum.

We use GADGET4[208], operating in TreePM mode, to perform simulations. The initial conditions are generated using second-order Lagrangian perturbation theory. Similar to the scale-free simulations, we employ a force split radius of $r_s = 2$ and adaptive time steps with a maximum step size of 0.005. The softening length ϵ_0 varies across different boxes, as indicated in table 2.2. To construct the halo catalogue, we use the built-in FoF-SUBFIND algorithm [201] of GADGET4, using a linking length of 0.2 along with its default parameters.

The simulations are initialised at $z_{\text{start}} = 150$, and we record particle positions and velocities at 20 different scale factors: 0.059, 0.067, 0.077, 0.083, 0.091, 0.1, 0.111, 0.125, 0.143, 0.167, 0.2, 0.25, 0.333, 0.4, 0.5, 0.571, 0.667, 0.8, 0.909, and 1.0.

The fig. 2.4 depicts the large scale structure in these simulations, each showcasing a 10 Mpc/h slice. The cyan markers denote different length scales in various simulation boxes. This illustration shows the hierarchical structure formation and clustering across different scales in the Λ CDM cosmology. As resolution increases, structures at smaller scales become more discernible. Nonetheless, small box simulations are more susceptible to errors arising from cosmic variance for high mass halos and larger structures.

This chapter begins with a review of some statistical tools essential for large scale structure studies. Following that, we delve into the fundamentals of N-body simulations. Further, we explore the advantages of scale-free and Λ CDM cosmologies and provide a catalogue of simulations for both. These tools are the foundation for addressing problems in structure formation studies. We discuss these problems in subsequent chapters 3, 4 and 5.

3 Halo mass function in scale-invariant models

This chapter is based on the following paper: Swati Gavas et al. “Halo mass function in scale invariant models”. In: *Monthly Notices of the Royal Astronomical Society* 521.4 (Mar. 2023), pp. 5960–5971. ISSN: 1365-2966.

Understanding the theoretical foundations of the large scale structure mapped by galaxies and galaxy clusters with precision has become crucial with the advent of state-of-the-art observations. The halo model [213–215] asserts that all mass in the Universe is contained within virialised dark matter halos. This model is developed by quantifying the spatial distribution of halos and the dark matter distribution within each halo. It has been effectively used to make a wide range of predictions [216, 217] and to interpret observations [218–220].

The spatial distribution of halos is quantified using the excursion set formalism, which also underpins the theory of the halo mass function. These are essential components in various studies, such as semi-analytical theories of galaxy formation [19, 221], constraining cosmological parameters using galaxy cluster abundance [95, 222–224], calculating halo merger rates [12, 191, 225–228], gravitational lensing [229–231], and constraining non-Gaussianity in the primordial power spectrum of matter perturbations [232–234]. [235] investigates the effects of variations in the halo mass function on the modelling of matter power spectra and the structure growth parameter σ_8 , finding that varying halo abundance may provide an alternative solution to the σ_8 tension. Therefore, understanding the theory of mass functions and its validity is critical for precision cosmology.

Press-Schechter presented the first detailed theory of the mass function of collapsed objects [236]. This model identifies spherical regions in initial linear Gaussian random fields representing peaks in the density field dense enough to eventually collapse into non-linear structures. The criterion for collapse is given by the spherical collapse mechanism [30]. [116] established a formal correspondence between the number density of peaks in the matter density field and the number density of halos, providing a quantitative connection. However, their theory underestimates the mass function by a factor of two because it does not account for peaks that are part of bigger peaks. Initial numerical results were consistent with the Press-Schechter mass function theory [191, 237, 238]. [148] discussed conceptual problems with the Press-Schechter mass function, such as not accounting for mass in underdense regions, the assumption that the formalism is independent of the choice of the filter function,

and the lack of detail on how overdense regions relate to bound objects. Several attempts were made to revisit the derivation of the mass function, but none resolved these conceptual issues satisfactorily. Larger and more accurate simulations have shown clear discrepancies when compared with theory, particularly the over-prediction of halos near the characteristic mass (m_*) and under-prediction at both lower and higher masses [239–242].

[149] adopted an approach based on excursion set formalism to calculate the mass function. This method employs Markovian random walks of mass elements and a spherical up-crossing absorbing barrier to determine the number density of halos. These calculations, initially in Lagrangian space, were later transformed into Eulerian space by [243] to address issues related to the clustering of halos. This approach naturally resolves the factor-of-two discrepancy in earlier theories by utilizing a top-hat filter in k-space [244].

The non-zero tidal field around a collapsing region naturally results in an ellipsoidal collapse rather than a spherical one. Additionally, the shape of equal-density contours around peaks in the density field is often not spherical. These concepts were investigated to modify the Press-Schechter mass function [117, 239, 245]. Sheth-Tormen [117] provided a functional form for the ellipsoidal collapse barrier required in the excursion set formalism and derived the mass function from it.

It is possible to develop the Press-Schechter formalism without explicitly referencing the cosmological model, the power spectrum of matter perturbations, or the epoch. Sheth-Tormen preserves this universality of the mass function. However, these are simple theoretical predictions that do not capture the intricacies of halo formation. The accurate mass function can be obtained using cosmological N-body simulations. Studies using simulations have revealed a 5–20% systematic deviation from theoretical predictions [246–249]. Further, attempts to examine the universality of the mass function in various settings—such as different cosmologies [250], power spectra [251], epochs [252], mass definitions [253], and mass intervals [247]—have shown a systematic trend in each case. This indicates that a complete understanding requires us to go beyond the excursion set formalism [254].

Understanding the physical origin of non-universality in the mass function is a complex issue that may involve factors such as cosmology, the power spectrum, or a combination of both. The threshold density for collapse introduces cosmology dependence [255]. The Cold Dark Matter (CDM) class of power spectra has a gradually varying slope; the spectral index $n(k)$ decreases with decreasing scale (increasing k). As perturbations at smaller scales collapse earlier, the effective index of the power spectrum is smaller at earlier times and increases towards later times. Therefore, variations in the mass function and deviations from a universal mass function may arise due to the changing slope of the power spectrum, cosmology, or both. The alternative is to use fitting functions for each of these quantities derived from N-body simulations [246–250, 252, 253, 256–266]. The effect of baryons on the mass function is observed in large cosmological hydrodynamic simulations [267, 268]. [269] use a deep-learning model to study partial non-universality of the halo mass function due to cos-

mological parameters.

Since CDM spectra lack the simplicity of scale-free spectra, we approach the problem of non-universality differently. We examine departures from non-universality in the mass function for scale-free power spectra of initial fluctuations with an Einstein-de Sitter (EdS) background. This approach helps us determine if non-universal behaviour can be attributed to spectrum dependence. Our choice of cosmology does not introduce any scale into the problem, and the threshold overdensity remains constant over time in a straightforward manner. Thus, we can isolate the non-universality of the mass function that arises from the slope of the power spectrum. We provide spectrum-dependent fits for the parameters in the Sheth-Tormen mass function and demonstrate that this allows for a better fit to simulation data.

[135] explored this idea using a similar approach. The key results of their work overlap with our findings, although, as we will demonstrate, our results are robust across many variations in the analysis. A similar analysis was also conducted by [73], who observed a similar departure from the universality of the mass function.

In §3.1, we summarise the basic framework of mass function theory and our simulations setup. In §3.2, we present results from our data analysis. We extend this work to Λ CDM models in §3.3. Finally, we conclude our findings in §3.4.

3.1 Methodology

3.1.1 Mass function

The form of equations thought to be universal [191] for Press-Schechter and Sheth-Tormen mass function is given by eqs. (2.14a) and (2.14b). In eq. (2.14b) $p \approx 0.3$ and $q \approx 0.75$ are the standard values of parameters proposed by Sheth-Tormen. These mass functions can be connected to the number density of halos of a given mass (per unit comoving volume),

$$\frac{dn}{d \ln m} = \frac{\bar{\rho}}{m} \frac{d \ln \sigma^{-1}}{d \ln m} f(\nu). \quad (3.1)$$

The dependence of the mass function on cosmology comes mainly through the growth factor \mathcal{G} and small dependence from threshold density δ_{sc} . All the power spectrum dependence lies in mass variance $\sigma(m)$ (see §2.2.4, and §2.2.3). Both of these dependencies are absorbed in the definition of ν . Thus eqs. (2.14a) and (2.14b) have a universal form. The Press-Schechter mass function has only one parameter, δ_{sc} , and this is fixed by theoretical considerations. The Sheth-Tormen mass function has two more free parameters p and q . We aim to study whether these parameters are independent of the slope of the power spectrum. If the theory of mass function can be constructed in a truly universal fashion, then these should be independent of the power spectrum.

| n | z_{start} | N_{box} | $\sqrt[3]{N_{\text{part}}}$ | realisations | r_{nl}^{mf} | r_{nl}^{\max} |
|------|--------------------|------------------|-----------------------------|--------------|---------------|-----------------|
| 0.0 | 453 | 512 | 512 | 10 | 7.87-29.2 | 58.7 |
| -0.5 | 270 | 512 | 512 | 10 | 6.06-38.01 | 42.3 |
| -1.0 | 160 | 512 | 512 | 10 | 4.66-22.49 | 27.4 |
| -1.5 | 96 | 1024 | 1024 | 1 | 2.76-22.49 | 27.8 |
| -1.8 | 70 | 1024 | 1024 | 1 | 2.12-13.30 | 14.5 |
| -2.0 | 57 | 1536 | 1536 | 1 | 1.25-7.88 | 11.4 |
| -2.2 | 46 | 1536 | 1536 | 1 | 0.96-3.58 | 4.4 |

Table 3.1: *Simulation Setup*: Column 1: power-law power spectrum index for the model, Column 2: Initial redshift used to start the simulation, Column 3: Side length of the cubical simulation box, Column 4: Cube root of the total number of particles put in the simulation, Column 5: Number of realisations run for the model, Column 8: Range of r_{nl} used to compute the mass function, Column 7: Maximum limit on r_{nl} considering the finite box size effect.

3.1.2 Simulations

We use a suite of dark-matter-only simulations for seven power-law power spectrum initial conditions. Power-law indices (n) of the power spectrum ($P(k) = Ak^n$) are shown in column 1 of table 3.1. We use GADGET4 [208] to run the simulation suite. These simulations are discussed in detail in §2.4.2.

Einstein-de-Sitter (EdS) is chosen as the background cosmology, it is the simplest model where scale factor grows with single slope and threshold collapse overdensity is constant. We identify epochs in terms of scale of non-linearity r_{nl} ($r_{nl} \propto a^{2/(n+3)}$) as discussed in §2.4.2.

The initial redshift z_{start} for the simulation is estimated from the epoch where the mass variance per unit length is about 10^{-2} , as discussed in §2.4.2. However, as per [171], systematic errors can arise in simulations due to initial conditions. Therefore, we have verified that the results presented here are not affected by the chosen z_{start} ; analysis is presented in appendix C.1.

Column 5 of table 3.1 lists the number of realisations run for a given model. For the first three models, we use multiple realisations of smaller boxes to account for sample variance. For others, we use a single realisation of larger boxes to cover a broad mass range.

Finite volume in N-body simulations causes significant errors, as modes greater than the box size are not considered while generating initial conditions and during the evolution. The errors in the mass variance can become arbitrarily large as the power spectrum slope approaches -3.0 . Furthermore, as demonstrated by [187], the finite box size can propagate errors in the mass function for Λ CDM cosmologies. These factors should be considered when determining the appropriate box size and the regime where the results can be regarded as reli-

able (see §2.3.2). We take it into account using the prescription provided by [134] for deciding the box size and the range of r_{nl} . Figure 2.1 demonstrate that as we move towards more steep slope models, larger box sizes must be selected to keep errors below 1%. We have confirmed the stability of our results against variations in the box size and shown in appendix C.2.

Column 6 of table 3.1 lists the r_{nl} range covered to compute the mass function. Column 7 lists the maximum limit on the scale of non-linearity within our criteria. Refer to §2.4.2 for details.

Finite box size effects induce error in non-linear spectra, where Fourier mode coupling becomes dominant. During this process, power is transferred from large to small scales; the effect becomes significant as the spectral index approaches -3 , as there is equal power in all modes in this limit. We are able to avoid errors arising from such effects by working with self-similar (power-law) models and EdS background. This approach has been used in many simulation studies to ensure that results are free from errors arising from finite box size or other effects, e.g., [211]. We require a similar self-evolution of clustering across epochs for the analysis in this study. We check for self-similarity using the volume-averaged two-point correlation function $\bar{\xi}$, demonstrated using fig. 2.3.

3.1.3 Finding halos

The inbuilt `FOF-SUBFIND` [201] algorithm of `GADGET4` is used to construct the halo catalogue with a linking length equal to 0.2 and all its default parameters. It first identifies Friends-Of-Friends (FOF) [188] groups in a given distribution and then decomposes each found object into substructures using the excursion set algorithm. We only consider isolated halos when computing the halo mass function. The halo-finding approach uses spherical overdensity (SO) to define halo masses. Hence, SO masses may differ from the FOF masses, and the finite box size effect and discreteness noise can affect halo masses at the high and low mass ends, respectively. Care is required to ensure that such effects do not creep into our analysis. We describe our approach in the following section.

3.2 Results

Here, we present the halo mass function from our simulations of power-law models and compare them with theory predictions. First, we bin the available halo catalogue with mass to compute the halo mass function. We construct adaptive bins in mass such that each bin contains a fixed number of halos. This gives us a halo count per mass bin, $dn/d\ln m$. With this, eq. (3.1) can now be rewritten as,

$$f(\nu) = \frac{6}{n+3} \frac{m}{\bar{\rho}_m} \frac{dn}{d\ln m}. \quad (3.2)$$

Note that $\bar{\rho}_m = 1$ in units used here. Equation (2.34) is used to arrive at eq. (3.2).

We fit the Sheth-Tormen mass function to our data using χ^2 minimisation with p and q as free parameters. To determine the errors in our analysis, we considered the Poisson errors associated with the number of halos in each bin. Furthermore, we assessed the suitability of our fits by employing Jackknife and Sample Variance errors, discussed in appendix C.3. We verify that using Poisson errors is a conservative approach and that the outcomes obtained using this method are equally robust as those obtained using other error types. In subsequent sections, we discuss mass function plots, tolerance to the filtering criteria, the dependence of Sheth-Tormen parameters on power-law n , and epoch dependence.

3.2.1 Mass function

We define spherical overdensity (SO) halo mass $M_{\Delta c}$ as the mass within a radius r_{Δ} inside which the average density is Δ times the critical density. We use masses with $\Delta = 200$ to compute the mass function. M_{vir} is the mass corresponding to the virial overdensity ($\Delta = 18\pi^2$). We only consider halos with more than 100 particles to avoid discreteness noise. The fixed count for bins is chosen to be 300 halos per bin. At the high mass end, the number counts of halos drop sharply and the fixed bin size requirement can create problems. The halo count at the high mass end is also affected strongly by cosmic variance. To remedy this, we remove the most massive 0.1% halos during the computation of the mass function. We discuss our choice of these values in the §3.2.3, along with variation in the best-fit values as we vary these thresholds/parameters.

Figure 3.1 and 3.2 show mass function results. The left panel shows mass function, while the middle and the right panel show the residual of the data with respect to best-fit Sheth-Tormen and standard Sheth-Tormen, respectively. Dotted orange and dashed red lines represent Press-Schechter and Sheth-Tormen mass functions, respectively. A solid black line represents the best-fit Sheth-Tormen mass function after χ^2 analysis. Blue-green data points show mass function calculated from simulations. Data points are color coded for r_{nl} , and color changes from blue to green as r_{nl} increases, as shown in the adjacent color-bar. The gray filled area in the right panel shows a standard deviation confidence interval (throughout this chapter, we quote a single standard deviation confidence interval) around the best-fit. We find the following points noteworthy in these plots:

- The best-fit curve (black) in the left panels or deviation of residual Press-Schechter and Sheth-Tormen curves in the middle panels show a systematic pattern in the shape of the mass function. Its variation with the power-law power spectrum index n can be seen.
- The right panels show data points, and the data point fit, moving away from standard Sheth-Tormen systematically. This points to the inadequacy of the Sheth-Tormen mass function in fully explaining the theory of mass function.

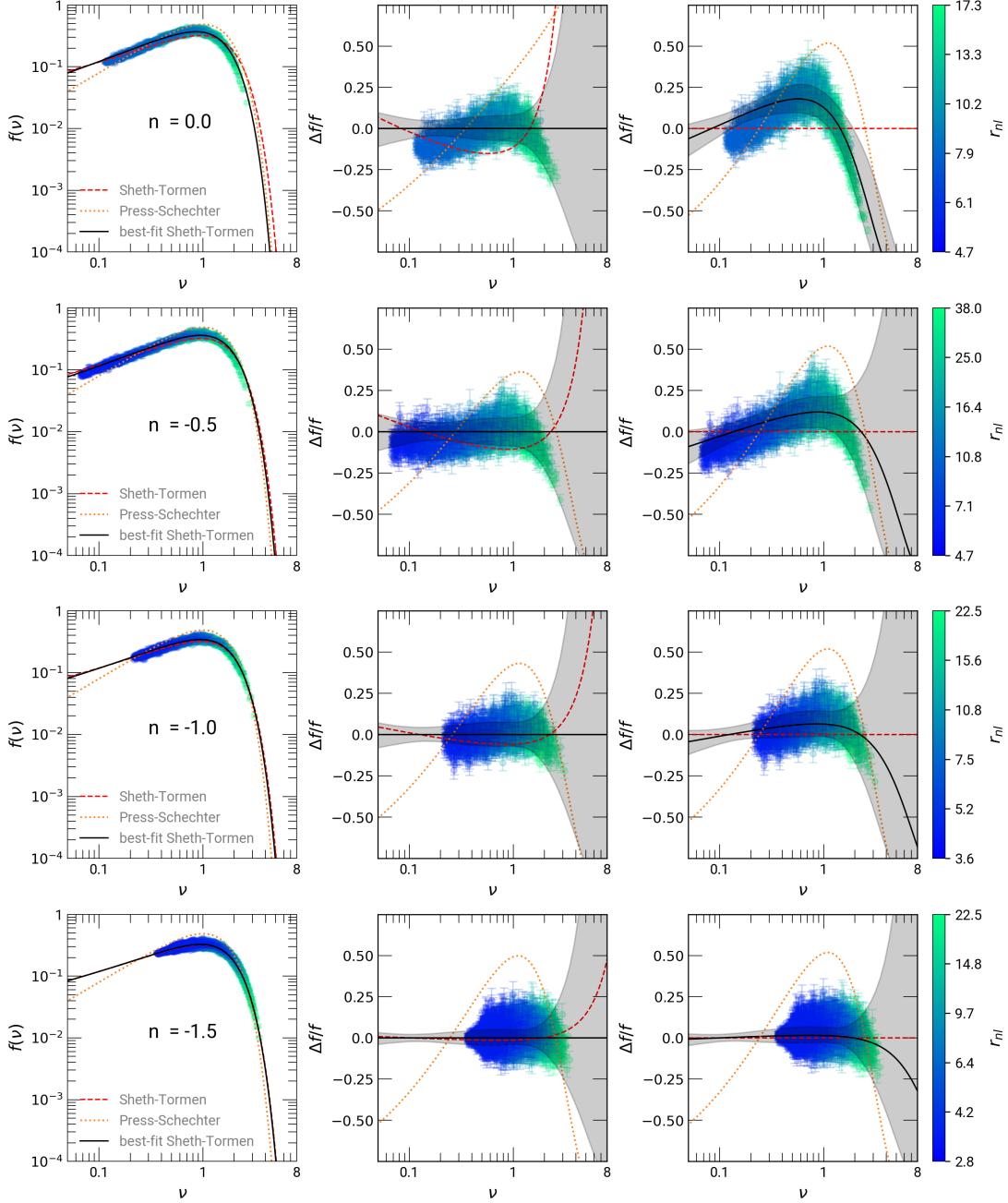


Figure 3.1: *Mass function*: The mass function from the simulations (left column), their residuals after fitting the Sheth-Tormen function (middle column) and residuals with respect to standard Sheth-Tormen function (right column). Rows 1,2,3 and 4 are for power-law indices 0.0, -0.5 , -1.0 and -1.5 respectively. Dotted orange and dashed red lines represent the Press-Schechter and the Sheth-Tormen mass function, respectively. Solid black lines represent the best-fit Sheth-Tormen mass function after χ^2 analysis. Blue-green data points show the mass function calculated from the simulations. Data points color coded with respect to r_{nl} , color changes from blue to green as r_{nl} increases as shown in adjacent color-bar. Gray filled area in the middle and right panel shows a standard deviation confidence interval for the fit.

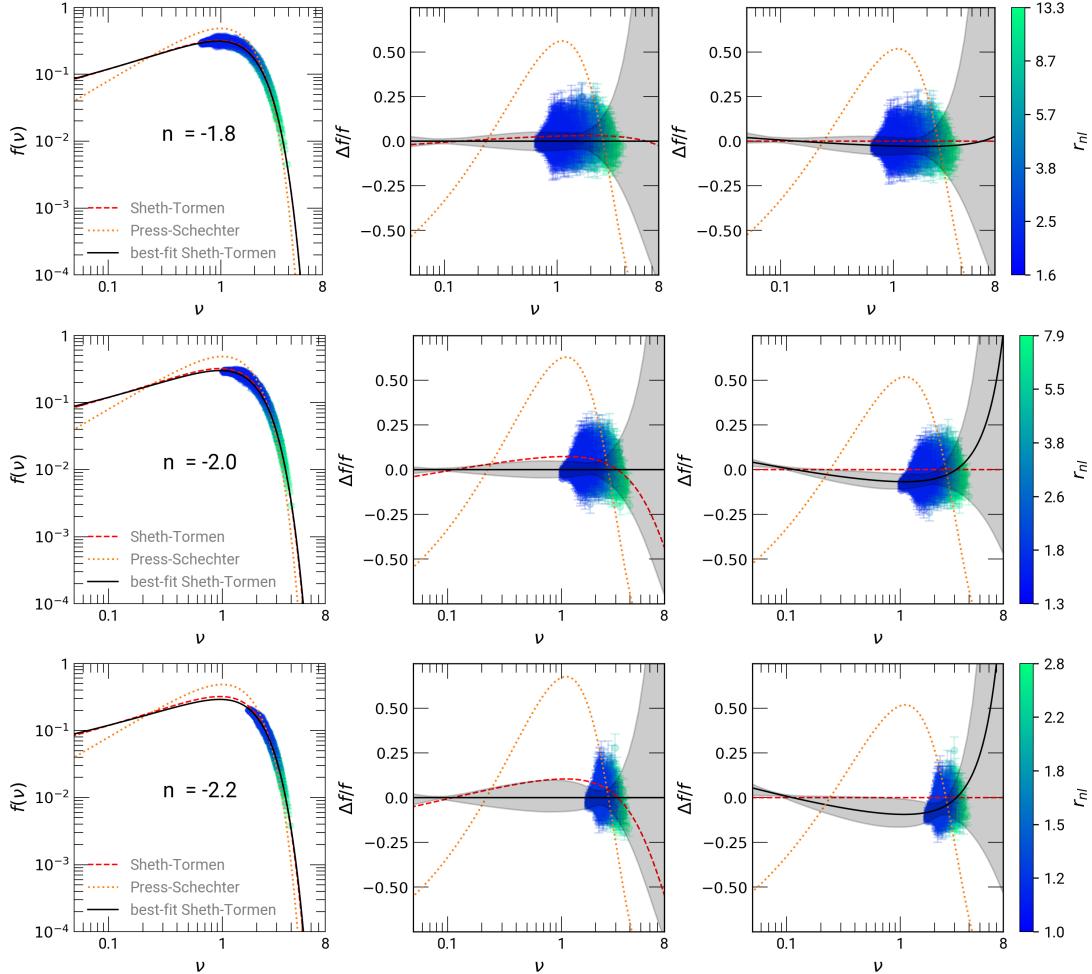


Figure 3.2: *Mass function*: The mass function from the simulations (left column), their residuals after fitting the Sheth-Tormen function (middle column) and residuals with respect to standard Sheth-Tormen function (right column). Rows 1,2 and 3 are for power-law indices -1.8 , -2.0 and -2.2 respectively. Dotted orange and dashed red lines represent the Press-Schechter and the Sheth-Tormen mass function, respectively. Solid black lines represent the best-fit Sheth-Tormen mass function after χ^2 analysis. Blue-green data points show the mass function calculated from the simulations. Data points color coded with respect to r_{nl} , color changes from blue to green as r_{nl} increases as shown in adjacent color bar. Gray filled area in the middle and right panel shows a standard deviation confidence interval for the fit.

- We use adaptive binning (fixed number of halos per bin) to construct mass function. We verify that it gives better justification to the high mass end as fits are less biased to the low mass end here compared to fixed bin width in $\log_{10}(\nu)$. However, it is still insufficient, as data points show deviations at the high mass end, especially for $n = 0, -0.5, -1.0$. One probably needs additional theory parameters to explain the high mass end. Alternative methods, such as the maximum likelihood approach by [270], can avoid binning entirely and provide a more direct comparison with theoretical predictions.
- The range of the mass function is smaller as we move towards models with more negative power-law power spectrum indices. We have not evolved simulations with these indices over a broad range of epochs considering finite box size effects. A smaller range in ν for $n = 0.0$ compared to $n = -0.5$ is due to a smaller range of r_{nl} . Evolution in the non-linear regime becomes dramatically slow as n approaches zero.
- We notice $n = -1.5$ as an index around which deviation of the standard Sheth-Tormen function from the best-fit shifts sign. This may be related to the effective index in the simulations where the Sheth-Tormen function was calibrated.
- As we progress towards late time, characteristic mass m_* increases, leading to a ν interval shift towards high values (data points changing color from blue to green).
- χ^2 is biased by the low mass end as the data points are clustered there; we see a deviation of data points at the high mass end from the best-fit curve, especially for $n = 0.0, -0.5$ and -1.0 .
- Choosing a fixed number of halo counts per mass bin leads to equal Poisson error bars; however, we see some data points have smaller error bars at the low mass end. This is because there are more halos in a particular mass interval than the fixed count value.
- The low mass end behaviour of the Sheth-Tormen mass function is decided by p , while the high mass end exponential cutoff depends on q . This leads to a large variation in mass function at the low (high) mass end due to q (p) change. In the transition region, the range of p and q in the fits is such that the Sheth-Tormen mass function does not vary much. Due to this, we see tapering of a standard deviation band around $\nu \approx 0.2$.

3.2.2 Dependence of Sheth-Tormen parameters on power-law index and epoch

We looked at the variation of best-fit Sheth-Tormen parameters p and q with power spectrum index n . We find the following approximate n dependence of p and q after fitting straight lines. The choice of a linear relation appears to be a good first approximation, as seen in fig. 3.3.

$$\begin{aligned} p(n) &\simeq -0.045 n + 0.231 \\ q(n) &\simeq 0.095 n + 0.922 \end{aligned} \tag{3.3}$$

The left and middle panels of fig. 3.3 show the relation between Sheth-Tormen parameters and power-law power spectrum index n . Errors in p and q are derived from a standard deviation confidence in the fitting. Black lines show linear fit (eq. (3.3)) to p and q scatter. Gray filled areas show a standard deviation confidence for the linear fits. Dashed red lines in the left and middle panels show standard Sheth-Tormen parameters. The right panel of fig. 3.3 shows a standard deviation contours of mass function in pq -space. Triangular data points show best-fit values from [271]. This shows a correlation between Sheth-Tormen parameters. The plot shows a clear trend of Sheth-Tormen parameters with power spectrum index n . Thus, we have clear evidence in these simulations of the dependence of the parameters of the mass function on the slope of the initial power spectrum. This indicates a departure from the universality of mass functions. This is a significant finding and requires theoretical investigations. Understanding this will lead to improved theoretical modelling of the collapse of halos and the theory of mass functions.

The propagation of errors from small to large scales can introduce transient features in N-body simulations. The sources of these errors can be mismatched mass and force resolution, discreteness effects, and stochastic noise. We study these in §5, where we explore the impact of a finite shortest scale up to which the desired power spectrum is realised in simulations. We find that the impact of small-scale cutoff in the initial power spectrum and discreteness increases with $(n + 3)$. For the $n \geq -1.0$, resolving a halo with mass function errors within 1% requires at least 300 particles (see fig. 5.1). However, in this analysis, we use a 100-particle criterion. Consequently, for $n \geq -1.0$, spurious transients likely affect earlier epochs due to the lower particle threshold we use. Hence, we see $n = 0$ as an outlier with the linear fit. If we eliminate this data point, we get a revised linear fit as in eq. (3.4). Eliminating the $n = 0$ data point does not affect the p trend at all but affects the slope of the q line significantly. Revised fit and corresponding standard deviation confidence interval is shown in the middle panel (black dashed line) of fig. 3.3. It is a better predictor for Λ CDM models, as seen in §3.3.

$$\begin{aligned} p(n) &\simeq -0.045 n + 0.231 \\ q(n) &\simeq 0.065 n + 0.859 \end{aligned} \tag{3.4}$$

We separately fit epochs available for a model to investigate the time dependence of the shape of the mass function. Figure 3.4 shows variation of p and q with r_{nl} . p tend to remain stable with r_{nl} . q varies with power-law index n ; variation in q with r_{nl} tends to increase as n approaches zero. This indicates the presence of some transient effects. Thus, we conclude that the shape of the mass function does not vary with time in a significant manner for $n < -1$.

Trends in p and q presented by [271] match qualitatively with our findings. In the present study, we have explored more models and used larger simulations. We have also studied the

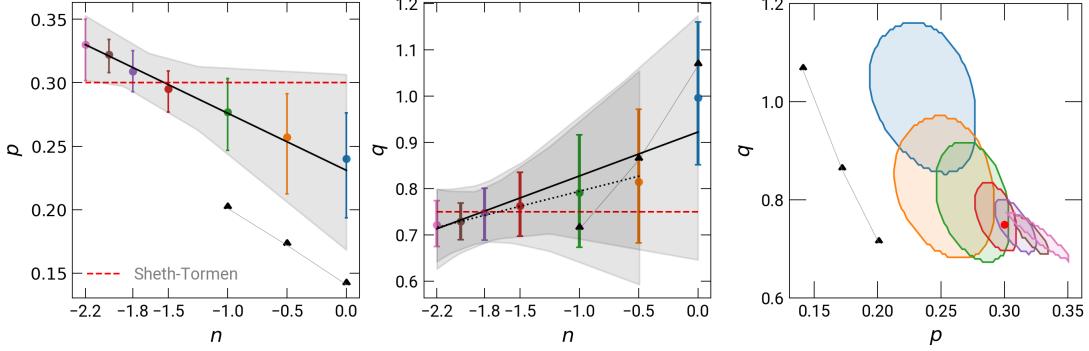


Figure 3.3: *Power spectrum dependence of the Sheth-Tormen parameters:* The variation of the Sheth-Tormen parameters with the power-law index (p : left panel, q : right panel). Errors in the p , q corresponds to a standard deviation confidence in the fitting. A solid black line represents a linear fit to the scatter with the gray area as a standard deviation confidence of this fitting. The right panel shows a standard deviation contours of the mass function fittings in the pq -space. Dashed red lines in the left and middle panels show standard Sheth-Tormen parameters, the corresponding red marker is shown in the right panel. The dotted black line in the middle panel shows a revised linear fit to the scatter. Triangular data points show best-fit values from [271].

tolerance towards various parameters used in the study, as discussed in the following section.

We notice an n -dependent slope in the p - q correlation. The possible reason could be the minimum mass cut in the simulations. At early times, halos above this cut mainly sample the exponential tail of the mass function, constraining q better than p . As the halo mass fraction increases with time, the Sheth-Tormen mass function induces a correlation between p and q [270]. This correlation slope can vary with epoch due to fixed mass cuts used here. Hence, it is possibly contributing to the observed weak trends in r_{nl} with n .

3.2.3 Tolerance to parameters

This subsection presents how the criteria used to prepare the halo catalogue obtained from simulations affect the mass function and the best fit. We test the criteria: removal of small mass halos, removal of massive halos, number of halos in each bin, and halo mass definitions.

The parameters that we vary are:

- N_{cut} is the minimum number of particles a halo must have for us to consider it in the halo catalogue used for fitting the mass function.
- Most massive halos are rare, and cosmic variance strongly affects the numbers. Thus, we remove a small fraction of halos from the high mass end of the catalogue and fit the mass function in the remaining bins. This is the second parameter used in the calculations.

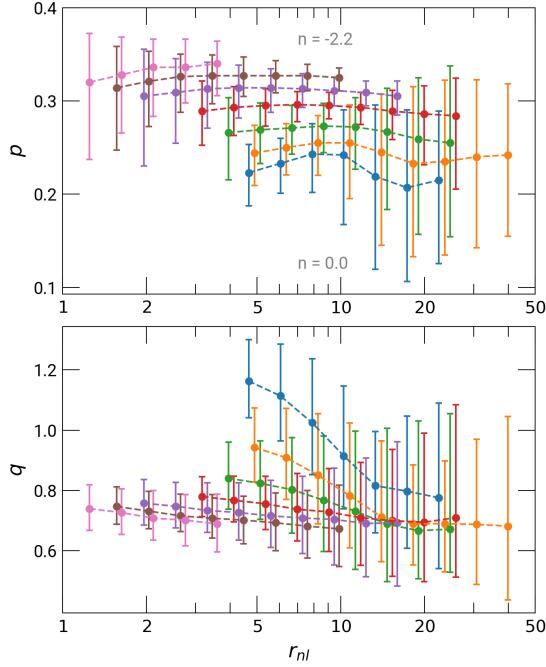


Figure 3.4: Epoch independence of the Sheth-Tormen parameters: The variation of the Sheth-Tormen parameters (p : upper panel, q : lower panel) with the scale on non-linearity (r_{nl}). Errors in the p , q corresponds to a standard deviation confidence in the fitting. Different colors used for the scatters represent different models with the power-law n as shown in the legend of fig. 2.1.

- The number of halos in an individual bin affects the size of the error bars. This is the third parameter that we have used.
- There are multiple ways to define the mass of halos. We consider three(m_{200c} , m_{vir} , m_{FOF}) possibilities in this chapter. We also study the tolerance of mass function to the choice of mass definition.

The fitting procedure gives us the best-fit values of the mass function parameters p and q and the reduced χ^2 . We study the variation of these three with our choices, and the results are shown in fig. 3.5. We have plotted curves for each of the power-law models in each panel. Columns in this figure correspond to our choices, and rows correspond to the computed values of χ^2_{red} , p and q . We note that in most panels, the values of these three parameters are stable, indicating that our choices are not critical and the results we present do not depend significantly on the choices. Specific panels that require a discussion are as follows:

- There is an increase in χ^2_{red} with the number of halos per bin for models with a steeper power spectrum slope. This appears to be caused by a reduction in the number of bins. Interestingly, there is no variation in values of other parameters p and q .

As the number of counts in a bin becomes very large, statistical fluctuations in the data become small compared to the bin size. This led to a systematic shift in the last bin.

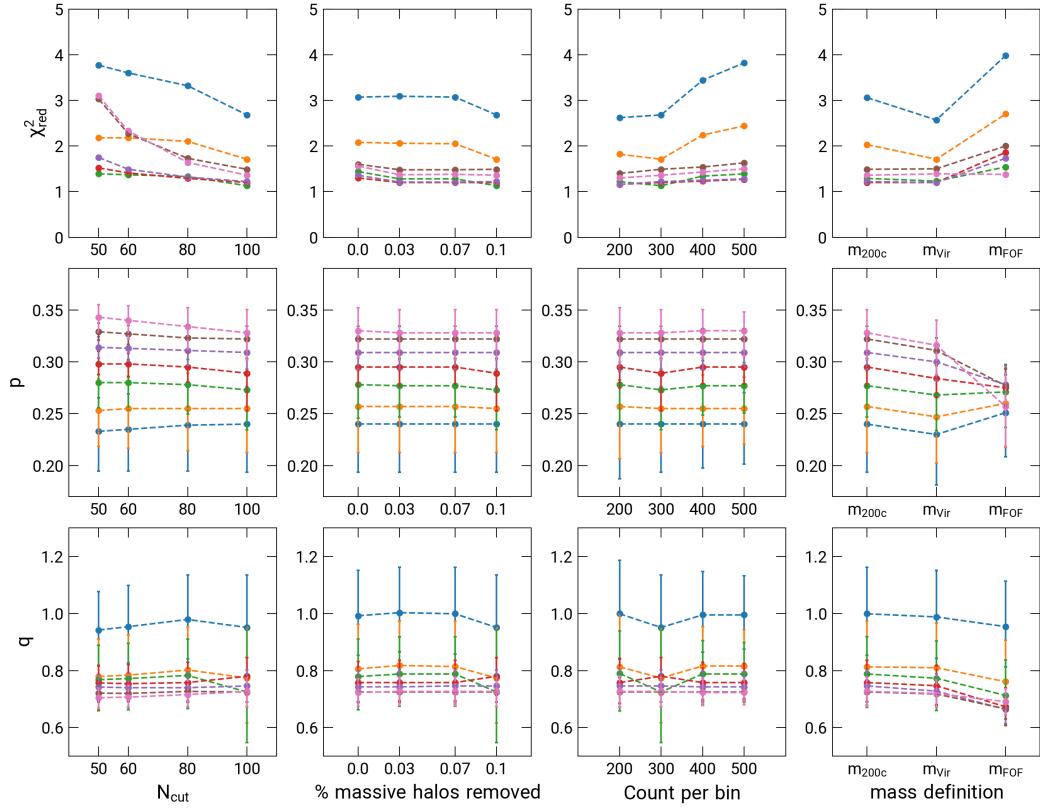


Figure 3.5: *Tolerance to parameters*: Effect of parameters like N_{cut} (column 1), percentage of massive halos removed from halo catalog (column 2), number of halos in each bin (column 3), and mass definitions (column 4) on χ^2_{red} (row 1) and fitting parameters (p : row 2, q : row 3). Different curves in each panel represent different power-law models, colors used for models are the same as shown in fig. 2.1.

Hence, we restrict our analysis to 500 halos per bin.

- The estimated variation of p with the power-law index is very small if we use m_{FOF} . Surprisingly, there is no corresponding effect in a variation of q with the index n . Notably, trends in p and q are intact if the $n = -2.2$ model is ignored.

3.3 Going beyond power-law and EdS

We have demonstrated conclusively that the mass functions for power-law models in an EdS background are not universal. Departure from universality is small but non-zero. The set of models used to demonstrate this departure from universality is simple; indeed, the simplicity allows us to demonstrate deviations from a universal behaviour. In this section, we seek to analyse how this departure from universality may affect the mass function in a more realistic model.

Departures from universality in mass function can potentially arise from cosmology via the growth factor or the threshold for collapse, or it can come from the shape of the power

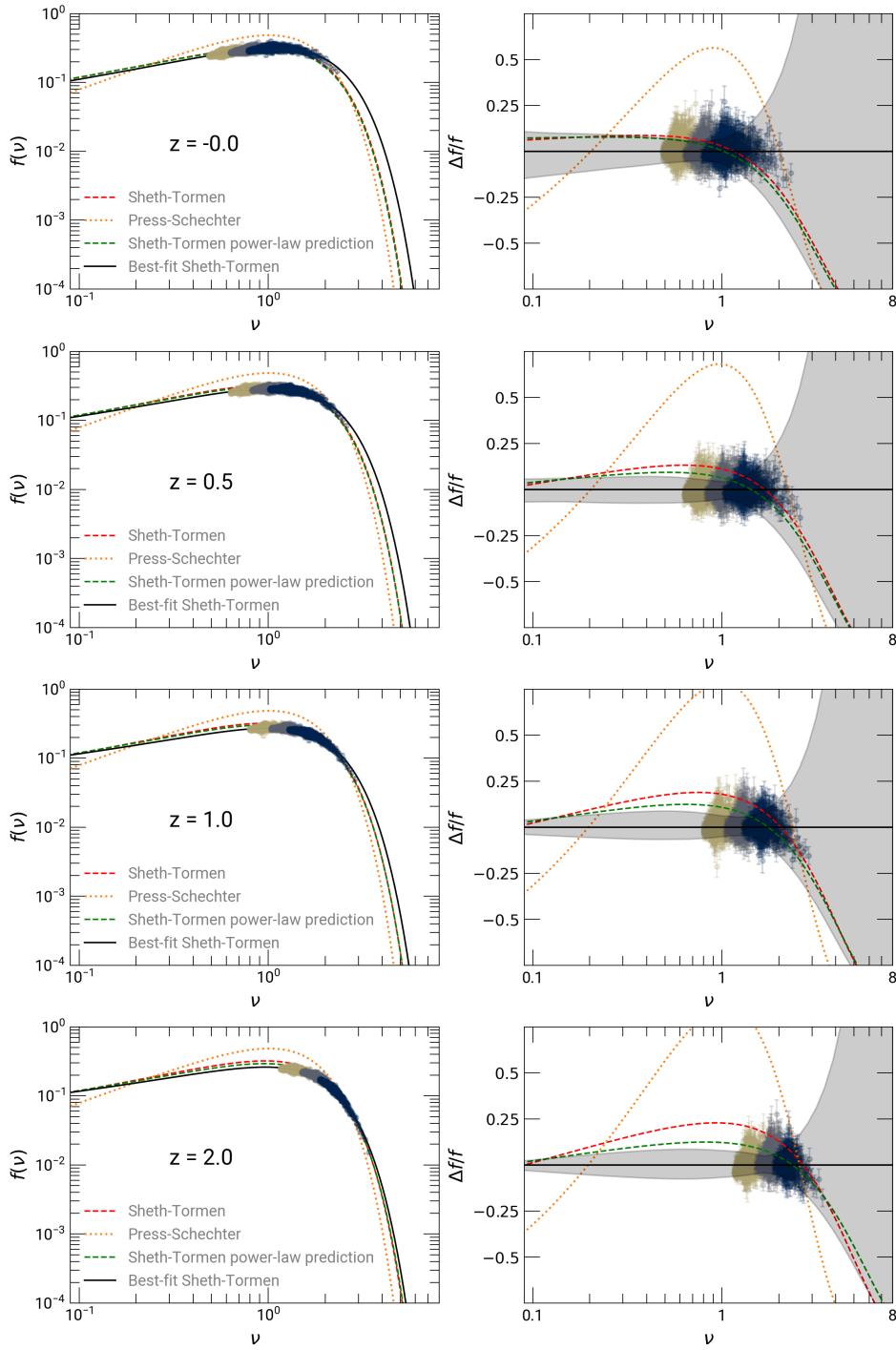


Figure 3.6: Mass function for Λ CDM models: The mass function for Λ CDM simulations (left column), and their residuals after fitting the Sheth-Tormen function (right column). Rows 1,2,3 and 4 are for redshifts 0.0, 0.5, 1.0 and 2.0 respectively. Dotted orange, dashed red, and dashed green lines represent the Press-Schechter, the Sheth-Tormen mass function, and power-law prediction using eq. (3.4). Solid black lines represent the best-fit Sheth-Tormen mass function after χ^2 analysis. Brown-gray data points show the mass function calculated from the simulations. Three different colors of data points represent different simulations used for analysis. Gray filled area shows a standard deviation confidence interval for the fit.

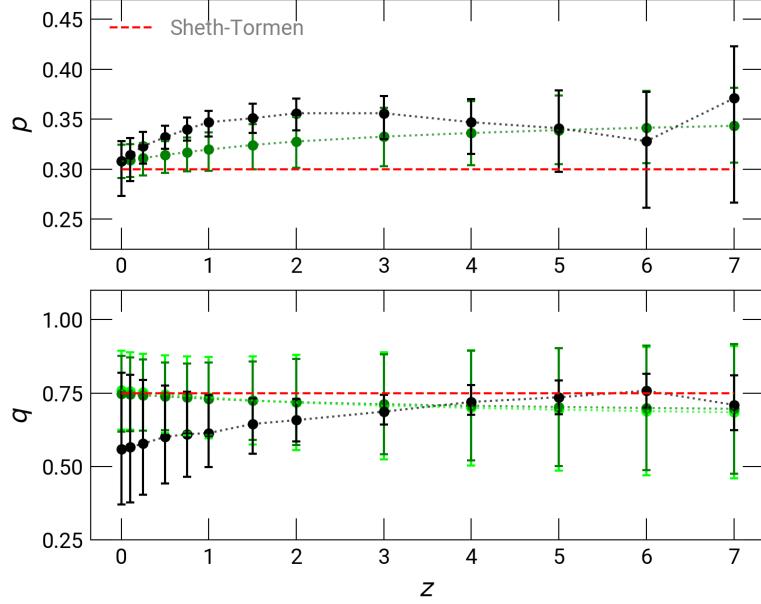


Figure 3.7: *Sheth-Tormen parameters in Λ CDM cosmology:* The figure shows the variation of the Sheth-Tormen parameters (p : upper panel, q : lower panel) with the redshift for the Λ CDM model. Black scatter represents best-fit Sheth-Tormen values while the Green and light-green data points represent predictions from power-law models with respect to eq. (3.4) and eq. (3.3) respectively. Errors in the p,q correspond to a standard deviation confidence. Dashed red lines show standard Sheth-Tormen parameters.

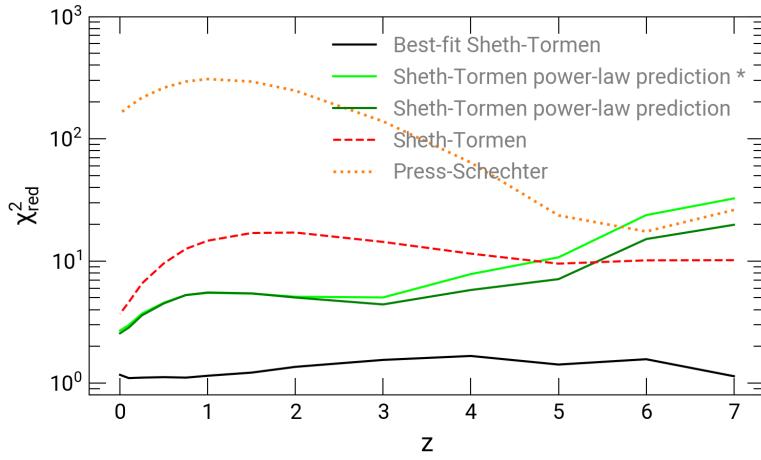


Figure 3.8: χ^2_{red} values for Λ CDM mass function fits: Black curve is χ^2_{red} for best-fit Sheth-Tormen mass function, Green and light-green curve are power-law models prediction of Sheth-Tormen mass function (see text), dashed red curve is χ^2_{red} for standard Sheth-Tormen mass function, and the dotted orange curve is χ^2_{red} for Press-Schechter mass function.

spectrum. Our study has given us a quantitative handle on the variation in mass function with the slope of the power spectrum. In this section, we use the dependence of p and q on n to estimate the expected variation of the mass function with the shape of the power spectrum. To map the varying slope of the Λ CDM power spectrum, we define an effective index at any redshift z by computing

$$\frac{d \log \sigma(m, z)}{d \log m} \Big|_{\sigma=1} = -\frac{n_{\text{eff}} + 3}{6}. \quad (3.5)$$

Then, we fit free Sheth-Tormen to Λ CDM simulations epoch-wise and compare the best-fit p and q with eq. (3.3). We use three Planck18 Cosmology [95] simulations with box sizes 150, 300 and 500 Mpc/h with 1024^3 particles, described in table 2.2. Small box simulations provide better mass resolution, essential to capture low mass halos. Further capturing low mass halos allows us to probe early epochs, as massive halos are rare. On the other hand, large box simulations provide us with high mass halos. We use one intermediate box to connect both mass ends. We calculate mass functions using the M_{200b} halo mass definition, 300 halos per bin, $N_{\text{cut}} = 100$, and removing the top 0.1% of massive halos from each box. We verify that the mass function has converged with regard to box size by fitting individual boxes and comparing the results to a combined fit. Figure 3.6 shows the mass function for Λ CDM simulations (left column), and their residuals after fitting the Sheth-Tormen function (right column). Different panels are for different representative epochs. Dotted orange, dashed red, and dashed green lines represent the Press-Schechter, the Sheth-Tormen mass function, and power-law prediction using eq. (3.4). Solid black lines represent the best-fit Sheth-Tormen mass function after χ^2 analysis. Brown-gray data points show the mass function calculated from the simulations. Three different colors of data points represent different simulations used for analysis. Gray filled area shows a standard deviation confidence interval for the fit.

We show these fitted parameters, and in fig. 3.7, black data points show the best-fit Sheth-Tormen values of p and q for redshift. Green data points represent predictions from power-law models according to eq. (3.4). Light-green scatter in the bottom panel shows eq. (3.3) prediction. Errorbars represent a standard deviation errors. Dashed red lines show p and q values from standard Sheth-Tormen. We find that the trend in p for both predicted and fitted values is similar. Their actual values are off by some factor, possibly because of cosmological contributions. q in both cases is consistent within errorbars. As we increase redshift, the halo count decreases, so it is challenging to probe that part. It is not easy to probe mass function below $\nu < 1$.

In fig. 3.8 we present χ^2_{red} values of Λ CDM mass function fit with epoch. The black curve shows χ^2_{red} for the best-fit Sheth-Tormen mass function. Green and light-green curves represent power-law models prediction of Sheth-Tormen mass function from eq. (3.3) and eq. (3.4) respectively. The dashed red and dotted orange curves show χ^2_{red} for standard Sheth-Tormen and Press-Schechter mass functions.

Power-law prediction better fits simulation data over standard Sheth-Tormen for $z \leq 4.0$.

For $z > 4.0$, we have less number of data points resulting in large χ^2_{red} values. Using revised fits for p and q (eq. (3.4)), χ^2_{red} for power-law prediction improves further. This indicates the scope of improvement in Sheth-Tormen mass function.

3.4 Conclusions

We have used power-law models in the Einstein-de Sitter background to investigate the mass function of collapsed halos in cosmological N-body simulations. We fit the mass function obtained from simulations with the Sheth-Tormen form to obtain the parameters p and q . This is done for several models with different power-law indices in the $-2.2 \leq n \leq 0$. We summarise the key findings of this work below.

- We find clear evidence that the halo mass function is not universal and has explicit power spectrum dependence.
- The Sheth-Tormen parameter values that are obtained by fitting the mass function show a systematic trend with the power-law index (fig. 3.1 and fig. 3.2). We also show a correlation between these parameters (fig. 3.3).
- The halo mass function has a weak epoch dependence: the best-fit value of Sheth-Tormen parameters varies with redshift. However, the values are consistent with a constant value for a given power-law model within a standard deviation limit.
- We have shown that these results are not sensitive to the process used for constructing the halo catalogue or subsequent analysis.
- best-fit functions and the corresponding χ^2_{red} values indicate, as is also seen in the plots for residuals, that the Sheth-Tormen mass function form is inadequate. Better modelling of collapse with more parameters may be required.
- We show that the Sheth-Tormen mass function with parameter values derived from a matched power-law EdS cosmology provides a better fit to the Λ CDM mass function than the standard Sheth-Tormen mass function. We have not studied departures from universal mass function with cosmology, so we cannot rule out a similar contribution from such dependence, especially at low redshifts.

While we have not studied the cause of non-universality in mass functions, we can speculate about its origin. The variation we have seen is with the slope of the power spectrum, and in these models, the only contribution that can cause this can arise from the coupling of modes between the collapsing structures and the large scale density field. While such coupling is not significant for spherical collapse, it becomes an essential ingredient for ellipsoidal collapse. Recent studies have shown that shifting from density maxima to energy minima and positive-definite 'energy shear' tensors provides a better alignment with halo formation

in simulations [272–274]. [275] integrates excursion set theory with tidal effects to predict dark halo abundances and clustering, revealing nonlocal halo bias. The halo properties that correlate with the tidal environment are also examined [276, 277]. A detailed coupling between halo formation and large scale tidal fields can be analysed in N-body simulations, and we are investigating it.

4 Dispersion in the Hubble-Lemaître constant measurements from gravitational clustering

This chapter is based on the following manuscript: Swati Gavas, J. S. Bagla, and Nishikanta Khandai. “Dispersion in the Hubble-Lemaître constant measurements from gravitational clustering”. In: (July 2024). arXiv: 2407.10139 [astro-ph.CO]

Hubble-Lemaître constant (H_0) is a key parameter in relativistic cosmology. It encapsulates the rate of expansion of the Universe at present [278–280]. Being a dimensional quantity, it is also a measure of the age of the Universe in combination with other cosmological parameters. Measurements of H_0 at scales of a few to a few tens of megaparsecs are done using standard candles like Cepheid variables. Several distance indicators are used to extend these measurements to larger distances. Measurements of H_0 are also made using time delay measurements in gravitational lensed systems and CMBR anisotropies [281]. There is, at present, a strong tension between the measured values of H_0 using local and distant sources [95, 282–284]. This has been given the name Hubble tension [136, 280, 285]. Many possible extensions of the standard cosmological model have been proposed to account for this discrepancy, though there appears to be no single solution that accounts for all observables [286, 287]. A number of studies have also been carried out to determine whether the contribution of systematic errors can help alleviate tension [288–292].

Local inhomogeneities, such as supervoids and peculiar velocities, significantly influence the measurement of the H_0 [293–300]. Peculiar velocities, resulting from gravitational interactions between nearby galaxies, distort the smooth Hubble flow, affecting the observed recession velocities and subsequent distance measurements [301]. Despite continuous efforts to improve distance measurement techniques and methodologies, discrepancies persist [280, 302, 303], highlighting the importance of robust statistical analyses.

In this chapter, we present a study of the effect of inhomogeneities and gravitational clustering on the measurements of H_0 . Our approach is inspired by [304], where we use cosmological N-body simulations to estimate the effect of peculiar velocities on the measurement of H_0 . Similar studies have been carried out by others [305–307]. In [305], it was found that observers in random dark matter halos typically measure lower expansion rates than those in voids, with differences more pronounced in halo rest frames compared to the CMB frame.

[306] explored how H_0 measurements are influenced by factors such as survey depth, size, sky coverage, and observer position. In [307], results from linear perturbation theory were compared with non-linear velocity power spectra derived from N-body simulations, which typically show smaller variances. Although none of these approaches produces a variance large enough to explain the H_0 discrepancy fully, the differences are significant given the precision of local H_0 measurements. We use N-body simulations and consider galaxies to be embedded in dark matter halos. Considering observers to be located in galaxies, we compute the *measured* H_0 for each galaxy pair by considering the peculiar velocities. This allows us to construct and analyse the distribution for H_0 across different relative distance intervals for all observers.

The chapter is structured as follows: In §4.1, we discuss the methodology to estimate local H_0 measurements. §4.2 outlines our simulations. In §4.3, we present results derived from our data analysis. Finally, we conclude our findings and discuss future directions in §4.4.

4.1 Methodology

Assuming galaxies reside within dark matter halos, we use halos as a representative for both observers and observed galaxies. We use the distribution of halos identified in N-body simulations to create a group of hypothetical observers positioned on the halos. We assign a single galaxy or observer for each halo: this is clearly a simplification, especially for massive halos, and leads to an under-estimate the effect of peculiar velocities as we ignore the contribution of the intra-halo velocity distribution and the halo occupation distribution. We then derive the local Hubble-Lemaître constant (H_L) for observers in their neighbourhood volume by considering distances and peculiar velocities of halos within the volume. As a result, we get a set of local Hubble-Lemaître constant measurements, creating a distribution of measured values for different observers. This distribution is then compared with the global Hubble-Lemaître constant (H_G) value used in the simulation model. This approach closely follows the methodology used by [304].

The contribution of cosmological expansion and peculiar velocities is combined to obtain redshifts that are then used to compute the local value of Hubble-Lemaître's constant for a given observer. To set the notation, we define \mathbf{r}_i as the proper coordinate of the i^{th} halo. This is related to the comoving coordinate as $\mathbf{r}_i = a(t)\mathbf{x}_i$. Hence, the velocities in the two coordinates are related as $\mathbf{V}_i = \dot{a}\mathbf{x}_i + a(t)\dot{\mathbf{x}}_i = H\mathbf{r}_i + a(t)\dot{\mathbf{x}}_i$. For the present epoch where we take $a(t_0) = 1$, we can write this as: $\mathbf{V}_i = H_0\mathbf{r}_i + \dot{\mathbf{x}}_i$. The "observed" Hubble-Lemaître's by an observer at the origin is then $\mathbf{V}_i \cdot \mathbf{r}_i/r_i^2$. For a pair of halos i and k , the measured Hubble-Lemaître's is given as,

$$H_{ik} = H_0 + \frac{(\mathbf{x}_i - \mathbf{x}_k) \cdot (\dot{\mathbf{x}}_i - \dot{\mathbf{x}}_k)}{x_{ik}^2}. \quad (4.1)$$

Here, $x_{ik} = |(\mathbf{x}_i - \mathbf{x}_k)|$, \mathbf{x} represents the comoving coordinates and $\dot{\mathbf{x}}$ is the peculiar velocity.

Note that we have $a_0 = 1$ where a_0 is the present value of the scale factor. The measured Hubble-Lemaître's constant for the k^{th} galaxy is the sum of this quantity over a set of galaxies (halos). The deviation of the measured Hubble-Lemaître's constant by the k^{th} galaxy is then,

$$\delta_{H,k} = \frac{1}{N} \sum_{i=1}^N \frac{\hat{\mathbf{x}}_{ik} \cdot (\dot{\mathbf{x}}_i - \dot{\mathbf{x}}_k)}{H_G |\mathbf{x}_i - \mathbf{x}_k|}. \quad (4.2)$$

The distribution of deviations from the reference value is described using the variable δ_H defined below in eq. (4.3). δ_H is a fractional deviation of the local H_0 value from the global H_0 value

$$\delta_H \equiv \frac{H_L - H_G}{H_G}. \quad (4.3)$$

Here, N represents the total number of halos within volume V , $\hat{\mathbf{x}}_{ik}$ corresponds to the unit vector pointing in the direction of the i^{th} halo from the observer's location. Averaging over all angles clearly reduces the dispersion in values, and hence δ_k is expected to be smaller than the corresponding estimation done in some patches in the sky for an observer.

The size and the distance to the local volume affect the distribution δ . We define this local volume as a spherical shell centred on the observer and use multiple concentric shells to explore how δ varies with distance from the observer. The lower radius (r_{\min}) and upper radius (r_{\max}) of the used volumes are displayed in the sixth column of table 4.1.

While the estimation of Hubble-Lemaître's constant in shells is useful to understand the impact of peculiar velocities as a function of distance, real measurements are done over a wide range of scales using multiple distance indicators and cross-calibration. To mimic this, we also conduct mock observations over a wide range of shells to understand the impact of peculiar velocities on the measurement of H_0 in an ideal scenario. We restrict this part of the study to observers in Milky Way-size halos by mass.

4.2 Simulations

We use three dark matter-only simulations with a different cubical comoving box size: 150 Mpc/h, 500 Mpc/h, and 1000 Mpc/h, described in §2.4.3. These simulations are based on Planck18 cosmology [95]. We analyse data using the simulation output at redshift $z = 0$ for this work. The large scale structure evolved in the simulations has significant cosmic variance on the scales comparable to box size and is also affected by discreteness errors at scales comparable to the grid size. Therefore, we use three box sizes to probe a sufficient range of distance scales with improved reliability.

All three simulations use 1024^3 particles and are subject to periodic boundary conditions. To ensure the consistency of results, we maintain overlap across the simulations in the spherical shells used as local volumes. Detailed simulation information is given in table 4.1.

The simulations were carried out using the GADGET4 code [208], and the halo catalogues were produced using its built-in SUBFIND algorithm. The halo finder identifies Friends-Of-

| L_{box} (Mpc/h) | N_{part} | N_{halos} | m_{low} (M_{\odot}/h) | Spherical shells: (r_{\min} - r_{\max}) (Mpc/h) |
|-----------------------------|-------------------|--------------------|--|--|
| 150 | 1024^3 | 1,081,171 | 8.68×10^9 | (15-20), (20-25), (25-30), (30-35), (35-40), (40-45), (45-50), (50-55) |
| 500 | 1024^3 | 1,599,064 | 3.22×10^{11} | (40-55), (55-70), (70-85) (85-100), (100-115), (115-130), (130-145), (145-160) |
| 1000 | 1024^3 | 1,973,556 | 2.57×10^{12} | (135-160), (160-185), (185-210), (210-235), (235-260), (260-285), (285-310), (310-335) |

Table 4.1: *Simulation details:* Column 1: Side length of cubical simulation box, Column 2: Number of particles in the simulation, Column 3: Total number of halos in each simulation (also equal to the number of individual observers), Column 4: Lowest Halo mass, Column 5: Spherical shells used as a local volume.

Friends (FOF) groups within the given distribution. It decomposes each detected object into substructures using the excursion set algorithm. A linking length of 0.2 times the grid length is used, and all other default parameters of the code are used. Here, we focus on isolated halos, disregarding any halo part of a larger group or cluster. Substructures or sub-halos are not considered distinct entities; thus, each halo has only one observer, and additional galaxies/observers are not assigned to subhalos. Note that this leads to eliminating the finger of god effect, leading to an underestimate of the impact of peculiar velocities. Thus, our results are expected to underestimate the actual effect. We verify that including substructures increases observers by $\sim 30\%$ and increases dispersion by a few percent (see appendix D).

One of the issues that can affect our results is the difference between a single time snapshot as compared to a light cone snapshot from simulations. We can estimate the variation of peculiar velocities at large scales using the Zel'dovich approximation [5, 29]. This is especially pertinent within our approach where we do not consider sub-halos. It can be shown that the overall variation is that peculiar velocities come down over time [308], and hence, for the largest separations used here, we underestimate the effect of peculiar velocities by a few percent. Note that the peculiar velocities are underestimated by a few percent, hence the overall effect on the dispersion in Hubble constant measurements is expected to be much smaller and can be ignored at this level. Similarly, we disregard the impact of baryonic processes on peculiar velocities, as studies have shown that this effect is well below 1% at the scales considered in our analysis [309–311].

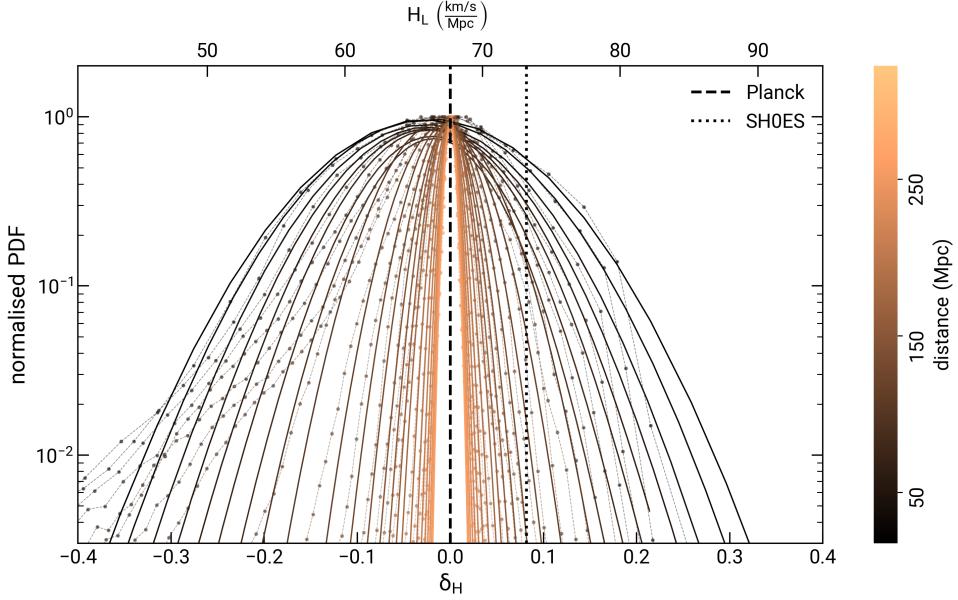


Figure 4.1: *Probability distribution functions:* normalised PDFs of the fractional deviation within a spherical local volume in redshift space. The scatter points represent the PDF obtained from the simulation data, while the solid lines represent Gaussian functions characterised by the first two moments of the PDF. The colors of lines and scatter points correspond to the distance to the local volume from the observer, as shown in the adjacent colorbar.

4.3 Results

4.3.1 The δ_H distributions

The distribution of δ_H obtained using eq. (4.3) and eq. (4.2), considering different spherical shells in the three simulations, are illustrated in fig. 4.1. Here, the distribution is obtained by averaging over all halos with a pairwise distance in redshift space corresponding to the given bin. The figure has points corresponding to the probability distribution in each bin and the Gaussian corresponding to the first two moments of the PDF. The colors of lines and scatter points correspond to the distance of local volume from the observer, as shown in the adjacent colorbar.

We observe a significant dispersion in the local H_0 values at all probed distances, with a maximum deviation on negative(positive) sides varying from $\sim 50\%(70\%)$ at 40 Mpc/h to $\sim 5\%(6\%)$ at 335 Mpc/h. As we approach larger distances, the PDFs become narrower, and the peaks shift to zero due to the increasing contribution of the Hubble flow to velocities.

PDFs show heavy negative halves, which is more evident for small-volume shells. Differences in peak location between PDFs (scatter) and Gaussian functions (solid lines) exhibit deviations of varying magnitude. These point towards a significant non-Gaussianity of the distributions.

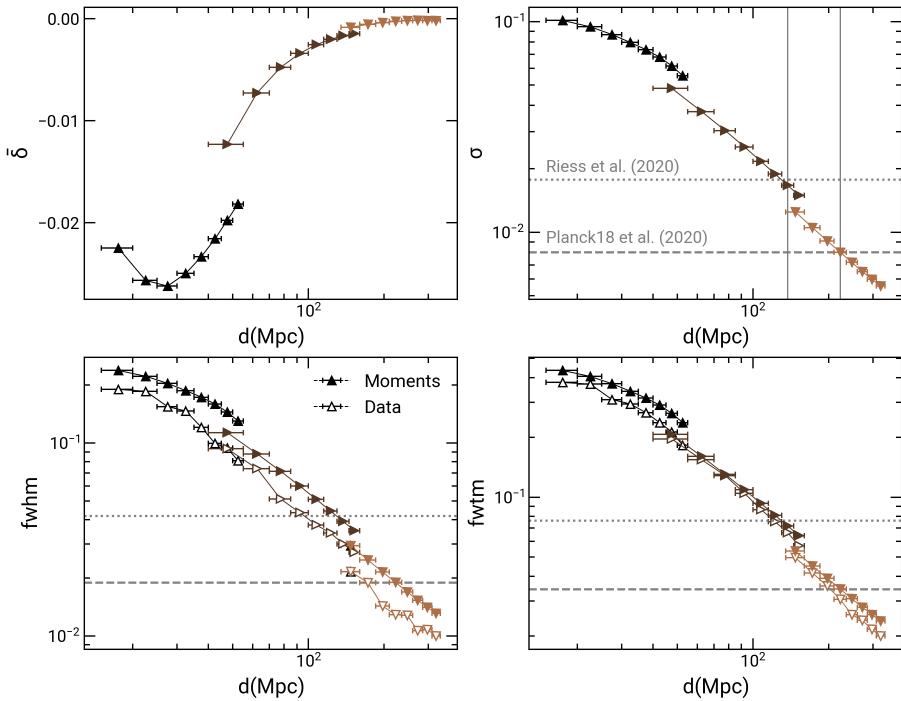


Figure 4.2: *Statistics of distribution with distance:* Mean ($\bar{\delta}$), standard deviation (σ), fwhm, and fwtm as indicated on the y-axis with distance in the x-axis. Filled markers represent quantities derived from the moments of PDF, and empty markers represent quantities computed from the discrete PDFs. The marker orientations up, right, and down correspond to simulations with box sizes of 150 Mpc/h, 500 Mpc/h, and 1000 Mpc/h, respectively. The dotted and dashed gray lines depict quantities reported by [282] and [95].

4.3.2 Statistics of distributions with distance

We analyse the δ_H distributions by calculating various statistical quantities such as mean ($\bar{\delta}$), standard deviation (σ), full-width half maximum (fwhm), and full-width tenth maximum (fwtm). In fig. 4.2, we present these quantities as a function of distance in separate panels. In the figure, filled markers represent moments and quantities derived from the moments of PDF assuming a Gaussian distribution, while empty markers represent quantities directly computed from the discrete PDFs. Up, right, and down orientations of triangular markers correspond to simulations with box sizes of 150 Mpc/h, 500 Mpc/h, and 1000 Mpc/h, respectively. Additionally, the dotted and dashed grey lines depict quantities reported by [282] and [95], providing a reference for comparison.

From the first panel of fig. 4.2, we notice the mean approaches zero from the negative side with an increase in distance. This behaviour is expected as halos reside in denser environments than the average density of the Universe. As a result, the velocities of nearby halos are slowed, leading to an overall decrease in the measured value of H_0 . This effect diminishes with an increase in the distance and size of the local volume. A jump in the halo mass

threshold causes the jump between the points from simulations of different sizes.

In the second panel of fig. 4.2, we show the variation of σ , which consistently decreases as distance increases. The parameter σ , represents the expected statistical error in the measurement of the Hubble-Lemaître constant, originating from the varying positions of observers within the large scale structure. It indicates that deviations from the global value, greater than $\sigma(d)$, in the measured Hubble-Lemaître constant H_L at a given distance d , are less likely to occur. As we investigate larger distances, this expected deviation diminishes, eventually leading to the convergence of the local (H_L) and global (H_G) Hubble-Lemaître constant values. The dotted and dashed gray lines in the plot represent the standard deviations quoted in observational measurements for local [282] and global [95] Hubble-Lemaître constant values, respectively. Given these factors and the existing errors in observational measurements, our analysis suggests a lower limit of ~ 135 Mpc/h for SH0ES and ~ 220 Mpc/h for Planck to get a robust convergence of these two Hubble-Lemaître constant values.

Given the non-Gaussian nature of distributions, we use fwhm and fwtm in addition to mean and standard deviation for better description. Panels 3 and 4 of fig. 4.2 display the behavior of fwhm and fwtm. These plots reinforce our earlier conclusion regarding the lower limit for converging local and global Hubble constant values. Furthermore, they indicate that at a distance of 100 Mpc/h, the probabilities (according to Gaussian) of encountering deviations exceeding 4% and 10% are $\sim 24\%$ and $\sim 3\%$, respectively. These probabilities steeply decline as we move to distances beyond 100 Mpc/h. Notably, the observed deviation of 8-9% in observational measurements is rare in this context.

From the fig. 4.2, we observe an offset in quantities as we transition across three simulations. The primary source of these offsets arises from the incomplete representation of the large-scale velocity field in simulations due to the finite box size. Smaller box sizes are more susceptible to this limitation, as the lack of large-scale information impacts peculiar velocities more significantly than the density field. The variation in the number of independent modes on small and large scales in simulations can be responsible for some of the contribution to the offsets in sigma between simulations. Additionally, finite mass resolution can be a secondary source of offsets across different simulations. We demonstrate this in appendix D.

4.3.3 Correlation with local density

We identify halos that exhibit a deviation of more than 5% in their distributions between 25-40 Mpc/h. Approximately 10-15% of all observers show this characteristic. To visually represent these halos, we have included 10 Mpc/h thick slices of the simulation box in the left panels of fig. 4.3. The scatter points in white denote all halos within the slice, while those in red and green correspond to halos with deviations exceeding 5%. We use the color red for halos with negative deviation from the global Hubble-Lemaître value and green for those with positive deviation. We use orange and light green colors to highlight Milky Way-sized halos, with mass ranging from 6×10^{11} to $3 \times 10^{12} M_\odot/h$. The sizes of the halo markers are

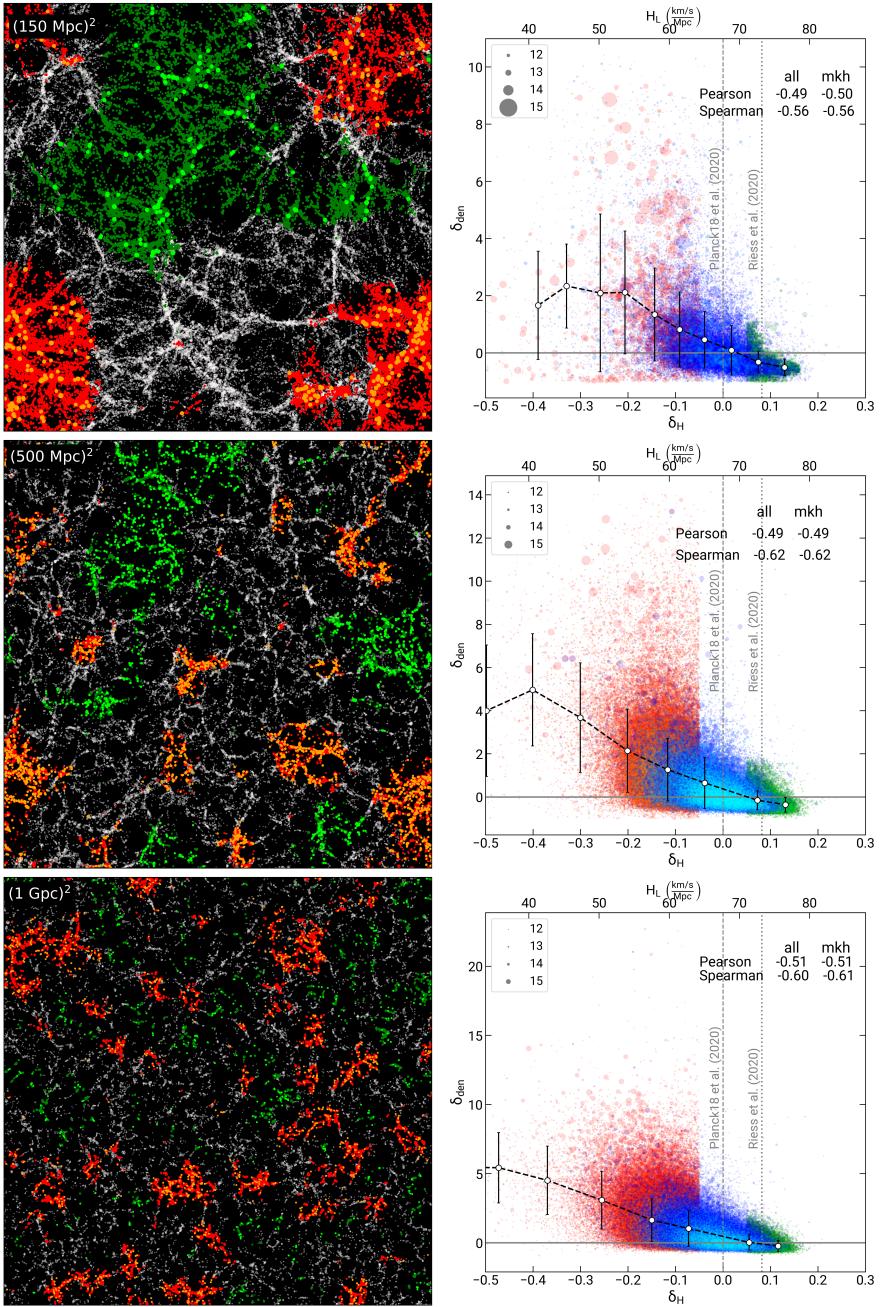


Figure 4.3: *Correlation with local overdensity.* Left panel: A 10 Mpc/h slice of simulation box showing all halos in blue color, red and green correspond to halos with deviations exceeding 5%. The color red (green) for halos with negative (positive) deviation with respect to the global Hubble-Lemaître value. The sizes of the halos are proportional to their masses, with the \log_{10} exponent indicated in the legend of the left panel. Right panel: The plot shows the local overdensity of the halos within 10 Mpc/h on the y-axis against the deviation in Hubble-Lemaître constant measurement for it on the x-axis. The mean and standard deviation in bins for three scatters(red+green, red, green) are plotted over scatters. The solid black line represents the combined scatter, while the dashed black lines, with red and green markers, represent the individual red and green scatter. The Pearson and Spearman correlation coefficient is displayed for combined scatter (§4.3.3).

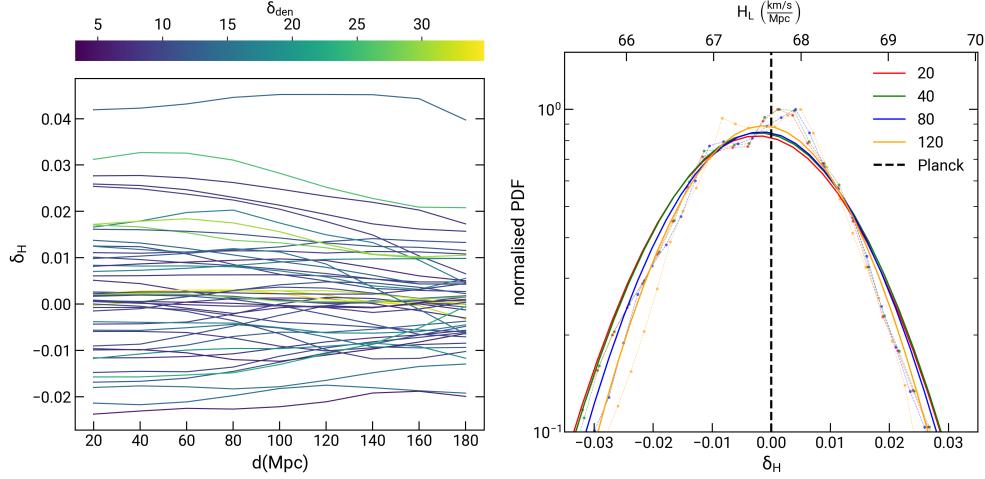


Figure 4.4: *Milky way-sized halos*: The left panel shows the deviations in local measurements for 50 randomly selected Milky Way-sized halos across various spherical shells with lower radii (r_{\min}) shown on the x-axis. The upper radius (r_{\max}) for all measurements is fixed at 200 Mpc/h. The colors of the lines represent the local overdensity of these halos, as shown in the colorbar. The right panel show normalised PDFs for Milky-sized halos within spherical shells (20-200), (40-200), (80-200), and (120-200).

proportional to the halo mass, with the \log_{10} exponent indicated in the legend of the right panels.

We see regions with a lower than global Hubble-Lemaître constant, as determined above, lie in and around clusters, whereas regions with a higher determination are away from rich clusters. To quantify this, we examine the local surroundings of these identified halos to quantify their correlation. Specifically, we calculate the local over-density (δ_{den}) of a halo within a 10 Mpc/h radius and plot this against the deviation in Hubble-Lemaître constant measurement (δ_H) for it within 25-40 Mpc/h volume. The scatter from all halos is presented in the right panels of fig. 4.3, with the same formatting as described for the left panels. We segment the data points for the all and Milky Way-sized halos into separate bins. We overplot the mean and standard deviation for each bin for the data sets. The dashed black line represents all the scatter. Further, we computed the Pearson and Spearman correlation coefficients for these. The results are tabulated in the right panel of fig. 4.3, where columns ‘all’ and ‘mkk’ represent all and Milky Way-sized halos. We find a moderate negative correlation between δ_H (deviation in Hubble-Lemaître constant) and δ_{den} (local over-density).

4.3.4 Milky Way-sized halos

The analysis presented above is limited to spherical shells. However, the observational determination of Hubble-Lemaître’s constant relies on observations across a wide range of scales, often using multiple distance indicators. To emulate this, we randomly select 50 Milky Way-

| Shell(r_{\min} - r_{\max}) | (20-200) | (40-200) | (80-200) | (120-200) |
|----------------------------------|-----------------|-----------------|-----------------|-----------------|
| 1σ | [-0.045, 0.037] | [-0.044, 0.037] | [-0.042, 0.038] | [-0.046, 0.036] |
| 2σ | [-0.042, 0.034] | [-0.042, 0.034] | [-0.040, 0.033] | [-0.039, 0.034] |
| 3σ | [-0.042, 0.034] | [-0.042, 0.034] | [-0.040, 0.033] | [-0.037, 0.032] |

Table 4.2: 1σ , 2σ and 3σ confidence intervals for PDFs shown in fig. 4.4

sized halos within our 500 Mpc/h simulation box. We ensure that these halos are not located close to each other to ensure the statistical independence of our analysis. The left panel of fig. 4.4 depicts the deviations in local measurements across various spherical shells observed by these individual observers. On the y-axis, we represent these deviations, and the x-axis denotes the lower radius (r_{\min}) of the spherical shell. The upper radius (r_{\max}) for all measurements is fixed at 200 Mpc/h. We color the trajectories of these observers according to their local overdensity, as shown in the colorbar at the top.

In the right panel of fig. 4.4, using a 500 Mpc/h simulation box, we show normalised PDFs for Milky-sized halos within spherical shells (20-200), (40-200), (80-200), and (120-200). 1σ , 2σ and 3σ intervals for these PDFs is tabulated in table 4.2. We note that the standard deviation of the distribution of δ_H is about a factor 2 higher when compared to the stated standard deviation in SH0ES. We also note that a low probability tail extends to fairly large deviations from the global value.

4.4 Conclusion and Discussion

In this work, we have considered the effect of peculiar velocities on the error budget of the measurement of Hubble-Lemaître's constant in the local Universe. We have worked with a model where we consider observers and galaxies to be located in halos. The halo distribution and velocities are obtained from N-body simulations. We make a simplifying assumption of one observer and one target galaxy per halo. This is an oversimplification as it ignores subhalos, however this is expected to lead to a conservative estimates of the contribution of peculiar velocities as we ignore internal velocities within halos and the finger of god effect.

Our key conclusions may be summarised as follows.

- We reproduce the result that the local measurements of the Hubble-Lemaître constant exhibit significant and systematic differences from the global value due to gravitational clustering and peculiar motions.
- We observe a negative correlation between the deviation of Hubble-Lemaître constant measurements from the global value and the local halo population. Observers in high-density regions measure smaller H_0 values, while those in low-density regions measure

larger values. While earlier studies indicated such a correlation [305–307], this work provides a novel result by explicitly quantifying it.

- The study suggests that, regardless of how accurately we measure the local value of the Hubble-Lemaître constant, it cannot be directly interpreted as a measure of the global value and deviations of the order of 3-5% may be expected.

We would like to note that observations have also revealed coherent bulk flow at scales in excess of 100 Mpc/h in the region around the Milky Way galaxy. This puts our region as an outlier in the expected distribution of galaxies and flows in the Λ CDM model [312–314]. An assessment of how these flows may contribute to the measurement of Hubble-Lemaître’s constant requires improved simulations as we will need to simulate a much larger physical region than we have done so far while retaining high resolution. We plan to address this issue in a follow-up study while noting that if large-scale bulk flows indicate that we are away from the average, then it may very well be that we are in the tail of the distribution for observed H_0 . Whether this can help alleviate the Hubble tension is something that can only be answered with more detailed studies.

In summary, we would like to recommend that the impact of peculiar velocities and large scale structure needs to be studied in greater detail while interpreting results from observations.

5 On the origin of transient features in cosmological N-body Simulations

This chapter is based on the following manuscript: J. S. Bagla and Swati Gavas. *On the origin of transient features in cosmological N-Body Simulations*. 2024. arXiv: 2408.05118 [astro-ph.CO]

It is believed that structures like galaxies and clusters of galaxies formed by the growth of small perturbations due to gravitational clustering [4–6, 8, 9]. The same processes are thought to be responsible for the formation of the large scale structure, i.e., the clustered distribution of galaxies in space. Dark matter is believed to be the dominant form of matter in galaxies and clusters of galaxies: dark matter is generally assumed to be made up of non-relativistic, weakly interacting massive particles (WIMPs) [95, 315, 316]. In view of the weak interactions within the dark matter sector, it responds mainly to gravitational forces. As dark matter dominates over baryonic matter in terms of total matter content, assembly of matter into halos and the large scale structure is driven mainly by gravitational amplification of initial perturbations. In the standard cosmological model, it is believed that galaxies form when gas in highly over-dense halos forms stars due to cooling, collapse and fragmentation of gas [18–20, 317–320]. Thus, the evolution of density perturbations due to gravitational clustering in an expanding universe is a key process for the study of large scale structure and its evolution. The basic equations for this can be derived from first principles [321]. These equations can be solved easily when the amplitude of perturbations is small. At this stage, perturbations at each scale evolve independently, and mode coupling is sub-dominant. Once the amplitude of perturbations at relevant scales becomes large, the coupling with perturbations at other scales becomes important and cannot be ignored, though a few results have been obtained. The equation for the evolution of density perturbations cannot be solved for generic perturbations in the non-linear regime. One can use dynamical approximations for studying mildly non-linear perturbations [9, 29, 32–37, 322–328]. Scaling relations can be used if the aim is to describe the system using statistical descriptors [151–153, 155–157, 329–334]. We require cosmological N-body simulations [35, 47, 51, 335] to follow the detailed evolution of generic systems.

Cosmological N-body simulations work with a representative region of the universe. This region is typically large but finite. Perturbations at scales smaller than the mass resolution of

the simulation, and at scales larger than the box cannot be represented or taken into account. Indeed, perturbations at scales comparable to the box are under-sampled with only a handful of independent modes. Perturbations at scales much larger than the simulation volume can affect the results of N-body simulations [133–135, 171, 336–342]. It is possible to estimate whether a given simulation volume is large enough to be representative or not [133, 134]. It has also been shown that for gravitational dynamics in an expanding universe, perturbations at small scales do not influence the collapse of large scale perturbations in a significant manner if the scale of non-linearity is larger than the smaller scales under consideration [138, 321, 343–345]. This is certainly valid for the correlation function or power spectrum at large scales. This has led to a belief that ignoring perturbations at scales much smaller than the scales of interest does not affect the results of N-body simulations. In earlier work, we have shown that if large scale collapse is highly symmetric, then the presence of perturbations at much smaller scales affects the evolution of density perturbations at large scales [346]. However, such effects are not obvious when we study the evolution of perturbations with generic initial conditions.

Substructure is known to play an important role in the dynamical relaxation of halos. It can induce mixing in phase space [347, 348], or change halo profiles by introducing transverse motions [349, 350]. Further, gravitational interactions between substructures can introduce an effective collisionality even for a collisionless system [351, 352]. It is, therefore, important to understand the role played by substructure in gravitational clustering and dynamical relaxation in the cosmological context.

The key mechanism for perturbations across scales affecting each other is mode coupling. We know the following from earlier work.

- Perturbations at large scales influence perturbations at small scales in a significant manner. If the initial conditions are modified by filtering out perturbations at small scales or other modifications restricted to small scales, then mode coupling generates small scale power. If the scale of filtration is smaller than the scale of non-linearity at the final epoch, then the non-linear power spectrum as well as the appearance of large scale structure is similar to the original initial conditions without modification [138, 343–345].
- Non-linear evolution of density perturbations *drives* every model towards a weak attractor ($P(k) \simeq k^{-1}$) in the mildly non-linear regime ($1 \leq \bar{\xi} \leq 200$) [138, 353].
- If there are no initial perturbations at large scales, mode coupling generates power with ($P(k) \simeq k^4$) that grows very rapidly at early times [138]. This can be explained in many ways. The second-order perturbation theory as well as momentum and mass conserving motion of a group of particles provide an adequate explanation. The k^4 tail can also be derived from the full non-linear equation for density [4, 321, 354].
- In case the large scale perturbations are highly symmetric, e.g. planar, then small scale

fluctuations play an important role in the non-linear evolution of perturbations at large scales [346].

The effect of large scales on small scales is clearly significant, particularly if the larger scales are comparable to the scale of non-linearity. On the other hand, the effect of small scales on larger scales is known to be small in most situations. Though this effect has not been studied in detail, many tools have been developed that exploit the presumed smallness of the influence of small scales on large scales [150, 355, 356].

Pre-initial conditions refer to the distribution of particles that is laid out before imposing initial density and velocity perturbations. The choice of pre-initial conditions is known to impact the results [167, 171, 184, 357–366]. Pre-initial conditions are expected to have no density perturbations or symmetry. It is clear, however, that at least one of these requirements must be relaxed in practice. This leads to growth of some modes deviating from expectations in perturbation theory. The work presented here allows us to estimate the effect such discrepant modes can have on the non-linear evolution of clustering across scales.

The representation and evolution of perturbations at small scales depend strongly on the mass and force resolution in the simulation. A high force resolution without a matching mass resolution can lead to better modelling of dense halos but gives rise to two body collisions and misleading results in some situations [182, 183, 367–372]. Further, discreteness and stochasticity limit our ability to measure physical quantities in simulations [51, 371, 373]. The question we address here is: can these errors remain significant at late times, and, can these errors propagate to larger scales through mode coupling? Propagation of these errors can potentially introduce transient features in N-body simulations that are likely to remain dominant till the mass scale of non-linearity is well resolved.

We present our analysis in two parts. In §5.1, we discuss mode coupling and the expected effect of collapsed halos at small scales on perturbations at much larger scales. In §5.2, we present an analysis of the effect of unrepresented perturbations at small scales. We also present an analysis of the same using N-body simulations. This is followed by discussion and summary in §5.3.

5.1 Mode Coupling

The evolution of density perturbations in a system of particles interacting gravitationally in an expanding Universe [4] can be written as,

$$\ddot{\delta}_{\mathbf{k}} + 2\frac{\dot{a}}{a}\dot{\delta}_{\mathbf{k}} = A_{\mathbf{k}} - B_{\mathbf{k}}, \quad (5.1)$$

$$A_{\mathbf{k}} = \frac{1}{M} \sum_j m_j \left[\iota \mathbf{k} \cdot \left(-\frac{\nabla \phi_j}{a^2} \right) \exp(\iota \mathbf{k} \cdot \mathbf{x}_j) \right], \quad (5.2)$$

$$B_{\mathbf{k}} = \frac{1}{M} \sum_j m_j (\mathbf{k} \cdot \dot{\mathbf{x}}_j)^2 \exp[\iota \mathbf{k} \cdot \mathbf{x}_j]. \quad (5.3)$$

The usual linear term in $\delta_{\mathbf{k}}$ is absorbed in $A_{\mathbf{k}}$. It has been shown that a single cluster of particles in virial equilibrium does not contribute to $A_{\mathbf{k}} - B_{\mathbf{k}}$ at wave numbers much smaller than the inverse of the cluster size. Here, we consider the effect of interaction between two clusters in order to estimate the effect of mode coupling and transfer of power from small scales to large scales.

The first mode coupling term can be written as,

$$A_{\mathbf{k}} = \frac{1}{M} \sum_j m_j \left[\iota \mathbf{k} \cdot \left(-\frac{\nabla \phi_j}{a^2} \right) \exp(\iota \mathbf{k} \cdot \mathbf{x}_j) \right] = \frac{\iota}{M} \sum_j (\mathbf{k} \cdot \mathbf{f}_j) \exp(\iota \mathbf{k} \cdot \mathbf{x}_j). \quad (5.4)$$

If there are two clusters C_1 and C_2 then the sum in RHS can be written separately for particles in two clusters,

$$A_{\mathbf{k}} = \frac{\iota}{M} \sum_{j \in C_1} (\mathbf{k} \cdot \mathbf{f}_j) \exp(\iota \mathbf{k} \cdot \mathbf{x}_j) + \frac{\iota}{M} \sum_{j \in C_2} (\mathbf{k} \cdot \mathbf{f}_j) \exp(\iota \mathbf{k} \cdot \mathbf{x}_j), \quad (5.5)$$

representing the force acting on particles in halo '1' due to halo '2' by \mathbf{f}^{12} , and the force due to particles within the same halo by \mathbf{f}^{11} , we can rewrite the sum¹ as,

$$\begin{aligned} A_{\mathbf{k}} &= \frac{\iota}{M} \sum_{j \in C_1} (\mathbf{k} \cdot \mathbf{f}_j^{11}) \exp(\iota \mathbf{k} \cdot \mathbf{x}_j) + \frac{\iota}{M} \sum_{j \in C_1} (\mathbf{k} \cdot \mathbf{f}_j^{12}) \exp(\iota \mathbf{k} \cdot \mathbf{x}_j) \\ &\quad + \frac{\iota}{M} \sum_{j \in C_2} (\mathbf{k} \cdot \mathbf{f}_j^{22}) \exp(\iota \mathbf{k} \cdot \mathbf{x}_j) + \frac{\iota}{M} \sum_{j \in C_2} (\mathbf{k} \cdot \mathbf{f}_j^{21}) \exp(\iota \mathbf{k} \cdot \mathbf{x}_j). \end{aligned} \quad (5.6)$$

If the wave numbers of interest are such that $|\mathbf{k} \cdot \mathbf{r}| \ll 1$, then we can expand the exponential in each term,

$$\begin{aligned} A_{\mathbf{k}} &= \frac{\iota}{M} \sum_{j \in C_1} (\mathbf{k} \cdot \mathbf{f}_j^{11}) \left[1 + \iota (\mathbf{k} \cdot \mathbf{x}_j) + \mathcal{O}(k^2) \right] + \frac{\iota}{M} \sum_{j \in C_1} (\mathbf{k} \cdot \mathbf{f}_j^{12}) \left[1 + \iota (\mathbf{k} \cdot \mathbf{x}_j) + \mathcal{O}(k^2) \right] \\ &\quad + \frac{\iota}{M} \sum_{j \in C_2} (\mathbf{k} \cdot \mathbf{f}_j^{22}) \left[1 + \iota (\mathbf{k} \cdot \mathbf{x}_j) + \mathcal{O}(k^2) \right] + \frac{\iota}{M} \sum_{j \in C_2} (\mathbf{k} \cdot \mathbf{f}_j^{21}) \left[1 + \iota (\mathbf{k} \cdot \mathbf{x}_j) + \mathcal{O}(k^2) \right] \\ &= \frac{\iota}{M} \mathbf{k} \cdot \left[\sum_{j \in C_1} \mathbf{f}_j^{11} + \sum_{j \in C_1} \mathbf{f}_j^{12} + \sum_{j \in C_2} \mathbf{f}_j^{22} + \sum_{j \in C_2} \mathbf{f}_j^{21} \right] \\ &\quad - \frac{1}{M} \mathbf{k} \cdot \left[\sum_{j \in C_1} \mathbf{f}_j^{11} (\mathbf{k} \cdot \mathbf{x}_j) + \sum_{j \in C_1} \mathbf{f}_j^{12} (\mathbf{k} \cdot \mathbf{x}_j) + \sum_{j \in C_2} \mathbf{f}_j^{22} (\mathbf{k} \cdot \mathbf{x}_j) + \sum_{j \in C_2} \mathbf{f}_j^{21} (\mathbf{k} \cdot \mathbf{x}_j) \right] \\ &\quad + \mathcal{O}(k^3). \end{aligned} \quad (5.7)$$

The first and the third term at $\mathcal{O}(k)$ are zero as there is no net internal force. Second and fourth terms at the same order cancel by virtue of Newton's third law. Thus, there is no contribution at this order, and the lowest order contribution is at $\mathcal{O}(k^2)$ from this term. The

¹Note that there is no self-interaction implied, each particle experiences a force due to all the others.

external force may be assumed to be constant across the cluster². This allows us to simplify the second and fourth terms,

$$\begin{aligned} A_{\mathbf{k}} &\simeq -\frac{1}{M}\mathbf{k} \cdot \left[\sum_{j \in C_1} \mathbf{f}_j^{11}(\mathbf{k} \cdot \mathbf{x}_j) + \sum_{j \in C_1} \mathbf{f}_j^{12}(\mathbf{k} \cdot \mathbf{x}_j) + \sum_{j \in C_2} \mathbf{f}_j^{22}(\mathbf{k} \cdot \mathbf{x}_j) + \sum_{j \in C_2} \mathbf{f}_j^{21}(\mathbf{k} \cdot \mathbf{x}_j) \right] \\ &\simeq -\frac{1}{M}\mathbf{k} \cdot \left[\sum_{j \in C_1} \mathbf{f}_j^{11}(\mathbf{k} \cdot \mathbf{x}_j) + \mathbf{F}^{12}(\mathbf{k} \cdot \mathbf{X}_1) + \sum_{j \in C_2} \mathbf{f}_j^{22}(\mathbf{k} \cdot \mathbf{x}_j) + \mathbf{F}^{21}(\mathbf{k} \cdot \mathbf{X}_2) \right] \\ &= -\frac{1}{M}\mathbf{k} \cdot \left[\sum_{j \in C_1} \mathbf{f}_j^{11}(\mathbf{k} \cdot \mathbf{x}_j) + \sum_{j \in C_2} \mathbf{f}_j^{22}(\mathbf{k} \cdot \mathbf{x}_j) \right] - \frac{1}{M}\mathbf{k} \cdot \mathbf{F}^{12}(\mathbf{k} \cdot (\mathbf{X}_1 - \mathbf{X}_2)). \end{aligned} \quad (5.8)$$

Here \mathbf{F}^{ij} is the force due to the j^{th} cluster in the i^{th} cluster, and \mathbf{X}_i is the centre of mass of the i^{th} cluster. It is easy to see that the combination of reduced terms scales inversely with the separation of the two clusters. Terms that refer to the internal forces can also be rewritten by switching to the centre of mass frame of each cluster. We use centre of mass coordinates $\mathbf{y}_j^i = \mathbf{x}_j - \mathbf{X}_i$ for the j^{th} particle in the i^{th} cluster. We will drop the superscript as it is obvious from the context,

$$\begin{aligned} A_{\mathbf{k}} &= -\frac{1}{M}\mathbf{k} \cdot \left[\sum_{j \in C_1} \mathbf{f}_j^{11}(\mathbf{k} \cdot (\mathbf{y}_j + \mathbf{X}_1)) + \sum_{j \in C_2} \mathbf{f}_j^{22}(\mathbf{k} \cdot (\mathbf{y}_j + \mathbf{X}_2)) \right] \\ &\quad - \frac{1}{M}\mathbf{k} \cdot \mathbf{F}^{12}(\mathbf{k} \cdot (\mathbf{X}_1 - \mathbf{X}_2)) \\ &= -\frac{1}{M}\mathbf{k} \cdot \left[\sum_{j \in C_1} \mathbf{f}_j^{11}(\mathbf{k} \cdot \mathbf{y}_j) + \sum_{j \in C_2} \mathbf{f}_j^{22}(\mathbf{k} \cdot \mathbf{y}_j) \right] \\ &\quad - \frac{1}{M}\mathbf{k} \cdot \mathbf{F}^{12}(\mathbf{k} \cdot (\mathbf{X}_1 - \mathbf{X}_2)) + \mathcal{O}(k^3). \end{aligned} \quad (5.9)$$

The second step follows as the internal forces cancel out for each cluster. This is the final expression for $A_{\mathbf{k}}$, ignoring the tidal interaction terms between the two clusters.

Now, we consider the second mode coupling term,

$$B_{\mathbf{k}} = \frac{1}{M} \sum_j m_j (\mathbf{k} \cdot \dot{\mathbf{x}}_j)^2 \exp[\iota \mathbf{k} \cdot \mathbf{x}_j]. \quad (5.10)$$

It is evident that this term will contribute at $\mathcal{O}(k^2)$ or higher orders to the evolution of $\delta_{\mathbf{k}}$. We can again write the sum in two parts, one for each cluster,

$$\begin{aligned} B_{\mathbf{k}} &= \frac{1}{M} \sum_j m_j (\mathbf{k} \cdot \dot{\mathbf{x}}_j)^2 \exp[\iota \mathbf{k} \cdot \mathbf{x}_j] \\ &= \frac{1}{M} \sum_{j \in C_1} m_j (\mathbf{k} \cdot \dot{\mathbf{x}}_j)^2 \exp[\iota \mathbf{k} \cdot \mathbf{x}_j] + \frac{1}{M} \sum_{j \in C_2} m_j (\mathbf{k} \cdot \dot{\mathbf{x}}_j)^2 \exp[\iota \mathbf{k} \cdot \mathbf{x}_j] \end{aligned}$$

²In principle, we can evaluate the tidal interaction term as well, but that is subdominant in most situations, and we choose to ignore it here.

$$\simeq \frac{1}{M} \sum_{j \in C_1} m_j (\mathbf{k} \cdot \dot{\mathbf{x}}_j)^2 + \frac{1}{M} \sum_{j \in C_2} m_j (\mathbf{k} \cdot \dot{\mathbf{x}}_j)^2 + \mathcal{O}(k^3). \quad (5.11)$$

We can now switch to the centre of mass frame for each cluster. Let \mathbf{V} be the velocity of the centre of mass and $\dot{\mathbf{y}}_j$ the velocity of the j^{th} particle in this frame,

$$\begin{aligned} B_{\mathbf{k}} &= \frac{1}{M} \sum_{j \in C_1} m_j [\mathbf{k} \cdot (\dot{\mathbf{y}}_j + \mathbf{V}_1)]^2 + \frac{1}{M} \sum_{j \in C_2} m_j [\mathbf{k} \cdot (\dot{\mathbf{y}}_j + \mathbf{V}_2)]^2 + \mathcal{O}(k^3) \\ &= \frac{1}{M} \sum_{j \in C_1} m_j [(\mathbf{k} \cdot \dot{\mathbf{y}}_j)^2 + (\mathbf{k} \cdot \mathbf{V}_1)^2 + 2(\mathbf{k} \cdot \dot{\mathbf{y}}_j)(\mathbf{k} \cdot \mathbf{V}_1)] \\ &\quad + \frac{1}{M} \sum_{j \in C_2} m_j [(\mathbf{k} \cdot \dot{\mathbf{y}}_j)^2 + (\mathbf{k} \cdot \mathbf{V}_2)^2 + 2(\mathbf{k} \cdot \dot{\mathbf{y}}_j)(\mathbf{k} \cdot \mathbf{V}_2)] + \mathcal{O}(k^3) \\ &= \frac{1}{M} \sum_{j \in C_1} m_j (\mathbf{k} \cdot \dot{\mathbf{y}}_j)^2 + \frac{1}{M} \sum_{j \in C_2} m_j (\mathbf{k} \cdot \dot{\mathbf{y}}_j)^2 \\ &\quad + \frac{1}{M} [M_1 (\mathbf{k} \cdot \mathbf{V}_1)^2 + M_2 (\mathbf{k} \cdot \mathbf{V}_2)^2] + \mathcal{O}(k^3). \end{aligned} \quad (5.12)$$

The total contribution to mode coupling can be written as,

$$\begin{aligned} A_{\mathbf{k}} - B_{\mathbf{k}} &= -\frac{1}{M} \mathbf{k} \cdot \left[\sum_{j \in C_1} \mathbf{f}_j^{11} (\mathbf{k} \cdot \mathbf{y}_j) + \sum_{j \in C_2} \mathbf{f}_j^{22} (\mathbf{k} \cdot \mathbf{y}_j) \right] \\ &\quad + \frac{1}{M} \sum_{j \in C_1} m_j (\mathbf{k} \cdot \dot{\mathbf{y}}_j)^2 + \frac{1}{M} \sum_{j \in C_2} m_j (\mathbf{k} \cdot \dot{\mathbf{y}}_j)^2 \\ &\quad - \frac{1}{M} \mathbf{k} \cdot \mathbf{F}^{12} (\mathbf{k} \cdot (\mathbf{X}_1 - \mathbf{X}_2)) + \frac{1}{M} [M_1 (\mathbf{k} \cdot \mathbf{V}_1)^2 + M_2 (\mathbf{k} \cdot \mathbf{V}_2)^2] \\ &\quad + \mathcal{O}(k^3). \end{aligned} \quad (5.13)$$

Using the fact that $A_{\mathbf{k}} - B_{\mathbf{k}}$ for a virialised cluster, computed in its centre of mass is zero at this order, we find,

$$A_{\mathbf{k}} - B_{\mathbf{k}} = -\frac{1}{M} \mathbf{k} \cdot \mathbf{F}^{12} (\mathbf{k} \cdot (\mathbf{X}_1 - \mathbf{X}_2)) + \frac{1}{M} [M_1 (\mathbf{k} \cdot \mathbf{V}_1)^2 + M_2 (\mathbf{k} \cdot \mathbf{V}_2)^2] + \mathcal{O}(k^3). \quad (5.14)$$

Thus, the contribution of interacting clusters to mode coupling leads to the influence of small scales on larger scales, which can be thought of in terms of clusters acting as point masses to the leading order. Here, we have ignored the subdominant tidal interactions between the clusters. Both the leading and subleading terms lead to the generation of a power spectrum that goes as k^4 at small k . This expression encodes the fact that the internal structure of virialised interacting clusters at small scales does not affect the evolution of perturbations at much larger scales to the leading order. This equation can be generalised to an arbitrary number of interacting clusters.

Going further with this analysis, especially to put these in the context of specific models is non-trivial. However, the message to take home from here is that there is a weak coupling

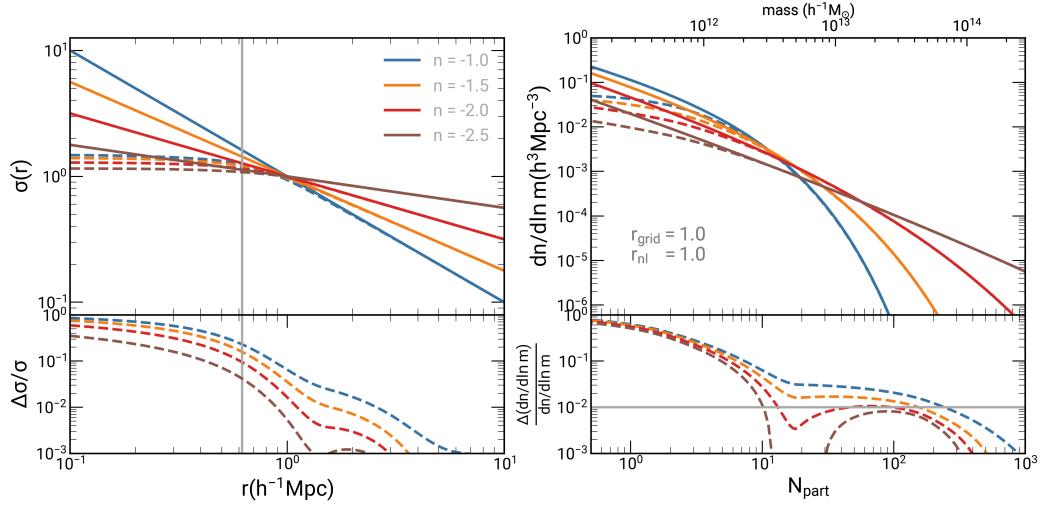


Figure 5.1: *Effect of missing small scale power on root-mean-square fluctuations and mass function:* Top left panel show mass variance for the power-law power spectrum of indices n . Solid lines are computed using full power spectrum, while dashed lines include constrained power spectrum up to Nyquist frequency. The bottom left panel show fractional errors in mass variance. The vertical grey line shows the size of a single particle. The top right panel shows the halo mass function at a time when the scale of non-linearity is equal to the unit grid length. Solid lines involve a full power spectrum, while dashed lines include a constrained power spectrum. The bottom right panel show fractional errors in mass function. Horizontal grey lines represent the 1% mark.

between non-linear structures at small scales and perturbations at very large scales. This coupling can be represented correctly in cosmological N-body simulations once halos are resolved at small scales. Therefore, we can expect simulations to match our expectations once the scale of non-linearity exceeds the mass resolution scale by more than an order of magnitude. At earlier times, there are likely to be errors in the resulting distribution of particles, and these may impact different indicators like power spectrum, mass function, etc., differently.

These expressions provide a firm footing for the ideas of the renormalisability of gravitational clustering in an expanding universe [37, 322, 328, 333, 345, 374].

5.2 Sources of errors at small scales

Cosmological N-body simulations have many potential sources of errors in the results at small scales. We have seen an estimate of the mode coupling between small scale virialised structures and large scale density fluctuations. Any errors or unrealistic representation of density fluctuations and halos at small scales will lead to such errors at large scales via mode coupling. The foremost source of errors at small scales is due to discreteness: particles in a simulation have a finite mass and it is not possible to use these to describe the density field at mass scales that are comparable to or smaller than this mass. Indeed, the density field can only be

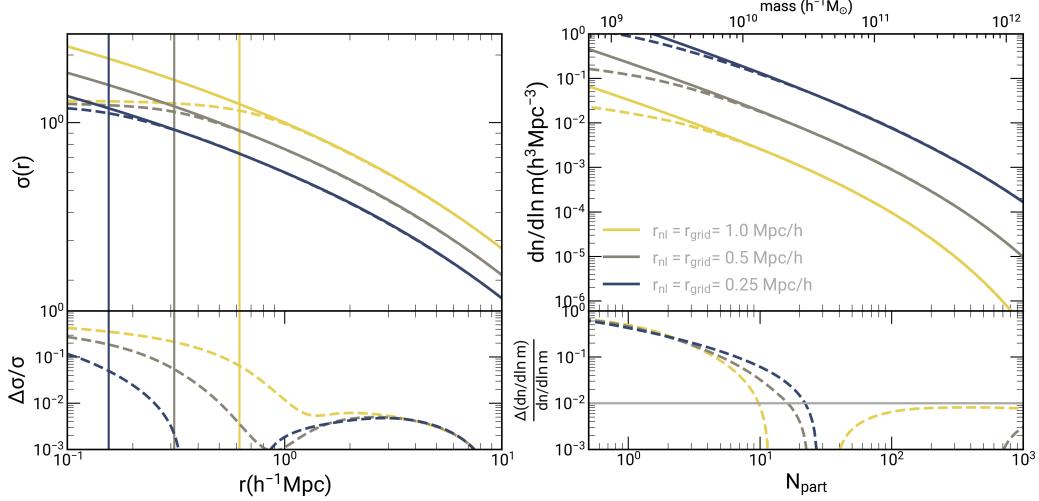


Figure 5.2: *Effect of missing small scale power for Λ CDM models* The top panels show mass variance and mass function in the Λ CDM models with three values of $r_{\text{nl}} = r_{\text{grid}} = 1.0, 0.5, 0.25 \text{ Mpc}/\text{h}$. Solid lines are computed using full power spectrum, while dashed lines include constrained power spectrum up to Nyquist frequency. The bottom panels show fractional errors in mass variance and mass function. The vertical lines in the left panels mark the size of a single particle. Horizontal grey lines represent the 1% errors.

defined at mass scales much larger than the mass of an individual N-body particle. Further, simulations cannot resolve collapsed and virialised halos of masses lower than (at least) tens of N-body particles.

To be realistic, we also need to worry about errors in inter-particle interaction force, but we drop this aspect as it depends on the specific code being used.

A corollary of the inability to represent the density field at small mass scales is that no initial perturbation is imprinted in the density distribution at these scales. This missing power translates into modified root-mean-square fluctuations at scales comparable to the grid scale in simulations. We can estimate the effect of this missing power on the mass function of halos using the approach where the root-mean-square fluctuations computed from the power spectrum realised in the initial conditions is used [134] and compared with the expected mass function without this limitation. Figure 5.1 shows the expected outcome of such a calculation. It shows the shift in root-mean-square fluctuations for power law models in the left panel and the corresponding shift in the mass function in the right panel. Here, we have chosen to plot these for the epoch where the scale of non-linearity is one grid length. The lower panels show fractional error as a result of the missing power. We notice that the root-mean-square fluctuations have an error that is larger for models with a larger $(n + 3)$, with n being the index of the power spectrum. For some models, the errors in the root-mean-square exceed 1% out to many grid cells. The corresponding errors in the mass function are also significant,

and it is clear that for some models, the error exceeds 1% for halos up to a few hundred particles. Figure 5.2 displays the case Λ CDM cosmology. The plots are presented for the epoch at which the non-linearity scale equals one grid length. As the effective spectral index is such that $(n_{\text{eff}} + 3)$ is small at very small scales in this model, the errors drop to 1% very quickly and hence the missing power is not a serious problem for Λ CDM models.

Figure 5.3 shows the mass function for three models at three epochs each. Each row is for a given index: the top row is for $n = -1$, the middle row is for $n = -1.8$, and the bottom row is for $n = -2.2$. Each row has plots for three epochs corresponding to an increasing scale of non-linearity as we go from left to right. Each panel shows the expected mass function shown in eq. (5.15) and the mass function from N-body simulations. The lower part of each panel shows the fractional error in the mass function. We note that the error is indeed highest at early epochs for $n = -1$, which corresponds to the largest $(n + 3)$ in this set. This is in agreement with the impact of the missing power at small scales estimated above. We see that the errors come down over a substantial range in mass at later epochs as a number of data points distributed around the zero-deviation mark show a systematic shift with the epoch. Some of the remaining disagreement results from a poor fitting function for the theoretical estimate, discussed in §3. Figure 3.1, fig. 3.2, and fig. 3.3 illustrate the poor fit of the halo mass function to scale-free simulations.

$$\frac{dn}{d \ln m} = \frac{\bar{\rho}}{m} \frac{d \ln \sigma^{-1}}{d \ln m} f(\nu), \quad (5.15a)$$

$$f(\nu) = A(p)\nu \sqrt{\frac{2q}{\pi}} [1 + (q\nu^2)^{-p}] \exp(-q\nu^2/2), \quad (5.15b)$$

$$p(n) = -0.045 n + 0.231, \quad (5.15c)$$

$$q(n) = 0.095 n + 0.922, \quad (5.15d)$$

where, ν is peak height $\nu \equiv \delta_c / (\mathcal{G}(z)\sigma(m))$. Here δ_c is the linearly extrapolated critical over density required for a spherically symmetric perturbation. \mathcal{G} is the growth factor. $\sigma(m)$ is the variance in the initial density fluctuation field, linearly extrapolated to the present epoch and smoothed with a top hat filter $W(k, m)$ of scale $R = (3m/4\pi\bar{\rho})^{1/3}$.

5.3 Discussion and Summary

We have seen in the last two sections that the mode coupling between small scales and large scales has a component arising from interactions of halos. This mode coupling arises from the departure of shapes of halos from spherical shapes and tidal interactions. These components clearly require the halos to be resolved in the simulations with a large number of particles. Thus, we expect that this is a source of errors in cosmology N-body simulations and that it arises from discreteness. It is reasonable to expect that the error will diminish at late times once the scale of non-linearity is sufficiently larger than the average inter-particle separation.

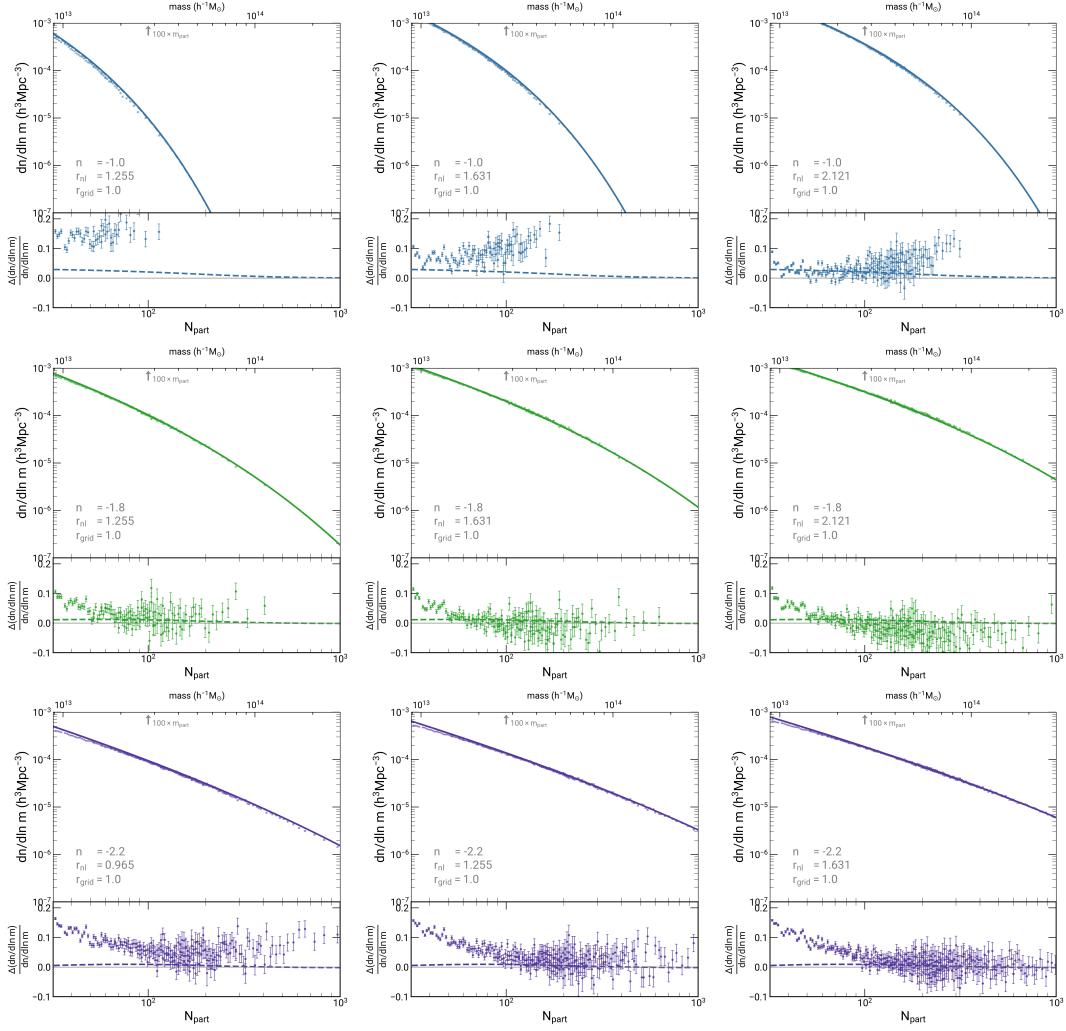


Figure 5.3: *Errors in the simulation mass function:* Each panel show the halo mass function and fractional errors in mass function as a function of a number of particles. Data points represent simulation, whereas solid lines show theoretical mass function. The top, middle and bottom rows represent the power-law power spectrum index n values -1.0 , -1.8 and -2.2 , respectively. Columns represent different scales of non-linearity (r_{nl}).

This is motivated by observations that clustering at small scales is dominated by the power spectrum around the scale of non-linearity [138, 344].

The second effect on the limitation of N-body simulations at small scales and at early times arises from missing power. It is not possible to represent density perturbations below the Nyquist frequency, and hence, there is a sharp departure from the scenario being simulated. We show that this results in an underestimation of the root-mean-square fluctuations at small scales in the simulation. Obviously, this affects models with steep power spectra much more than other models. We quantify this by computing the root-mean-square amplitude of fluctuations and the corresponding mass function of collapsed halos. We find that the errors percolate to an under-estimated root-mean-square amplitude of fluctuations out to several grid lengths and mass function for halos with more than 10^2 particle halos at early times.

We test this deficit in the mass function using N-body simulations and find that this is seen clearly at early times for the $n = -1$ power spectrum. We also note that the error diminishes at later times. The error is smaller for other models for halos with 50 or more particles, even at early times.

The summary of our findings is that there are significant errors in cosmological N-body simulations at small scales at early times. In general, the results of simulations are not reliable till the scale of non-linearity is larger than one grid length. For steeper power spectra with an index where $(n + 3) > 1$, it may be better to only use data once the scale of non-linearity is much larger than a grid length. Note that we have not considered any errors that may arise due to errors in force or numerical implementation, e.g., [167, 375]. These limitations on cosmological N-body simulations are complementary to the limitations arising from a finite box size [134].

A comparison with some of the other estimates of expected transients is also in order. [376] estimated transients arising from the deviations of initial conditions generated using the Zel'dovich approximation as compared to the expected evolution. Note that the Zel'dovich approximation gives correct evolution at linear order and hence any deviations arise from its use to set up initial conditions directly in the mildly non-linear regime. This estimation was done using higher order Lagrangian perturbation theory. We would like to point out that we have assumed that the generation of initial conditions as well as dynamical evolution is *ideal* within the limitations of mass resolution and the range of wave modes available in the simulation. Interestingly, analytical calculations using a gravitational dynamics approach also indicate a spectrum dependence in the influence of small scales on larger scales [377] where a similar factor of $(n + 3)$ makes an appearance³.

The primary concern raised in this work is that N-Body simulations are plagued by transient features introduced by two factors: discreteness induced errors in mode coupling and, missing power at small scales. We do not dwell on any other *non-ideal* aspects of evolution of perturbations in cosmological N-Body simulations. As we have shown, the errors aris-

³See end of §5 in [377].

ing out of these factors are significant at small scales and at early epochs where the scale of non-linearity is smaller than a few grid lengths. Once the scale of non-linearity is such that halos with tens of particles have collapsed and virialised, these errors become insignificant. Thus we are pointing out that the reliability of cosmological N-Body simulations is limited at smaller scales. When combined with the limitations arising from box size (See, e.g., [134]), we see that the dynamic range of cosmological simulations in length scales is about an order of magnitude smaller than the box size (in terms of number of cells). As these limitations cannot be overcome within the context of simulations, other approaches that help enhance the dynamic range of simulations must be studied. These results imply that the use of very large simulations to generate mock catalogues needs to be evaluated carefully and suitable corrections devised in order to ensure that these are reliable, e.g. [277]. Of course, as all such methods are calibrated with simulations, sufficient care and caution needs to be exercised. Approaches that suffice for some quantities of interest may fail for other observables and hence the validation of these methods has to be carried out with some care before deploying these for real applications.

One may wonder whether such errors arising at small scales will contribute to larger scales at late time. However, we know from many numerical experiments that once larger scales become non-linear these can erase differences at smaller scales [138, 344].

We may summarise the lessons from this work as follows: N-body simulations can be trusted for generic CDM/ Λ CDM type models once the scale of non-linearity exceeds the mass resolution by at least an order of magnitude. For such situations the *correctness* of cosmological N-body simulations and renormalisability of gravitational clustering can be justified.

6 Conclusion

The linear theory predictions of structure formation are based on the large scale homogeneity and isotropy of the Universe, which allows for simpler mathematical descriptions of perturbation growth. As structures grow and the amplitude of perturbations becomes large at small scales, the limitations of such assumptions become noticeable, necessitating non-linear investigations of structure formation [4, 83, 84, 155, 168, 334]. The non-universality of halo mass functions that depend on the underlying power spectrum indicates a need to go beyond the isolated ellipsoidal gravitational collapse assumption. Further, The systematic biases introduced by peculiar velocities in local Hubble-Lemaître constant measurements can arise due to inhomogeneities in the Universe. These findings in this thesis reassure that simplifying assumptions needs to be improved in capturing the details of large scale structure. Non-linear processes, such as mode coupling driven by halo interactions and the influence of large scale bulk flows, can play a role in modifying observables both on small and large scales. In simulations, discreteness and resolution errors can introduce significant biases on small scales. As large scale surveys and precision cosmology push for tighter constraints, it becomes evident that a more fine approach that accounts for mode coupling, tidal interactions, and local environmental factors is essential for accurate theoretical predictions [254, 275, 314, 378].

In this chapter, we provide a summary of our results and discuss their implications.

Summary

We first investigate the power spectrum dependence of the halo mass function using a suite of dark-matter-only N-body simulations for seven power-law models with an Einstein-de Sitter cosmology. Our findings demonstrate a clear non-universality in the mass function. We provide fits for the parameters of the Sheth-Tormen mass function across a range of power-law power-spectrum indices, observing a mild evolution in the overall shape of the mass function over epochs. We extend these results to a Λ CDM cosmology and show that the Sheth-Tormen mass function with parameters derived from a matched power-law EdS cosmology more accurately fits the Λ CDM mass function than the standard Sheth-Tormen model. These results suggest that an improved analytical theory is necessary to describe the mass function better.

We further quantify the impact of peculiar velocities on estimating the Hubble-Lemaître

constant H_0 . By considering observers located within dark matter halos, we compute the distribution of the estimated H_0 across all such observers. We find that the dispersion of this distribution is significant at small scales but diminishes with increasing separations, approaching the level of quoted statistical errors in SH0ES and Planck measurements at distances beyond 135 Mpc/h and 220 Mpc/h, respectively. A negative correlation is observed between the local over-density around an observer and the deviation between the local and global H_0 values. Deviations exceeding 5% of the global value occur frequently on scales up to 40 Mpc/h, which is considerably larger than in the range of distance scales where majority of large distance indicators like supernovae of type Ia are calibrated. We also analyse the cumulative impact of these errors on mock measurements of H_0 as deduced from Milky Way-sized halos. Our results show that this error is sensitive to the minimum distance used for measurements. The distribution of H_0 in mock measurements exhibits a large tail, suggesting that a deviation of a few percent from the global value is possible in the atypical regions of the Universe.

Lastly, we investigate the impact of small scale gravitational clustering on large scales by analysing mode coupling between virialised halos. Building on Peebles' (1974) [321] calculation, which shows that a virialised halo does not contribute mode coupling terms at small wave numbers k , we use a perturbative expansion in wave numbers to demonstrate that this effect is minor. We relate these findings to the limitations of finite mass resolution in cosmological N-body simulations on the evolution of perturbations at early times. We explore the effects of a finite shortest scale up to which the desired power spectrum is realised in simulations. We provide basic estimates of the magnitude of these effects and their dependence on the power spectrum. Our results indicate that the influence of small-scale cutoffs in the initial power spectrum and discreteness effects increases with $(n + 3)$, where n is the power spectrum index. Cosmological simulation data should be utilised only if the non-linearity scale is sufficiently larger than the average inter-particle separation.

Implications and outlook

§3 explores the test of the accuracy of theoretical modelling of the halo mass function, focusing on its dependence on the input power spectrum. The widely used assumption that the halo mass function is universal across different models has often been questioned [73, 250–253]. This challenges when predicting the abundance of halos at high masses and early redshifts, further impact galaxy formation models. The halo mass function is also used to make mock catalogues to estimate the efficacy of future observational programs and a covariance matrix to estimate errors. In the precision cosmology era, we need to ensure that underlying assumptions do not introduce errors in the system. Our long-term aim is to identify these sources of deviations, which will improve our understanding of underlying processes and enhance the precision of theoretical calculations. The observed non-universality in the mass functions may originate from mode coupling between collapsing structures and the large scale density

field, particularly in non-spherical collapse scenarios. It is important to note that the physical scenario relevant to excursion set theory is far more complex than the assumptions of the ellipsoidal collapse [254, 275, 378]. We are investigating this using N-body simulations.

The effect of peculiar velocities on measuring the Hubble-Lemaître constant H_0 in the local Universe has essential implications for understanding the Hubble tension [301, 305, 307]. Study in §4 supports the idea that local measurements of H_0 are systematically different from the global value due to gravitational clustering. It adds to the evidence that the Hubble tension may not only be due to observational systematics or unknown physics but could also be influenced by local arrangements of structures. Further, one needs to investigate how it impacts the calibrations involved in the observations. Observed large scale bulk flows around the Milky Way [314] suggest that our region might be an outlier in the expected distribution of galaxies and flows according to the concordance Λ CDM model. This implies that local measurements might be largely biased if we live in a region with atypical flows. However, more detailed studies are needed to evaluate this possibility.

N-body simulations are essential tools in cosmology and galaxy formation studies. However, understanding their limitations is necessary to avoid misinterpreting simulation data. The finite resolution of N-body simulations leads to missing power at small scales. In §5, we see that this results in errors in the root-mean-square fluctuations and the corresponding mass function. These errors are most significant at early times and on small scales. Analytical calculations suggest that the mode coupling between small and large scales is tied to the resolution of collapsed halos. Thus, accurate mode coupling estimates require sufficient halos in the simulation.

The thesis highlights the need to refine existing theoretical models, such as the halo mass function. Such refinement is essential for improving galaxy abundance and their distribution predictions, impacting the large scale surveys like Euclid [46, 73] and LSST [45, 72]. The recent discovery of a very large number of massive galaxies at high redshifts in JWST [379–381] data demands a closer look at galaxy abundances. Identifying sources of deviations in halo mass function can lead to more accurate theory-observation alignment. Moreover, the role of peculiar velocities in local Hubble-Lemaître constant (H_0) measurements needs to be reconsidered while analysing the Hubble tension. Linking local density to variations in H_0 can provide a statistical framework to contextualise discrepancies. Examining cosmological simulation limitations helps bridge the gaps in modelling the structure formation within theoretical frameworks. Resolving such errors is necessary to ensure simulations can reliably inform future observations.

A Consequence of expanding Universe on observables

The expansion of the Universe influences the measurement of physical quantities in various ways. Here, we discuss the Hubble scale, redshift and redshift-space distortions.

A.1 The Hubble scale/radius

The Hubble parameter sets a characteristic time scale for the expansion of the Universe, which in turn defines a corresponding length scale over which physical processes function coherently [4, 7]. Hubble scale R_H is defined as,

$$R_H = cH^{-1}, \quad (\text{A.1})$$

where, c denotes the speed of light, and H is the Hubble parameter which signifies the rate of cosmic expansion. As the Universe expands, the Hubble scale marks the threshold at which the curvature of spacetime, as governed by general relativity, begins to impact cosmic dynamics. On scales smaller than this, Newtonian gravity typically suffices.

A.2 Redshift

Redshift denotes the increase in the wavelength and a corresponding decrease in the frequency of light emitted from distant objects as a consequence of the expansion of the Universe. The redshift z can be measured as,

$$1 + z = \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}}, \quad (\text{A.2})$$

where, $\lambda_{\text{observed}}$ is the observed wavelength of light and λ_{emitted} is the wavelength of light emitted by the source. It can be shown that

$$1 + z(t) = \frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})}. \quad (\text{A.3})$$

The redshift function $z(t)$ and the scale factor $a(t)$ serve as interchangeable descriptors of cosmic expansion and can be used as time variables. At small redshifts ($z \ll 1$), the relationship between redshift and distance can be established using the Hubble's law ($v \approx cz = H_0d$).

A.3 Redshift-space distortions

Redshift-space distortions occur due to the peculiar velocities of galaxies within the large scale structure of the Universe. Velocities of galaxies along the line of sight result in distortions in their observed redshifts. This effect introduces anisotropy in the distribution of galaxies, particularly in galaxy surveys where redshift is used as a proxy for distance [382–385].

Two distinct effects contribute to redshift distortions. The first, known as the *Fingers of God*, results from the velocity dispersion of galaxies within rich clusters, elongating the cluster in redshift space. This effect influences only redshift, not sky position, so the elongation occurs solely along the radial direction. Hence, the term ‘fingers’ points to the observer. The other significant redshift distortion, the *Kaiser effect* [382], arises from galaxies bound to a central mass and undergoing infall. Unlike the Fingers of God, the peculiar velocities in the Kaiser effect are coherent, not random, and directed towards the central mass, although the effect is more nuanced.

In general, redshift has two components: the cosmological component (z_H) due to the expansion of the Universe, as discussed above, and the Doppler effect of peculiar velocities (z_{peculiar}). Therefore, if r and v_{peculiar} are the distance and the peculiar velocity of the source galaxy, the total observed redshift is

$$(1 + z_{\text{observed}}) = (1 + z_H)(1 + z_{\text{peculiar}}), \quad (\text{A.4})$$

where, $z_{\text{peculiar}} \approx v_{\text{peculiar}}/c$ for $v_{\text{peculiar}} \ll c$. Hence, for small distances ($cz_H \approx H_0 r$)

$$z_{\text{observed}} \approx \frac{1}{c} [rH_0 + v_{\text{peculiar}}]. \quad (\text{A.5})$$

We can define a redshift space $\mathbf{s} \equiv (s_1, s_2, s_3)$ which is a transform of the real space $\mathbf{r} \equiv (r_1, r_2, r_3)$, as

$$s_1 = r_1 = \frac{cz}{H_0}\theta_1, \quad s_2 = r_2 = \frac{cz}{H_0}\theta_2, \quad \text{and} \quad s_3 = r_3 + \frac{v_3}{H_0}, \quad (\text{A.6})$$

where, θ_1 and θ_2 are two angles on the sky and v_3 is peculiar velocity along the line of sight.

The radial axis is influenced by Doppler effects resulting from peculiar velocities. Further, these peculiar velocities arise from the clustering itself, introducing complexity. Consequently, the apparent clustering pattern observed in redshift space deviates systematically from that in real space. As a result, the spatial correlation function of galaxies, typically isotropic in real space, becomes anisotropic in redshift space.

B Fourier transform

The Fourier transform proves crucial in large-scale structure analyses, given that differentiation operations translate to simple multiplication in Fourier space. Here, we explore the Fourier transform of density contrast, the Vlasov equation, correlation functions, and various window functions. We will use the following Fourier convention:

$$F(\mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} F(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}. \quad (\text{B.1})$$

B.1 Density contrast

The Fourier component $\delta_{\mathbf{k}}$ of the density contrast $\delta_{\mathbf{x}} \equiv \delta(\mathbf{x})$ (eq. (1.6)) are related through a Fourier transform

$$\delta_{\mathbf{x}} = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}}, \quad \delta_{\mathbf{k}} = \int \frac{d^3\mathbf{k}}{V} \delta_{\mathbf{x}} e^{-i\mathbf{k}\cdot\mathbf{x}}. \quad (\text{B.2})$$

Since $\delta_{\mathbf{x}}$ is a real quantity $\delta_{\mathbf{k}}$ satisfies the reality condition, i.e. $\delta_{-\mathbf{k}} = \delta_{\mathbf{k}}^*$.

B.2 Vlasov equation

The Vlasov eq. (1.8) dictates the dynamics of density contrast eq. (1.6) in expanding Universe as discussed in §1.2.2. Assuming Universe as a periodic volume V , comprised of particles, with M representing the total mass within V , Fourier transform Vlasov equation is given as (refer §27 of [4])

$$\ddot{\delta}_{\mathbf{k}} + 2\frac{\dot{a}}{a}\dot{\delta}_{\mathbf{k}} = A_{\mathbf{k}} - B_{\mathbf{k}}, \quad (\text{B.3a})$$

$$A_{\mathbf{k}} = \frac{1}{M} \sum_j m_j \left[i\mathbf{k} \cdot \left(-\frac{\nabla \phi_j}{a^2} \right) e^{i\mathbf{k}\cdot\mathbf{x}_j} \right], \quad (\text{B.3b})$$

$$B_{\mathbf{k}} = \frac{1}{M} \sum_j m_j (\mathbf{k} \cdot \dot{\mathbf{x}}_j)^2 e^{i\mathbf{k}\cdot\mathbf{x}_j}. \quad (\text{B.3c})$$

The terms $A_{\mathbf{k}}$ and $B_{\mathbf{k}}$ denote nonlinear couplings between different modes. Specifically, $B_{\mathbf{k}}$ couples density contrasts indirectly via the peculiar velocities of particles.

B.3 n -point Fourier-space correlations

The correlation functions eq. (2.6) give information about the statistic of spatial arrangements of density contrast δ , as discussed in §2.2.1. Their Fourier space counterpart are called Fourier-space correlation functions are given as,

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2) \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) P(k_1), \quad (\text{B.4a})$$

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3) \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3), \quad (\text{B.4b})$$

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\delta(\mathbf{k}_4) \rangle_c = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4), \quad (\text{B.4c})$$

where, δ_D is the Dirac delta function, the quantities P , B and T are known as the power spectrum, bispectrum and trispectrum, respectively.

B.4 Filter/window functions

A common concept in manipulating cosmological density fields involves filtering, achieved through the convolution of the density field with a filter/window function: $\delta \rightarrow \delta * W$ or in Fourier space $\delta_k \rightarrow \delta_k * W_k$. For example, the filtered power spectrum is $P(k)|W_k|^2$. Below are three standard filter functions. Each filter allows the definition of a mass $M = f_W r^3 \rho_m$, where f_W is a constant dependent on the shape of the filter. Thus, a filter can be characterised by size R or mass M .

Gaussian filter:

The Gaussian window function $W(r, R)$ for characteristic scale R or M , its Fourier transform $\tilde{W}(k, R)$, and corresponding volume factor f_W is given as

$$W(r, R) = \frac{1}{(2\pi)^{3/2} R^3} e^{-\frac{r^2}{2R^2}}, \quad (\text{B.5a})$$

$$\tilde{W}(k, R) = e^{-\frac{k^2 R^2}{2}}, \quad (\text{B.5b})$$

$$f_W = (2\pi)^{3/2}. \quad (\text{B.5c})$$

Sharp k -space Filter:

The Sharp k -space window function $W(r, R)$ for characteristic scale R or M , its Fourier transform $\tilde{W}(k, R)$, and corresponding volume factor f_W is given as

$$W(r, R) = \frac{1}{2\pi^2 r^3} \left[\sin\left(\frac{r}{R}\right) - \left(\frac{r}{R}\right) \cos\left(\frac{r}{R}\right) \right], \quad (\text{B.6a})$$

$$\tilde{W}(k, R) = \begin{cases} 1 & k \leq \frac{1}{R} \\ 0 & k > \frac{1}{R} \end{cases}, \quad (\text{B.6b})$$

$$f_W = 6\pi^2. \quad (\text{B.6c})$$

Top-hat Filter:

The top-hat window function $W(r, R)$ for characteristic scale R or M , its Fourier transform $\tilde{W}(k, R)$, and corresponding volume factor f_W is given as

$$W(r, R) = \begin{cases} \frac{3}{4\pi R^3} & r \leq R \\ 0 & r > R \end{cases}, \quad (\text{B.7a})$$

$$\tilde{W}(k, R) = \frac{3}{k^3 R^3} [\sin(kR) - (kR) \cos(kR)], \quad (\text{B.7b})$$

$$f_W = \frac{4\pi}{3}. \quad (\text{B.7c})$$

C Mass function: supplementary analysis

We present an additional examination of the halo mass function, testing the robustness of our computed mass function for scale-free models discussed in §3.2.1. We explore its sensitivity to the choice of initial redshift (appendix C.1), box size (appendix C.2), and error types (appendix C.3). In appendix C.4, we provide box-wise fits for the mass function in the Λ CDM model, elaborated in §3.3. The color formatting for figures in this chapter follows the conventions specified in §3, specifically referencing fig. 3.1, 3.3, and 3.6. For all the mass functions variations presented in this chapter, we use the m_{200c} halo mass definition, with 300 halos per bin (a bin size of > 500 halos is found to produce systematic effects), $N_{\text{cut}} = 100$, and excluded the top 0.1% of massive halos.

C.1 Effect of initial redshift

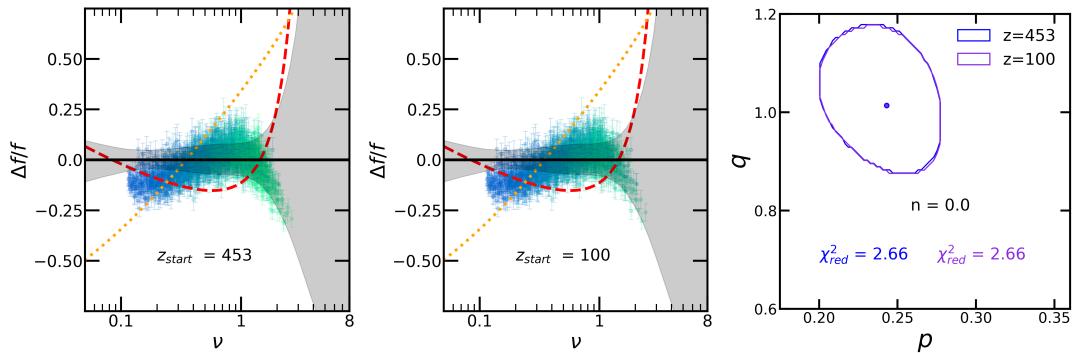


Figure C.1: Effect of initial redshift on mass function in scale-free model $n = 0.0$

Generally, varying the initial redshifts can introduce significant systematic errors in the mass function [171]. To address this, we used simulations with $n = 0$, starting at $z_{\text{start}} = 100$, and compared the resulting mass function with that from a simulation starting at $z_{\text{start}} = 453$, keeping all other parameters identical. The results are presented in fig. C.1. The left and middle panels show the residuals of the mass function data relative to the best-fit Sheth-Tormen model, while the right panel shows a standard deviation contours of the mass function for the two cases. The solid dots within the contours represent the best-fit values for the two simulations. We observe no change in the best-fit Sheth-Tormen parameters or a standard

deviation confidence levels. Additionally, both fits show no deviation from the best fit in the mass function computed using ten realisations.

C.2 Effect of Box size

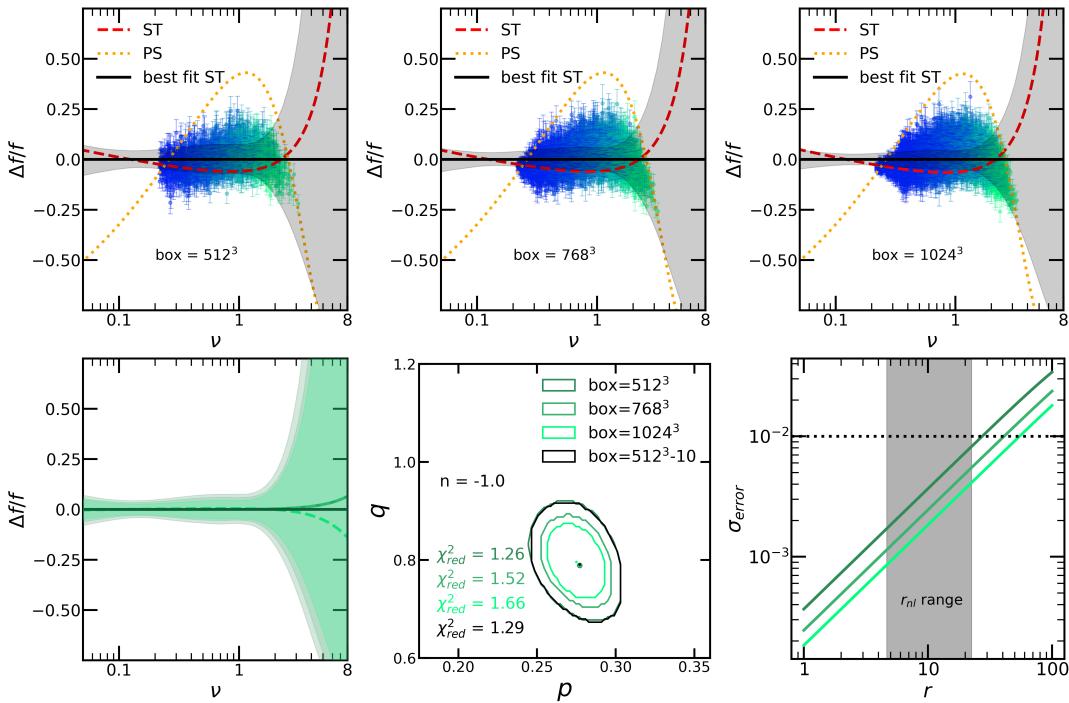


Figure C.2: Effect of Box size on mass function in scale-free model $n = -1.0$

The small box sizes used for some power-law models might introduce systematic errors. However, because the power-law slopes used in these smaller boxes are relatively smaller, the impact of finite box effects is expected to be minimal [134, 187]. To ensure these effects are negligible, we compared three simulations with $n = -1$, using box sizes of 512^3 , 768^3 , and 1024^3 . All three simulations share the same parameters, differing only in box size and the number of particles ($N_{\text{part}} = \text{box size}$). The results are shown in fig. C.2. The top row of panels displays the residuals of the mass function data relative to the best-fit Sheth-Tormen model for box sizes 512^3 (left), 768^3 (middle), and 1024^3 (right). The bottom left panel illustrates the deviations of the three fits relative to the best-fit (solid black) presented in §3.2.1 which is based on ten realisations. The bottom middle panel shows a standard deviation contours for the fits, with the solid dots representing the best-fit values for the corresponding fits. We observe no significant change in the best-fit values, though a standard deviation interval narrows as the box size increases.

The decrease in a standard deviation interval is due to including larger scales in simulations with larger box sizes. The theoretical errors in the mass variance for the three simu-

lations are shown in the bottom right panel of fig. C.2. This plot determines the upper limit of r_{nl} for calculating the mass function. The gray area in the plot represents the r_{nl} range based on a 512^3 box. The same r_{nl} range is applied across the three simulations. The upper limit of r_{nl} is set to ensure that σ_{error} remains below 1%, while the lower limit is determined by technical considerations, ensuring that each r_{nl} value includes at least 10 data points. The plot confirms that the chosen box sizes are sufficient for obtaining reliable results within the specified r_{nl} range outlined in table 3.1.

C.3 Effect of error-types

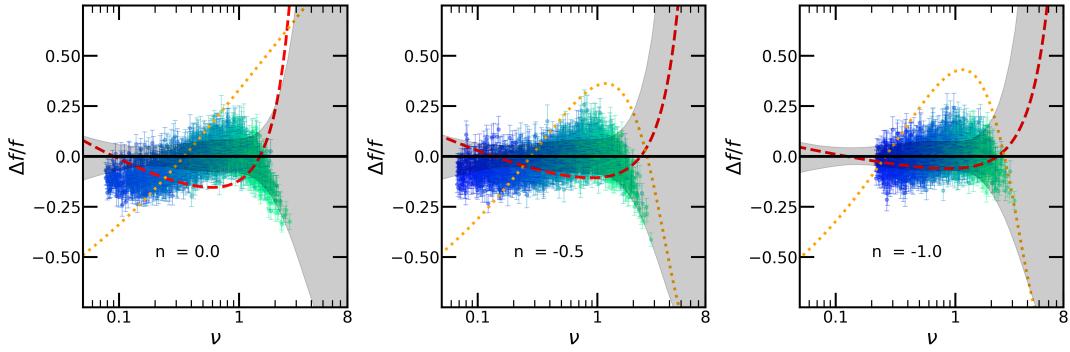


Figure C.3: Mass function: Poisson error in scale-free models $n = 0.0, -0.5, -1.0$

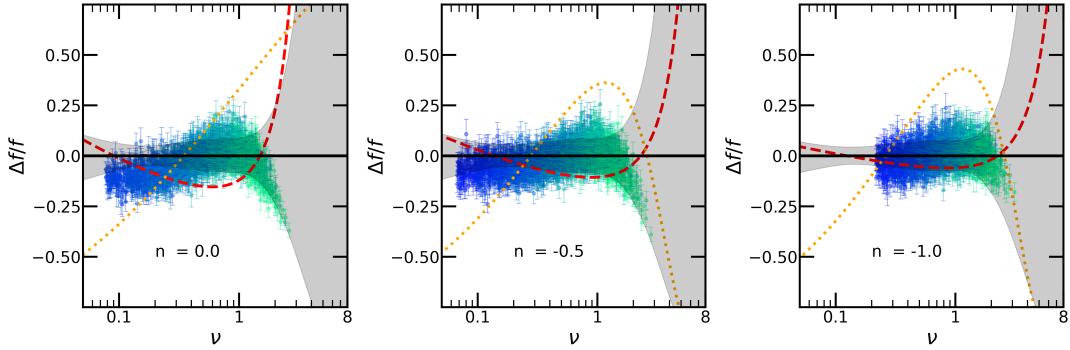


Figure C.4: Mass function: Jackknife error in scale-free models $n = 0.0, -0.5, -1.0$

Poisson errors for halo counts in bins may under-estimate the errors at the low mass end of the halo mass function. An alternate estimate can be obtained using jackknife errors. We have conducted Jackknife error estimates for scale-free models. We divide a simulation volume into eight equal subvolumes to generate jackknife samples. Furthermore, as we have multiple realisations for three models with power law index $n = 0, -0.5, -1.0$, we also estimated sample variance error and included them here. Figure C.4, C.5, and C.3 depict the residue of the mass function data with the best-fit Sheth-Tormen for Poisson errors, Jackknife, and sample variance, respectively.

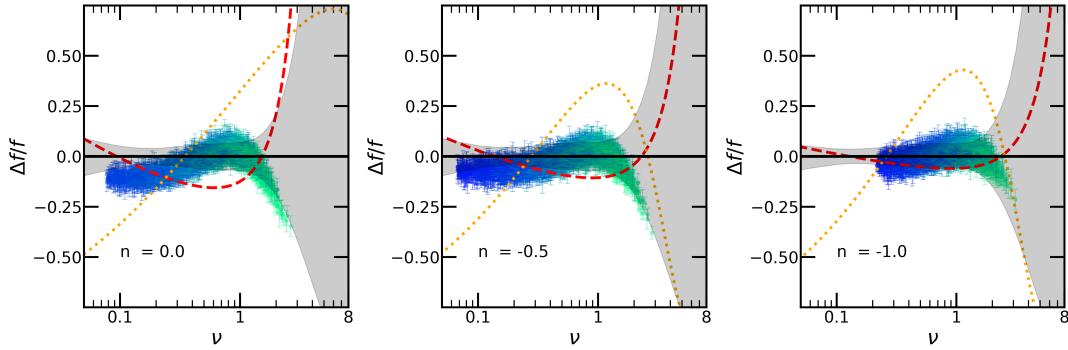


Figure C.5: *Mass function: Sample variance error in scale-free models $n = 0.0, -0.5, -1.0$*

We have found that Poisson errors and Jackknife errors are comparable and do not offer any benefit over one another. Sample variance is more effective as it uses information from multiple realisations. However, we do not implement it because the additional computational time outweighs the fit improvement.

The fig. C.6 shows the best-fit parameters of p (left panel), q (middle panel), and power law index n in three cases, with the right panel displaying a standard deviation contours. Solid, dotted, and dashed lines represent Poisson, Jackknife, and Sample Variance errors. The analysis reveals no substantial change in the best-fit values of p and q . A standard deviation contours for Poisson and Jackknife errors are relatively unchanged, but the Sample Variance contours are comparatively tighter. As a result, using Poisson errors for our purposes is sufficient.

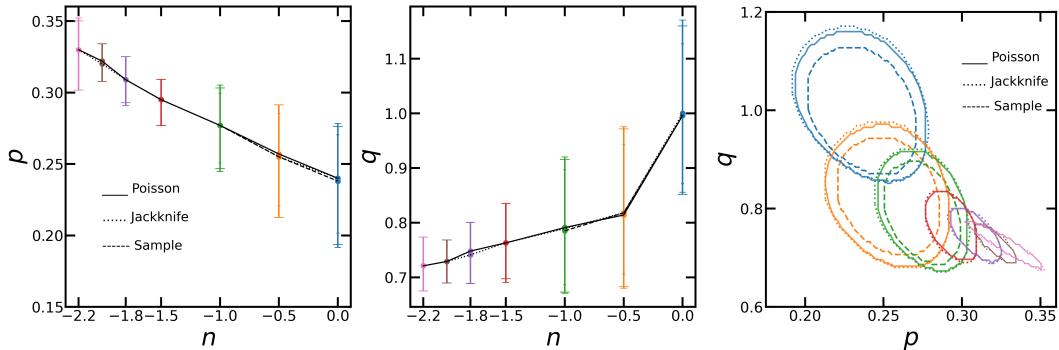


Figure C.6: *Mass function: Effect of error-types for all scale-free models*

C.4 Box-wise Λ CDM mass function fits

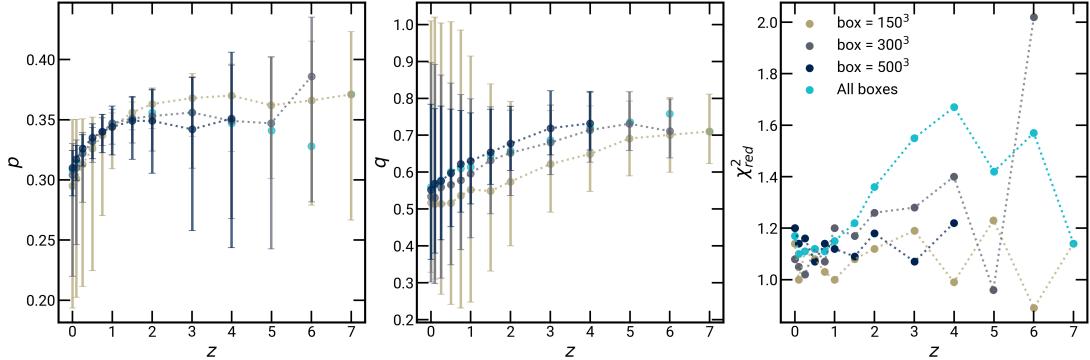


Figure C.7: *Box-wise Λ CDM mass function*

We present the box-wise mass function fit for Λ CDM models in fig. C.7. We confirm that the mass function remains consistent within the error bars across different box sizes by fitting each box individually and comparing the results with a combined fit.

D H_0 measurements: supplementary analysis

In this chapter, we further analyse H_0 measurements using simulations referencing the effects of substructure and halo mass cuts on the mean and standard deviation of the distributions.

D.1 Effect of substructure

We analyse the distributions by including substructures, using the $L_{box}(\text{Mpc}/\text{h}) / N_{part}$: $150/1024^3$ simulation. The first panel of fig. D.1 shows the halo mass function for central and satellite halos in our halo catalogue. The substructure fraction generally increases with host halo mass but remains a minority ($\sim 10\text{--}30\%$) of the total halo population in simulations [386–388]. The second and third panels show mean and standard deviations for five different shells with and without substructure. Including substructures increases the number of observers by 30% compared to when substructures are not considered. This leads to a slight increase in the dispersion by a few percent.

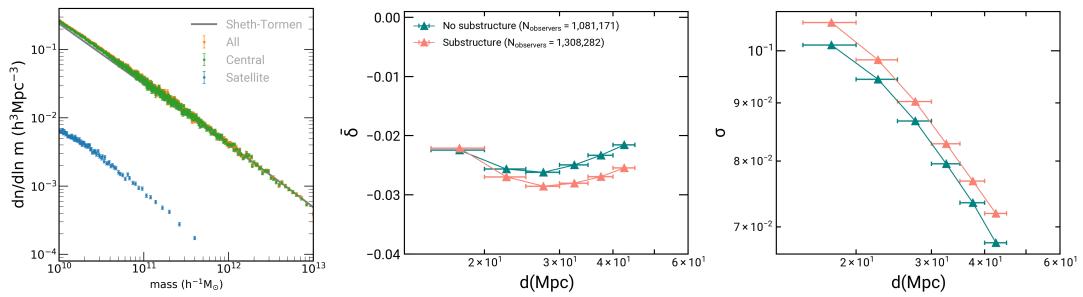


Figure D.1: H_0 PDF: Effect of substructure

D.2 Convergence in the simulations

We analyse the distributions by applying different mass cuts as a proxy for convergence study. We use four simulations with $L_{box}(\text{Mpc}/\text{h})/N_{part}$: $150/1024^3$, $300/1024^3$, $500/1024^3$ and $1000/1024^3$, with corresponding lowest halo masses (M_\odot/h) of 8.68×10^9 , 6.94×10^{10} , 3.22×10^{11} and 2.57×10^{12} . These lowest halo masses are used as mass cuts to compute the distributions, which are shown in the fig. D.2.

In the first row, we use a mass cut of 300 Mpc/h box simulation to the 300 and 150 Mpc/h box simulations, considering distance bins up to one-third of the box size. Similarly, the second row compares a mass cut from the 500 Mpc/h box simulation to the 500, 300, and 150 Mpc/h box simulations. The last row shows the mass cut from the 1000 Mpc/h box simulation, compared to the 1000, 500, 300, and 150 Mpc/h box simulations. The figure demonstrates that a larger box size reduces the mean and standard deviations offset significantly for the same mass cut. Additionally, the offset in both quantities increases with distance. We verify that the sample size of a number of observers ($N_{\text{observers}}$) does not alter the results.

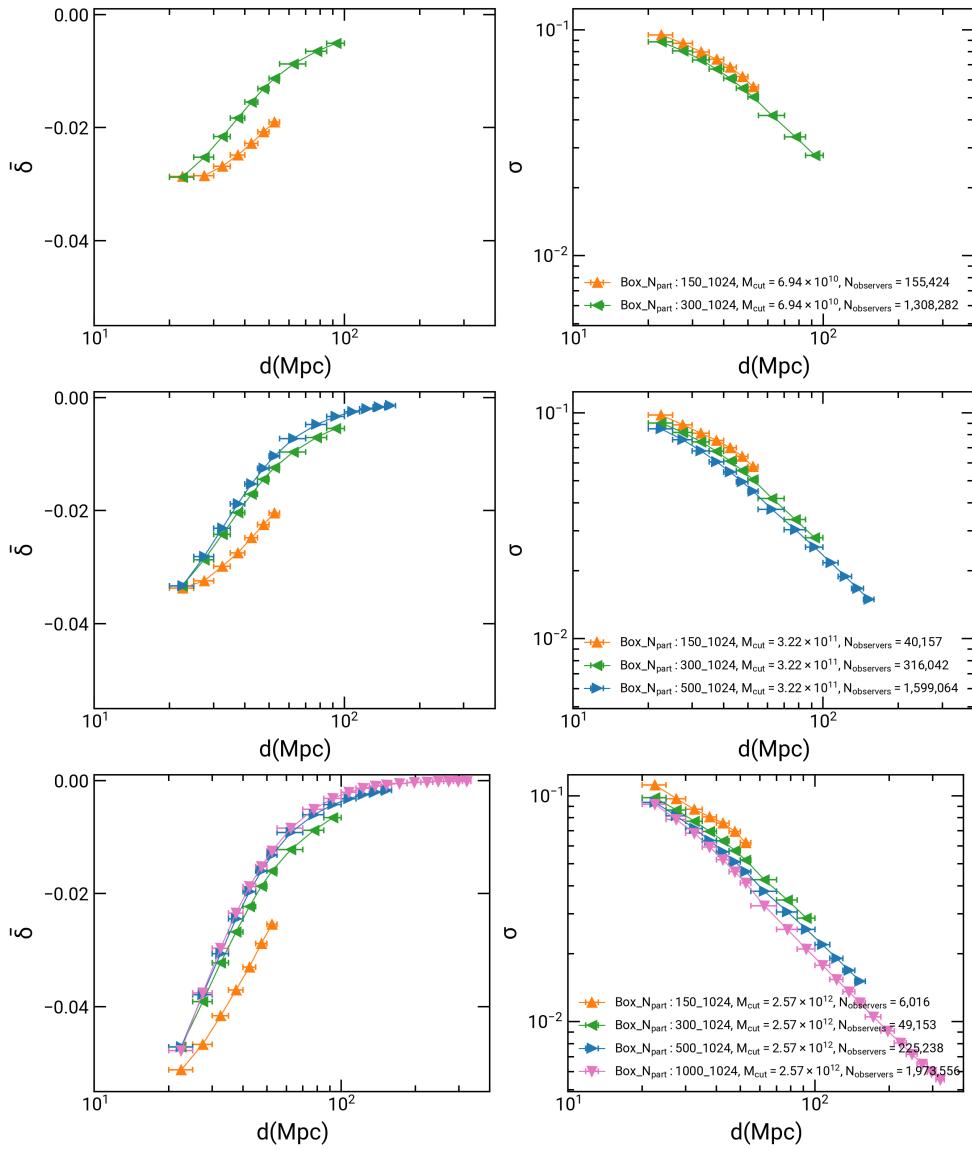


Figure D.2: H_0 PDF: Convergence in the simulations

E Simulations in §5

In this chapter, we describe the simulations used in §5 and fig. 5.1. We also demonstrate the convergence of the mass function computed in these simulations.

| n | z_{start} | N_{box} | $\sqrt[3]{N_{\text{part}}}$ | r_{nl}^{sim} | r_{nl}^{mf} | r_{nl}^{max} |
|------|--------------------|------------------|-----------------------------|-----------------------|----------------------|-----------------------|
| -1.0 | 160 | 512 | 512 | 1.25-49.4 | 1.265, 1.631, 2.121 | 27.4 |
| -1.0 | 160 | 768 | 768 | 1.25-49.4 | 1.265, 1.631, 2.121 | 27.4 |
| -1.0 | 160 | 1024 | 768 | 1.25-49.4 | 1.631, 2.121, 2.757 | 27.4 |
| -1.0 | 160 | 1024 | 1024 | 1.25-49.4 | 1.265, 1.631, 2.121 | 27.4 |
| -1.8 | 70 | 1024 | 1024 | 1.25-49.4 | 1.265, 1.631, 2.121 | 14.5 |
| -2.2 | 46 | 1536 | 1536 | 0.2-10.24 | 0.965, 1.265, 1.631 | 4.4 |

Table E.1: *Simulation Setup*: Column 1: Power law power spectrum index for the model, Column 2: Initial redshift used to start the simulation, Column 3: Side length of the cubical simulation box, Column 4: Cube root of the total number of particles put in the simulation, Column 5: Range of the scale of non-linearity(r_{nl}) covered in the simulation, Column 6: r_{nl} s used to compute the mass function, Column 8: Maximum limit on r_{nl} considering the finite box size effect.

We use six dark-matter-only simulations initialised with three power-law power spectrum conditions. Additional details about the simulations are provided in table E.1 and §2.4.2.

Figure E.1 shows a conversion of the mass function computed in the simulation with respect to finite mass resolution. The lower-resolution scatters show higher fractional errors in mass function at a given r_{nl} . However, these errors settle similarly with the r_{nl} . Likewise, fig. E.2 displays the impact of finite box size on the mass function. The box size affects the range of masses that can be probed, with larger box sizes allowing for better coverage of the higher-mass end.

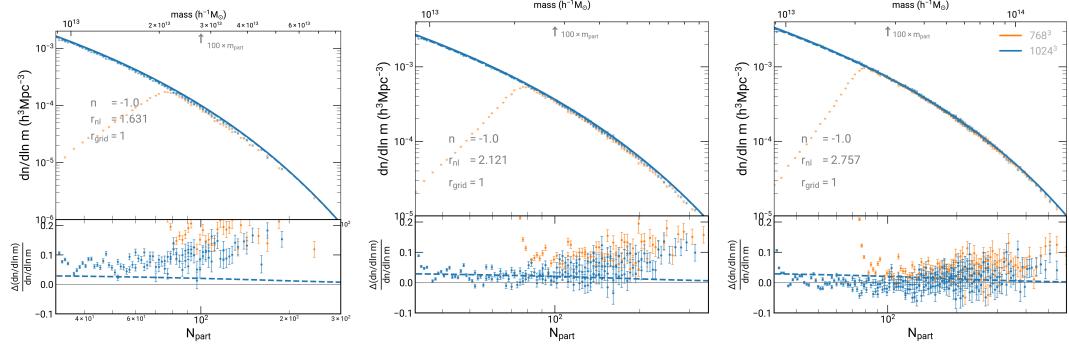


Figure E.1: *Mass function: Effect of finite mass resolution:* Each panel shows the halo mass function and fractional errors in mass function in the simulation for the $n = -1.0$ model with two values of $N_{\text{part}} = 768^3$ (orange), and 1024^3 (blue) of box size 1024^3 shown in Table E.1. Data points represent simulation, whereas solid lines show theoretical mass function. Three panels represent different scales on non-linearity (r_{nl}).

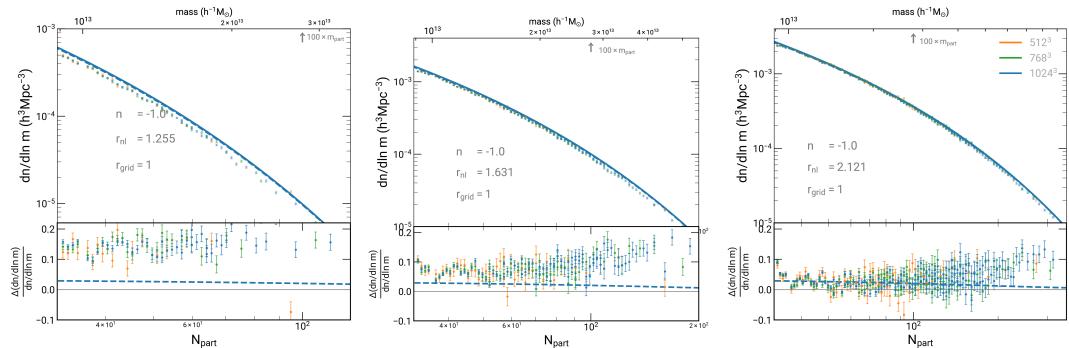


Figure E.2: *Mass function: Effect of finite box size:* Each panel shows the halo mass function and fractional errors in mass function in the simulation for three variations of $N_{\text{part}} = N_{\text{box}}^3 = 512^3$ (orange), 768^3 (green), 1024^3 (blue) for the $n = -1.0$ model shown in Table E.1. Data points represent simulation, whereas solid lines show theoretical mass function. Three panels represent different scales on non-linearity (r_{nl}).

Bibliography

- [1] Swati Gavas et al. “Halo mass function in scale invariant models”. In: *Monthly Notices of the Royal Astronomical Society* 521.4 (Mar. 2023), pp. 5960–5971. ISSN: 1365-2966.
- [2] Swati Gavas, J. S. Bagla, and Nishikanta Khandai. “Dispersion in the Hubble-Lemaître constant measurements from gravitational clustering”. In: (July 2024). arXiv: 2407 . 10139 [astro-ph.CO].
- [3] J. S. Bagla and Swati Gavas. *On the origin of transient features in cosmological N-Body Simulations*. 2024. arXiv: 2408 . 05118 [astro-ph.CO].
- [4] P. J. E. Peebles. *The Large-Scale Structure of the Universe*. Princeton University Press, Dec. 1981. ISBN: 9780691206714.
- [5] S. F. Shandarin and Ya. B. Zeldovich. “The large-scale structure of the universe: Turbulence, intermittency, structures in a self-gravitating medium”. In: *Reviews of Modern Physics* 61.2 (Apr. 1989), pp. 185–220. ISSN: 0034-6861.
- [6] J. A. Peacock. *Cosmological Physics*. Cambridge University Press, Dec. 1998. ISBN: 9780511804533.
- [7] T. Padmanabhan. *Structure Formation in the Universe*. Cambridge University Press, 1993. ISBN: 9780521424868.
- [8] T. Padmanabhan. *Theoretical Astrophysics*. Cambridge University Press, Oct. 2002. ISBN: 9780511840166.
- [9] F. Bernardeau et al. “Large-scale structure of the Universe and cosmological perturbation theory”. In: *Physics Reports* 367.1-3 (Sept. 2002), pp. 1–248. ISSN: 0370-1573.
- [10] Alison L. Coil. “The Large-Scale Structure of the Universe”. In: *Planets, Stars and Stellar Systems*. Springer Netherlands, 2013, pp. 387–421.
- [11] G. Kauffmann, S. D. M. White, and B. Guiderdoni. “The formation and evolution of galaxies within merging dark matter haloes*”. In: *Monthly Notices of the Royal Astronomical Society* 264.1 (Sept. 1993), pp. 201–218. ISSN: 1365-2966.
- [12] C. Lacey and S. Cole. “Merger rates in hierarchical models of galaxy formation”. In: *Monthly Notices of the Royal Astronomical Society* 262.3 (June 1993), pp. 627–649. ISSN: 1365-2966.

- [13] Kamal Bora, R. F. L. Holanda, and Shantanu Desai. “Probing the dark matter density evolution law with large scale structures”. In: *The European Physical Journal C* 81.7 (July 2021). ISSN: 1434-6052.
- [14] Ivan Debono. “Probing Inflation with Large-Scale Structure Data: The Contribution of Information at Small Scales”. In: *The 1st Electronic Conference on Universe*. ECU 2021. MDPI, Mar. 2021.
- [15] Ruth Durrer. “Testing general relativity with cosmological large scale structure”. In: *General Relativity and Gravitation* 54.8 (Aug. 2022). ISSN: 1572-9532.
- [16] Bartolomeo Fiorini, Kazuya Koyama, and Albert Izard. “Studying large-scale structure probes of modified gravity with COLA”. In: *Journal of Cosmology and Astroparticle Physics* 2022.12 (Dec. 2022), p. 028. ISSN: 1475-7516.
- [17] A. J. Benson et al. “The nature of galaxy bias and clustering”. In: *Monthly Notices of the Royal Astronomical Society* 311.4 (Feb. 2000), pp. 793–808. ISSN: 1365-2966.
- [18] Andrew J. Benson. “Galaxy formation theory”. In: *Physics Reports* 495.2-3 (Oct. 2010), pp. 33–86. ISSN: 0370-1573.
- [19] Rachel S. Somerville and Romeel Dave. “Physical Models of Galaxy Formation in a Cosmological Framework”. In: *Annual Review of Astronomy and Astrophysics* 53.1 (Aug. 2015), pp. 51–113. ISSN: 1545-4282.
- [20] Risa H. Wechsler and Jeremy L. Tinker. “The Connection Between Galaxies and Their Dark Matter Halos”. In: *Annual Review of Astronomy and Astrophysics* 56.1 (Sept. 2018), pp. 435–487. ISSN: 1545-4282.
- [21] Jess Domnguez-Gmez et al. “Stellar mass-metallicity relation throughout the large-scale structure of the Universe: CAVITY mother sample”. In: *Astronomy and Astrophysics* 680 (Dec. 2023), A111. ISSN: 1432-0746.
- [22] E. R. HARRISON. “Observational tests in cosmology”. In: *Nature* 260.5552 (Apr. 1976), pp. 591–592. ISSN: 1476-4687.
- [23] Matt Visser. “Cosmography: Cosmology without the Einstein equations”. In: *General Relativity and Gravitation* 37.9 (Sept. 2005), pp. 1541–1548. ISSN: 1572-9532.
- [24] Neal Dalal, Joseph F. Hennawi, and Paul Bode. “Noise in Strong Lensing Cosmography”. In: *The Astrophysical Journal* 622.1 (Mar. 2005), pp. 99–105. ISSN: 1538-4357.
- [25] Jun Qing Xia et al. “Cosmography beyond standard candles and rulers”. In: *Physical Review D* 85.4 (Feb. 2012). ISSN: 1550-2368.
- [26] Hlne M. Courtois et al. “COSMOGRAPHY OF THE LOCAL UNIVERSE”. In: *The Astronomical Journal* 146.3 (Aug. 2013), p. 69. ISSN: 1538-3881.

- [27] Salvatore Capozziello, Mariafelicia De Laurentis, and Orlando Luongo. “Curvature dark energy reconstruction through different cosmographic distance definitions”. In: *Annalen der Physik* 526.7-8 (July 2014), pp. 309–317. ISSN: 1521-3889.
- [28] Peter K. S. Dunsby and Orlando Luongo. “On the theory and applications of modern cosmography”. In: *International Journal of Geometric Methods in Modern Physics* 13.03 (Mar. 2016), p. 1630002. ISSN: 1793-6977.
- [29] Ya. B. Zel'dovich. “Gravitational instability: An approximate theory for large density perturbations.” In: *Astronomy and Astrophysics* 5 (Mar. 1970), pp. 84–89.
- [30] James E. Gunn and III Gott J. Richard. “On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution”. In: *The Astrophysical Journal* 176 (Aug. 1972), p. 1. ISSN: 1538-4357.
- [31] J. A. Fillmore and P. Goldreich. “Self-similar gravitational collapse in an expanding universe”. In: *The Astrophysical Journal* 281 (June 1984), p. 1. ISSN: 1538-4357.
- [32] S. N. Gurbatov, A. I. Saichev, and S. F. Shandarin. “The large-scale structure of the Universe in the frame of the model equation of non-linear diffusion”. In: *Monthly Notices of the Royal Astronomical Society* 236.2 (Jan. 1989), pp. 385–402. ISSN: 1365-2966.
- [33] S. Matarrese et al. “A frozen-flow approximation to the evolution of large-scale structures in the Universe”. In: *Monthly Notices of the Royal Astronomical Society* 259.3 (Dec. 1992), pp. 437–452. ISSN: 1365-2966.
- [34] Tereasa G. Brainerd, Robert J. Scherrer, and Jens V. Villumsen. “Linear Evolution of the Gravitational Potential: A New Approximation for the Nonlinear Evolution of Large-Scale Structure”. In: *The Astrophysical Journal* 418 (Dec. 1993), p. 570. ISSN: 1538-4357.
- [35] J. S. Bagla and T. Padmanabhan. “Non-linear evolution of density perturbations using the approximate constancy of the gravitational potential”. In: *Monthly Notices of the Royal Astronomical Society* 266.1 (Jan. 1994), pp. 227–237. ISSN: 1365-2966.
- [36] V Sahni. “Approximation methods for non-linear gravitational clustering”. In: *Physics Reports* 262.1-2 (Nov. 1995), pp. 1–135. ISSN: 0370-1573.
- [37] Lam Hui and Edmund Bertschinger. “Local Approximations to the Gravitational Collapse of Cold Matter”. In: *The Astrophysical Journal* 471.1 (Nov. 1996), pp. 1–12. ISSN: 1538-4357.
- [38] C.D. Shane and C.A. Wirtanen. “Publications of the Lick Observatory”. In: *Publications of the Lick Observatory* Vol. 22 Part 1 (1967).
- [39] Thomas Jarrett. “Large Scale Structure in the Local Universe - The 2MASS Galaxy Catalog”. In: *Publications of the Astronomical Society of Australia* 21.4 (2004), pp. 396–403. ISSN: 1448-6083.

- [40] R. P. Kirshner, Jr. Oemler A., and P. L. Schechter. “A study of field galaxies. I - Redshifts and photometry of a complete sample of galaxies”. In: *The Astronomical Journal* 83 (Dec. 1978), p. 1549. ISSN: 0004-6256.
- [41] M. Davis et al. “A survey of galaxy redshifts. II - The large scale space distribution”. In: *The Astrophysical Journal* 253 (Feb. 1982), p. 423. ISSN: 1538-4357.
- [42] Matthew Colless et al. “The 2dF Galaxy Redshift Survey: spectra and redshifts”. In: *Monthly Notices of the Royal Astronomical Society* 328.4 (Dec. 2001), pp. 1039–1063. ISSN: 1365-2966.
- [43] Donald G. York et al. “The Sloan Digital Sky Survey: Technical Summary”. In: *The Astronomical Journal* 120.3 (Sept. 2000), pp. 1579–1587. arXiv: astro-ph/0006396 [astro-ph].
- [44] DESI Collaboration et al. “The Early Data Release of the Dark Energy Spectroscopic Instrument”. In: (2023).
- [45] LSST Science Collaboration et al. *LSST Science Book, Version 2.0*. 2009. arXiv: 0912 . 0201 [astro-ph.IM].
- [46] Euclid Collaboration. *Euclid Definition Study Report*. 2011. arXiv: 1110 . 3193 [astro-ph.CO].
- [47] Edmund Bertschinger. “SIMULATIONS OF STRUCTURE FORMATION IN THE UNIVERSE”. In: *Annual Review of Astronomy and Astrophysics* 36.1 (Sept. 1998), pp. 599–654. ISSN: 1545-4282.
- [48] J. S. Bagla. “TreePM A code for cosmological N-body simulations”. In: *Journal of Astrophysics and Astronomy* 23.3-4 (Dec. 2002), pp. 185–196. ISSN: 0973-7758.
- [49] Volker Springel et al. “Simulations of the formation, evolution and clustering of galaxies and quasars”. In: *Nature* 435.7042 (June 2005), pp. 629–636. ISSN: 1476-4687.
- [50] Volker Springel et al. “First results from the IllustrisTNG simulations: matter and galaxy clustering”. In: *Monthly Notices of the Royal Astronomical Society* 475.1 (Dec. 2017), pp. 676–698. ISSN: 1365-2966.
- [51] Raul E. Angulo and Oliver Hahn. “Large-scale dark matter simulations”. In: *Living Reviews in Computational Astrophysics* 8.1 (Feb. 2022). ISSN: 2365-0524.
- [52] E. P. Hubble. “Extragalactic nebulae.” In: *The Astrophysical Journal* 64 (Dec. 1926), p. 321. ISSN: 1538-4357.
- [53] Harlow Shapley and Adelaide Ames. “A Survey of the External Galaxies Brighter than the Thirteenth Magnitude”. In: *Annals of Harvard College Observatory* 88 (Jan. 1932), pp. 41–76.
- [54] Edwin Hubble. “The Distribution of Extra-Galactic Nebulae”. In: *The Astrophysical Journal* 79 (Jan. 1934), p. 8. ISSN: 1538-4357.

- [55] F. Zwicky, E. Herzog, and P. Wild. *Catalogue of Galaxies and of Clusters of Galaxies*. Caltech, 1961.
- [56] Fritz Zwicky. *Morphological astronomy*. 1957.
- [57] C. D. Shane. “The distribution of extragalactic nebulae. II”. In: *The Astronomical Journal* 61 (Aug. 1956), p. 292. ISSN: 0004-6256.
- [58] M. J. Geller, M. J. Kurtz, and V. de Lapparent. “The Shane-Wirtanen counts”. In: *The Astrophysical Journal* 287 (Dec. 1984), p. L55. ISSN: 1538-4357.
- [59] M. Seldner et al. “New reduction of the Lick catalog of galaxies”. In: *The Astronomical Journal* 82 (Apr. 1977), p. 249. ISSN: 0004-6256.
- [60] G. de Vaucouleurs. “Sur la Contribution des Nébuleuses Extragalactiques à la Lumière du Ciel Nocturne”. In: *Annales d'Astrophysique* vol. 12, pg. 162 (1949).
- [61] P. J. E. Peebles and E. J. Groth. “Statistical analysis of catalogs of extragalactic objects. V - Three-point correlation function for the galaxy distribution in the Zwicky catalog”. In: *The Astrophysical Journal* 196 (Feb. 1975), p. 1. ISSN: 1538-4357.
- [62] S. J. Maddox et al. “The APM galaxy surveyI. APM measurements and star-galaxy separation.” In: *MNRAS* 243 (Apr. 1990), pp. 692–712.
- [63] S. A. Gregory and L. A. Thompson. “The Coma/A1367 supercluster and its environs”. In: *The Astrophysical Journal* 222 (June 1978), p. 784. ISSN: 1538-4357.
- [64] M. Joeveer, J. Einasto, and E. Tago. “Spatial distribution of galaxies and of clusters of galaxies in the southern galactic hemisphere”. In: *Monthly Notices of the Royal Astronomical Society* 185.2 (Nov. 1978), pp. 357–370. ISSN: 1365-2966.
- [65] A. Sandage and G. A. Tammann. “Steps toward the Hubble constant. V - The Hubble constant from nearby galaxies and the regularity of the local velocity field”. In: *The Astrophysical Journal* 196 (Mar. 1975), p. 313. ISSN: 1538-4357.
- [66] Gerard Henri de Vaucouleurs, Antoinette de Vaucouleurs, and Harlow Shapley. *Reference catalogue of bright galaxies*. 1964.
- [67] G. de Vaucouleurs, A. de Vaucouleurs, and J. R. Corwin. “Second reference catalogue of bright galaxies”. In: *Second reference catalogue of bright galaxies* 1976 (Jan. 1976), p. 0.
- [68] Margaret J. Geller and John P. Huchra. “Mapping the Universe”. In: *Science* 246.4932 (Nov. 1989), pp. 897–903. ISSN: 1095-9203.
- [69] Sergei F. Shandarin and Capp Yess. “Detection of Network Structure in the Las Campanas Redshift Survey”. In: *The Astrophysical Journal* 505.1 (Sept. 1998), pp. 12–17. ISSN: 1538-4357.

- [70] S. J. Lilly et al. "The Canada-France Redshift Survey. I. Introduction to the Survey, Photometric Catalogs, and Surface Brightness Selection Effects". In: *Astrophysical Journal* 455 (Dec. 1995), p. 50. arXiv: astro-ph/9507010 [astro-ph].
- [71] Kyle S. Dawson et al. "THE BARYON OSCILLATION SPECTROSCOPIC SURVEY OF SDSS-III". In: *The Astronomical Journal* 145.1 (Dec. 2012), p. 10. ISSN: 1538-3881.
- [72] Zeljko Ivezić et al. "LSST: From Science Drivers to Reference Design and Anticipated Data Products". In: *The Astrophysical Journal* 873.2 (Mar. 2019), p. 111. ISSN: 1538-4357.
- [73] Euclid Collaboration. "Euclid preparation: XXIV. Calibration of the halo mass function in L(v)CDM cosmologies". In: *Astronomy and Astrophysics* 671 (Mar. 2023), A100. ISSN: 1432-0746.
- [74] Nick Kaiser. "Weak gravitational lensing of distant galaxies". In: *The Astrophysical Journal* 388 (Apr. 1992), p. 272. ISSN: 1538-4357.
- [75] Yannick Mellier. "Probing the Universe with Weak Lensing". In: *Annual Review of Astronomy and Astrophysics* 37. Volume 37, 1999 (1999), pp. 127–189. ISSN: 1545-4282.
- [76] Matthias Bartelmann and Peter Schneider. "Weak gravitational lensing". In: *Physics Reports* 340.4-5 (Jan. 2001), pp. 291–472. ISSN: 0370-1573.
- [77] M Birkinshaw. "The Sunyaev-Zeldovich effect". In: *Physics Reports* 310.2-3 (Mar. 1999), pp. 97–195. ISSN: 0370-1573.
- [78] R. A. Sunyaev and Ya. B. Zeldovich. "Microwave Background Radiation as a Probe of the Contemporary Structure and History of the Universe". In: *Annual Review of Astronomy and Astrophysics* 18.1 (Sept. 1980), pp. 537–560. ISSN: 1545-4282.
- [79] A. G. Kim et al. *Testing Gravity Using Type Ia Supernovae Discovered by Next-Generation Wide-Field Imaging Surveys*. 2019. arXiv: 1903.07652 [astro-ph.CO].
- [80] David H. Weinberg. "The Lyman-alpha Forest as a Cosmological Tool". In: *AIP Conference Proceedings*. AIP, 2003.
- [81] S. Weinberg. *GRAVITATION AND COSMOLOGY: PRINCIPLES AND APPLICATIONS OF THE GENERAL THEORY OF RELATIVITY*. Wiley India Pvt. Limited, 2008. ISBN: 9788126517558.
- [82] Houjun Mo, Frank C. van den Bosch, and Simon White. *Galaxy Formation and Evolution*. 2010.
- [83] S. Dodelson and F. Schmidt. *Modern Cosmology*. Elsevier Science, 2020. ISBN: 9780128159484.
- [84] D. Baumann. *Cosmology*. Cambridge University Press, 2022. ISBN: 9781108838078.
- [85] P. Meszaros. "The behaviour of point masses in an expanding cosmological substratum." In: *Astronomy and Astrophysics* 37.2 (Dec. 1974), pp. 225–228.

- [86] P. Meszaros. “Primeval black holes and galaxy formation.” In: *Astronomy and Astrophysics* 38.1 (Jan. 1975), pp. 5–13.
- [87] Wayne Hu and Naoshi Sugiyama. “Small-Scale Cosmological Perturbations: An Analytic Approach”. In: *The Astrophysical Journal* 471.2 (Nov. 1996), pp. 542–570. ISSN: 1538-4357.
- [88] Joseph Silk. “Cosmic Black-Body Radiation and Galaxy Formation”. In: *Astrophysical Journal* 151 (Feb. 1968), p. 459.
- [89] R. K. Sachs and A. M. Wolfe. “Perturbations of a Cosmological Model and Angular Variations of the Microwave Background”. In: *The Astrophysical Journal* 147 (Jan. 1967), p. 73. ISSN: 1538-4357.
- [90] Uros Seljak and Matias Zaldarriaga. “A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies”. In: *The Astrophysical Journal* 469 (Oct. 1996), p. 437. ISSN: 1538-4357.
- [91] Antony Lewis, Anthony Challinor, and Anthony Lasenby. “Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models”. In: *The Astrophysical Journal* 538.2 (Aug. 2000), pp. 473–476. ISSN: 1538-4357.
- [92] Julien Lesgourges. “The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview”. In: (Apr. 14, 2011). arXiv: 1104.2932v2 [astro-ph.IM].
- [93] Daniel J. Eisenstein and Wayne Hu. “Baryonic Features in the Matter Transfer Function”. In: *The Astrophysical Journal* 496.2 (Apr. 1998), pp. 605–614. ISSN: 1538-4357.
- [94] Daniel J. Eisenstein and Wayne Hu. “Power Spectra for Cold Dark Matter and Its Variants”. In: *The Astrophysical Journal* 511.1 (Jan. 1999), pp. 5–15. ISSN: 1538-4357.
- [95] Planck Collaboration. “Planck2018 results: VI. Cosmological parameters”. In: *Astronomy and Astrophysics* 641 (Sept. 2020), A6. ISSN: 1432-0746.
- [96] A. Friedman. “über die Krummung des Raumes”. In: *Zeitschrift fur Physik* 10.1 (Dec. 1922), pp. 377–386. ISSN: 1434-601X.
- [97] H. P. Robertson. “ON THE FOUNDATIONS OF RELATIVISTIC COSMOLOGY”. In: *Proceedings of the National Academy of Sciences* 15.11 (Nov. 1929), pp. 822–829. ISSN: 1091-6490.
- [98] A. G. Walker. “On Milne’s Theory of World-Structure*”. In: *Proceedings of the London Mathematical Society* s2-42.1 (1937), pp. 90–127. ISSN: 0024-6115.
- [99] A. G. Lematre. “A Homogeneous Universe of Constant Mass and Increasing Radius accounting for the Radial Velocity of Extra-galactic Nebulae”. In: *Monthly Notices of the Royal Astronomical Society* 91.5 (Mar. 1931), pp. 483–490. ISSN: 1365-2966.
- [100] Albert Einstein. “The Field Equations of Gravitation”. In: *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* 1915 (1915), pp. 844–847.

- [101] Edward W. Kolb and Michael S. Turner. *The early universe*. Vol. 69. 1990.
- [102] Andrew R. Liddle and David H. Lyth. *Cosmological Inflation and Large Scale Structure*. 2000.
- [103] Julien Lesgourgues. “An overview of Cosmology”. In: (Sept. 17, 2004). arXiv: astro-ph/0409426v1.
- [104] Mark Trodden and Sean M. Carroll. “TASI Lectures: Introduction to Cosmology”. In: (Jan. 26, 2004). arXiv: astro-ph/0401547v1.
- [105] David Langlois. “Inflation, quantum fluctuations and cosmological perturbations”. In: (May 7, 2004). arXiv: hep-th/0405053v1.
- [106] D. Langlois. “Inflation and Cosmological Perturbations”. In: *Lecture Notes in Physics*. Springer Berlin Heidelberg, 2010, pp. 1–57. ISBN: 9783642105982.
- [107] Adam G. Riess et al. “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant”. In: *The Astronomical Journal* 116.3 (Sept. 1998), pp. 1009–1038. ISSN: 0004-6256.
- [108] Shadab Alam et al. “The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample”. In: *Monthly Notices of the Royal Astronomical Society* 470.3 (Mar. 2017), pp. 2617–2652. ISSN: 1365-2966.
- [109] F. R. Bouchet et al. “Perturbative Lagrangian Approach to Gravitational Instability”. In: (1994).
- [110] T. Buchert. “A class of solutions in Newtonian cosmology and the pancake theory”. In: *Astronomy and Astrophysics* 223.1-2 (Oct. 1989), pp. 9–24.
- [111] O. Hahn et al. “Properties of dark matter haloes in clusters, filaments, sheets and voids”. In: *Monthly Notices of the Royal Astronomical Society* 375.2 (Feb. 2007), pp. 489–499. ISSN: 1365-2966.
- [112] Guilhem Lavaux and Benjamin D. Wandelt. “Precision cosmology with voids: definition, methods, dynamics”. In: *Monthly Notices of the Royal Astronomical Society* 403.3 (Apr. 2010), pp. 1392–1408. ISSN: 1365-2966.
- [113] AL Melott, T Buchert, and AG Weiss. “Testing Higher Order Lagrangian Perturbation Theory Against Numerical Simulations 2. Hierarchical Models”. In: *arXiv preprint astro-ph/9404018* (1994).
- [114] E. Bertschinger. “Self-similar secondary infall and accretion in an Einstein-de Sitter universe”. In: *The Astrophysical Journal Supplement Series* 58 (May 1985), p. 39. ISSN: 1538-4365.

- [115] S. D. M. White and J. Silk. “The growth of aspherical structure in the universe - Is the Local Supercluster an unusual system”. In: *The Astrophysical Journal* 231 (July 1979), p. 1. ISSN: 1538-4357.
- [116] J. M. Bardeen et al. “The statistics of peaks of Gaussian random fields”. In: *The Astrophysical Journal* 304 (May 1986), p. 15. ISSN: 1538-4357.
- [117] R. K. Sheth, H. J. Mo, and G. Tormen. “Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes”. In: *Monthly Notices of the Royal Astronomical Society* 323.1 (May 2001), pp. 1–12. ISSN: 1365-2966.
- [118] Ravi K. Sheth and Giuseppe Tormen. “An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier”. In: *Monthly Notices of the Royal Astronomical Society* 329.1 (Jan. 2002), pp. 61–75. ISSN: 1365-2966.
- [119] Ayako Yoshisato et al. “Why is the Zeldovich Approximation So Accurate?” In: *The Astrophysical Journal* 637.2 (Feb. 2006), pp. 555–560. ISSN: 1538-4357.
- [120] Edmund Bertschinger and Bhuvnesh Jain. “Gravitational instability of cold matter”. In: *The Astrophysical Journal* 431 (Aug. 1994), p. 486. ISSN: 1538-4357.
- [121] Peter Coles, Adrian L. Melott, and Sergei F. Shandarin. “Testing approximations for nonlinear gravitational clustering”. In: (Apr. 1992).
- [122] Dipak Munshi and Alexei A. Starobinsky. “Nonlinear approximations to gravitational instability: A comparison in second-order perturbation theory”. In: *The Astrophysical Journal* 428 (June 1994), p. 433. ISSN: 1538-4357.
- [123] Lev A. Kofman and Sergey F. Shandarin. “Theory of adhesion for the large-scale structure of the Universe”. In: *Nature* 334.6178 (July 1988), pp. 129–131. ISSN: 1476-4687.
- [124] Lev Kofman et al. “Coherent structures in the universe and the adhesion model”. In: *The Astrophysical Journal* 393 (July 1992), p. 437. ISSN: 1538-4357.
- [125] Patrick Valageas and Francis Bernardeau. “Density fields and halo mass functions in the geometrical adhesion toy model”. In: *Physical Review D* 83.4 (Feb. 2011). ISSN: 1550-2368.
- [126] P. J. E. Peebles. *Open Problems in Cosmology*. 2003. arXiv: astro - ph / 0311435 [astro-ph].
- [127] Paul J. Steinhardt and Neil Turok. “Why the Cosmological Constant Is Small and Positive”. In: *Science* 312.5777 (May 2006), pp. 1180–1183. ISSN: 1095-9203.
- [128] G F R Ellis. “Dark energy and inhomogeneity”. In: *Journal of Physics: Conference Series* 189 (Oct. 2009), p. 012011. ISSN: 1742-6596.
- [129] G. Gilmore and H. M. Asiri. “Open Issues in the Evolution of the Galactic Disks”. In: *Stellar Clusters and Associations- A RIA Workshop on Gaia*. Jan. 2011, pp. 280–285.

- [130] Jacques Colin et al. "Evidence for anisotropy of cosmic acceleration". In: *Astronomy and Astrophysics* 631 (Nov. 2019), p. L13. ISSN: 1432-0746.
- [131] C Krishnan et al. "Does Hubble tension signal a breakdown in FLRW cosmology?" In: *Classical and Quantum Gravity* 38.18 (Aug. 2021), p. 184001. ISSN: 1361-6382.
- [132] Elcio Abdalla et al. "Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies". In: *Journal of High Energy Astrophysics* 34 (June 2022), pp. 49–211. ISSN: 2214-4048.
- [133] J. S. Bagla and S. Ray. "Comments on the size of the simulation box in cosmological N-body simulations". In: *Monthly Notices of the Royal Astronomical Society* 358.3 (Apr. 2005), pp. 1076–1082. ISSN: 1365-2966.
- [134] J. S. Bagla and J. Prasad. "Effects of the size of cosmological N-body simulations on physical quantities - I. Mass function". In: *Monthly Notices of the Royal Astronomical Society* 370.2 (Aug. 2006), pp. 993–1002. ISSN: 1365-2966.
- [135] J.S. Bagla, Jayanti Prasad, and Nishikanta Khandai. "Effects of the size of cosmologicalN-body simulations on physical quantities - III. Skewness". In: *Monthly Notices of the Royal Astronomical Society* 395.2 (May 2009), pp. 918–930. ISSN: 1365-2966.
- [136] Eleonora Di Valentino et al. "In the realm of the Hubble tension- a review of solutions". In: *Classical and Quantum Gravity* 38.15 (July 2021), p. 153001. ISSN: 1361-6382.
- [137] P. J. E. Peebles. "The Gravitational-Instability Picture and the Nature of the Distribution of Galaxies". In: *The Astrophysical Journal* 189 (Apr. 1974), p. L51. ISSN: 1538-4357.
- [138] J.S. Bagla and T. Padmanabhan. "Transfer of power in non-linear gravitational clustering". In: *Monthly Notices of the Royal Astronomical Society* 286.4 (Apr. 1997), pp. 1023–1031. ISSN: 1365-2966.
- [139] M. J. Kirkby. "The fractal geometry of nature. Benoit B. Mandelbrot. W. H. Freeman and co." In: *Earth Surface Processes and Landforms* 8.4 (July 1983), pp. 406–406. ISSN: 1096-9837.
- [140] J. Schmalzing, M. Kerscher, and T. Buchert. "Minkowski Functionals in Cosmology". In: (1995).
- [141] *Statistical Physics for Cosmic Structures*. Springer-Verlag, 2005. ISBN: 3540407456.
- [142] Ofer Lahav and Yasushi Suto. "Measuring our Universe from Galaxy Redshift Surveys". In: *Living Reviews in Relativity* 7.1 (July 2004). ISSN: 1433-8351.
- [143] Eric D. Feigelson et al. "Twenty-First-Century Statistical and Computational Challenges in Astrophysics". In: *Annual Review of Statistics and Its Application* 8.1 (Mar. 2021), pp. 493–517. ISSN: 2326-831X.

- [144] A.A. Starobinsky. “Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations”. In: *Physics Letters B* 117.3-4 (Nov. 1982), pp. 175–178. ISSN: 0370-2693.
- [145] James M. Bardeen, Paul J. Steinhardt, and Michael S. Turner. “Spontaneous creation of almost scale-free density perturbations in an inflationary universe”. In: *Physical Review D* 28.4 (Aug. 1983), pp. 679–693. ISSN: 0556-2821.
- [146] S. W. HAWKING. “THE DEVELOPMENT OF IRREGULARITIES IN A SINGLE BUBBLE INFLATIONARY UNIVERSE”. In: *Advanced Series in Astrophysics and Cosmology*. WORLD SCIENTIFIC, June 1993, pp. 144–146.
- [147] R. I. Epstein. “Proto-galactic perturbations”. In: *Monthly Notices of the Royal Astronomical Society* 205.1 (Nov. 1983), pp. 207–229. ISSN: 1365-2966.
- [148] J. A. Peacock and A. F. Heavens. “Alternatives to the Press-Schechter cosmological mass function”. In: *Monthly Notices of the Royal Astronomical Society* 243.1 (Oct. 1990), pp. 133–143. ISSN: 1365-2966.
- [149] J. R. Bond et al. “Excursion set mass functions for hierarchical Gaussian fluctuations”. In: *The Astrophysical Journal* 379 (Oct. 1991), p. 440. ISSN: 1538-4357.
- [150] J. R. Bond and S. T. Myers. “The Peak-Patch Picture of Cosmic Catalogs. I. Algorithms”. In: *The Astrophysical Journal Supplement Series* 103 (Mar. 1996), p. 1. ISSN: 1538-4365.
- [151] M. Davis and P. J. E. Peebles. “On the integration of the BBGKY equations for the development of strongly nonlinear clustering in an expanding universe”. In: *The Astrophysical Journal Supplement Series* 34 (Aug. 1977), p. 425. ISSN: 1538-4365.
- [152] A. J. S. Hamilton et al. “Reconstructing the primordial spectrum of fluctuations of the universe from the observed nonlinear clustering of galaxies”. In: *The Astrophysical Journal* 374 (June 1991), p. L1. ISSN: 1538-4357.
- [153] R. Nityananda and T. Padmanabhan. “Scaling properties of non-linear gravitational clustering”. In: *Monthly Notices of the Royal Astronomical Society* 271.4 (Dec. 1994), pp. 976–980. ISSN: 1365-2966.
- [154] J. A. Peacock and S. J. Dodds. “Reconstructing the linear power spectrum of cosmological mass fluctuations”. In: *Monthly Notices of the Royal Astronomical Society* 267.4 (Apr. 1994), pp. 1020–1034. ISSN: 1365-2966.
- [155] T. Padmanabhan. “Modelling the non-linear gravitational clustering in the expanding Universe”. In: *Monthly Notices of the Royal Astronomical Society* 278.2 (Jan. 1996), pp. L29–L33. ISSN: 1365-2966.
- [156] T. Padmanabhan et al. “Patterns in Nonlinear Gravitational Clustering: A Numerical Investigation”. In: *The Astrophysical Journal* 466 (Aug. 1996), p. 604. ISSN: 1538-4357.

- [157] R. E. Smith et al. “Stable clustering, the halo model and non-linear cosmological power spectra”. In: *Monthly Notices of the Royal Astronomical Society* 341.4 (June 2003), pp. 1311–1332. issn: 1365-2966.
- [158] Siyu He et al. “Learning to predict the cosmological structure formation”. In: *Proceedings of the National Academy of Sciences* 116.28 (June 2019), pp. 13825–13832. issn: 1091-6490.
- [159] Helen Meskhidze. “Can Machine Learning Provide Understanding? How Cosmologists Use Machine Learning to Understand Observations of the Universe”. In: *Erkenntnis* 88.5 (July 2021), pp. 1895–1909. issn: 1572-8420.
- [160] A Ciprijanovic et al. “DeepMerge - II. Building robust deep learning algorithms for merging galaxy identification across domains”. In: *Monthly Notices of the Royal Astronomical Society* 506.1 (June 2021), pp. 677–691. issn: 1365-2966.
- [161] Francisco Villaescusa-Navarro et al. “Multifield Cosmology with Artificial Intelligence”. In: (Sept. 20, 2021). arXiv: 2109.09747v1 [astro-ph.CO].
- [162] Davide Piras, Benjamin Joachimi, and Francisco Villaescusa-Navarro. “Fast and realistic large-scale structure from machine-learning-augmented random field simulations”. In: *Monthly Notices of the Royal Astronomical Society* 520.1 (Jan. 2023), pp. 668–683. issn: 1365-2966.
- [163] Francisco Villaescusa-Navarro et al. “The CAMELS Multifield Data Set: Learning the Universe’s Fundamental Parameters with Artificial Intelligence”. In: *The Astrophysical Journal Supplement Series* 259.2 (Apr. 2022), p. 61. issn: 1538-4365.
- [164] Rudiger Pakmor et al. “The MillenniumTNG Project: the hydrodynamical full physics simulation and a first look at its galaxy clusters”. In: *Monthly Notices of the Royal Astronomical Society* 524.2 (July 2023), pp. 2539–2555. issn: 1365-2966.
- [165] Euclid Collaboration. “Euclid preparation: VII. Forecast validation for Euclid cosmological probes”. In: *Astronomy and Astrophysics* 642 (Oct. 2020), A191. issn: 1432-0746.
- [166] R. W Hockney and J. W Eastwood. *Computer Simulation Using Particles*. Taylor and Francis, 1988. isbn: 9780852743928.
- [167] J S Bagla and T Padmanabhan. “Cosmological N-body simulations”. In: *Pramana* 49.2 (Aug. 1997), pp. 161–192. issn: 0973-7111.
- [168] T. Buchert. “Lagrangian theory of gravitational instability of Friedman-Lematre cosmologies - a generic third-order model for non-linear clustering”. In: *Monthly Notices of the Royal Astronomical Society* 267.4 (Apr. 1994), pp. 811–820. issn: 1365-2966.
- [169] M. Crocce, S. Pueblas, and R. Scoccimarro. “Transients from initial conditions in cosmological simulations”. In: *Monthly Notices of the Royal Astronomical Society* 373.1 (Nov. 2006), pp. 369–381. issn: 1365-2966.

- [170] Oliver Hahn and Tom Abel. “Multi-scale initial conditions for cosmological simulations: Multi-scale initial conditions”. In: *Monthly Notices of the Royal Astronomical Society* 415.3 (July 2011), pp. 2101–2121. ISSN: 0035-8711.
- [171] Michael Michaux et al. “Accurate initial conditions for cosmological N-body simulations: minimizing truncation and discreteness errors”. In: *Monthly Notices of the Royal Astronomical Society* 500.1 (Nov. 2020), pp. 663–683. ISSN: 1365-2966.
- [172] Oliver Hahn, Cornelius Rampf, and Cora Uhlemann. “Higher order initial conditions for mixed baryon-CDM simulations”. In: *Monthly Notices of the Royal Astronomical Society* 503.1 (Dec. 2020), pp. 426–445. ISSN: 1365-2966.
- [173] P. P. Ewald. “Die Berechnung optischer und elektrostatischer Gitterpotentiale”. In: *Annalen der Physik* 369.3 (Jan. 1921), pp. 253–287. ISSN: 1521-3889.
- [174] A. A. Klypin and S. F. Shandarin. “Three-dimensional numerical model of the formation of large-scale structure in the Universe”. In: *Monthly Notices of the Royal Astronomical Society* 204.3 (Oct. 1983), pp. 891–907. ISSN: 1365-2966.
- [175] G. Efstathiou et al. “Numerical techniques for large cosmological N-body simulations”. In: *The Astrophysical Journal Supplement Series* 57 (Feb. 1985), p. 241. ISSN: 1538-4365.
- [176] Josh Barnes and Piet Hut. “A hierarchical O($N \log N$) force-calculation algorithm”. In: *Nature* 324.6096 (Dec. 1986), pp. 446–449. ISSN: 1476-4687.
- [177] Abdulnour Y. Toukmaji and Jr. Board John A. “Ewald summation techniques in perspective: a survey”. In: *Computer Physics Communications* 95.2 (June 1996), pp. 73–92.
- [178] Thomas Quinn et al. *Time stepping N-body simulations*. 1997.
- [179] Volker Springel, Naoki Yoshida, and Simon D.M. White. “GADGET: a code for collisionless and gasdynamical cosmological simulations”. In: *New Astronomy* 6.2 (Apr. 2001), pp. 79–117. ISSN: 1384-1076.
- [180] Lehman H. Garrison et al. “Improving initial conditions for cosmological N-body simulations”. In: *Monthly Notices of the Royal Astronomical Society* 461.4 (July 2016), pp. 4125–4145. ISSN: 1365-2966.
- [181] Lehman H. Garrison, Michael Joyce, and Daniel J. Eisenstein. “Good and proper: self-similarity of N-body simulations with proper force softening”. In: *Monthly Notices of the Royal Astronomical Society* 504.3 (Apr. 2021), pp. 3550–3560. ISSN: 1365-2966.
- [182] Bryce Kuhlman, Adrian L. Melott, and Sergei F. Shandarin. “A Test of the Particle Paradigm in [ITAL]N/[ITAL]-Body Simulations”. In: *The Astrophysical Journal* 470.1 (Oct. 1996), pp. L41–L44. ISSN: 0004-637X.

- [183] Randall J. Splinter et al. “Fundamental Discreteness Limitations of Cosmological N-Body Clustering Simulations”. In: *The Astrophysical Journal* 497.1 (Apr. 1998), pp. 38–61. ISSN: 1538-4357.
- [184] Michael Joyce, Lehman Garrison, and Daniel Eisenstein. “Quantifying resolution in cosmological N-body simulations using self-similarity”. In: *Monthly Notices of the Royal Astronomical Society* 501.4 (Nov. 2020), pp. 5051–5063. ISSN: 1365-2966.
- [185] Joaquin Armijo et al. “Making use of sub-resolution haloes in N-body simulations”. In: *Monthly Notices of the Royal Astronomical Society: Letters* 510.1 (Nov. 2021), pp. 29–33. ISSN: 1745-3933.
- [186] Hideki Yahagi and Yuzuru Yoshii. “N-Body Code with Adaptive Mesh Refinement”. In: *The Astrophysical Journal* 558.1 (Sept. 2001), pp. 463–475. ISSN: 1538-4357.
- [187] Anatoly Klypin and Francisco Prada. “Effects of long-wavelength fluctuations in large galaxy surveys”. In: *Monthly Notices of the Royal Astronomical Society* 489.2 (Aug. 2019), pp. 1684–1696. ISSN: 1365-2966.
- [188] M. Davis et al. “The evolution of large-scale structure in a universe dominated by cold dark matter”. In: *The Astrophysical Journal* 292 (May 1985), p. 371. ISSN: 1538-4357.
- [189] Surhud More et al. “THE OVERDENSITY AND MASSES OF THE FRIENDS-OF-FRIENDS HALOS AND UNIVERSALITY OF HALO MASS FUNCTION”. In: *The Astrophysical Journal Supplement Series* 195.1 (June 2011), p. 4. ISSN: 1538-4365.
- [190] Michael S. Warren et al. “Dark halos formed via dissipationless collapse. I - Shapes and alignment of angular momentum”. In: *The Astrophysical Journal* 399 (Nov. 1992), p. 405. ISSN: 1538-4357.
- [191] Cedric Lacey and Shau Cole. “Merger rates in hierarchical models of galaxy formation - II. Comparison with N-body simulations”. In: *Monthly Notices of the Royal Astronomical Society* 271.3 (Dec. 1994), pp. 676–692. ISSN: 1365-2966.
- [192] Mark C. Neyrinck. “zobov: a parameter-free void-finding algorithm”. In: *Monthly Notices of the Royal Astronomical Society* 386.4 (June 2008), pp. 2101–2109. ISSN: 1365-2966.
- [193] Marius Cautun et al. “Evolution of the cosmic web”. In: *Monthly Notices of the Royal Astronomical Society* 441.4 (May 2014), pp. 2923–2973. ISSN: 0035-8711.
- [194] M. Ramella et al. “Finding galaxy clusters using Voronoi tessellations”. In: *Astronomy and Astrophysics* 368.3 (Mar. 2001), pp. 776–786. ISSN: 1432-0746.
- [195] E. Platen, R. Van De Weygaert, and B. J. T. Jones. “A cosmic watershed: the WVF void detection technique”. In: *Monthly Notices of the Royal Astronomical Society* 380.2 (Sept. 2007), pp. 551–570. ISSN: 1365-2966.

- [196] Yin Lin et al. “The distribution of bubble sizes during reionization”. In: *Monthly Notices of the Royal Astronomical Society* 461.3 (June 2016), pp. 3361–3374. ISSN: 1365-2966.
- [197] X. Xu et al. “Finding cosmic voids and filament loops using topological data analysis”. In: *Astronomy and Computing* 27 (Apr. 2019), pp. 34–52. ISSN: 2213-1337.
- [198] Alejandro Palomino et al. “Cosmic Web Classification through Stochastic Topological Ranking”. In: (Apr. 1, 2024). arXiv: 2404.01124v1 [astro-ph.CO].
- [199] S. Gottloeber. “Galaxy tracers in cosmological N-body simulations”. In: (Nov. 26, 1997). arXiv: astro-ph/9711312v1.
- [200] Mark C. Neyrinck, Nickolay Y. Gnedin, and Andrew J. S. Hamilton. “voboz: an almost-parameter-free halo-finding algorithm”. In: *Monthly Notices of the Royal Astronomical Society* 356.4 (Feb. 2005), pp. 1222–1232. ISSN: 1365-2966.
- [201] Volker Springel et al. “Populating a cluster of galaxies I. Results at z=0”. In: *Monthly Notices of the Royal Astronomical Society* 328.3 (Dec. 2001), pp. 726–750. ISSN: 1365-2966.
- [202] D. Aubert, C. Pichon, and S. Colombi. “The origin and implications of dark matter anisotropic cosmic infall on haloes”. In: *Monthly Notices of the Royal Astronomical Society* 352.2 (Aug. 2004), pp. 376–398. ISSN: 1365-2966.
- [203] D. Tweed et al. “Building merger trees from cosmologicalN-body simulations: Towards improving galaxy formation models using subhaloes”. In: *Astronomy and Astrophysics* 506.2 (Aug. 2009), pp. 647–660. ISSN: 1432-0746.
- [204] Peter S. Behroozi, Risa H. Wechsler, and Hao-Yi Wu. “THE ROCKSTAR PHASE-SPACE TEMPORAL HALO FINDER AND THE VELOCITY OFFSETS OF CLUSTER CORES”. In: *The Astrophysical Journal* 762.2 (Dec. 2012), p. 109. ISSN: 1538-4357.
- [205] Pascal J. Elahi et al. “Hunting for galaxies and halos in simulations with VELOCIraptor”. In: *Publications of the Astronomical Society of Australia* 36 (2019). ISSN: 1448-6083.
- [206] Shuang-Peng Sun et al. “HIKER: a halo-finding method based on kernel-shift algorithm”. In: *Research in Astronomy and Astrophysics* 20.4 (Apr. 2020), p. 046. ISSN: 1674-4527.
- [207] Boryana Hadzhiyska et al. “<scp>compaso</scp>: A new halo finder for competitive assignment to spherical overdensities”. In: *Monthly Notices of the Royal Astronomical Society* 509.1 (Oct. 2021), pp. 501–521. ISSN: 1365-2966.
- [208] Volker Springel et al. “Simulating cosmic structure formation with the GADGET-4 code”. In: *Monthly Notices of the Royal Astronomical Society* 506.2 (July 2021), pp. 2871–2949. ISSN: 1365-2966.
- [209] Jeremy L. Tinker. “A Self-Calibrating Halo-Based Galaxy Group Finder: Algorithm and Tests”. In: (July 23, 2020). arXiv: 2007.12200v1 [astro-ph.GA].

- [210] David Valles-Perez, Susana Planelles, and Vicent Quilis. “The halo-finding problem revisited: a deep revision of the ASOHF code”. In: *Astronomy and Astrophysics* 664 (Aug. 2022), A42. ISSN: 1432-0746.
- [211] Sara Maleubre et al. “Constraining accuracy of the pairwise velocities in N-body simulations using scale-free models”. In: *Monthly Notices of the Royal Astronomical Society* 525.1 (Aug. 2023), pp. 1039–1052. ISSN: 1365-2966.
- [212] Manodeep Sinha and Lehman H Garrison. “corrfunc - a suite of blazing fast correlation functions on the CPU”. In: *Monthly Notices of the Royal Astronomical Society* 491.2 (Nov. 2019), pp. 3022–3041. ISSN: 1365-2966.
- [213] J. A. Peacock and R. E. Smith. “Halo occupation numbers and galaxy bias”. In: *Monthly Notices of the Royal Astronomical Society* 318.4 (Nov. 2000), pp. 1144–1156. arXiv: astro-ph/0005010 [astro-ph].
- [214] Uroš Seljak. “Analytic model for galaxy and dark matter clustering”. In: *Monthly Notices of the Royal Astronomical Society* 318.1 (Oct. 2000), pp. 203–213. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/318/1/203/3943998/318-1-203.pdf>.
- [215] A COORAY and R SHETH. “Halo models of large scale structure”. In: *Physics Reports* 372.1 (Dec. 2002), pp. 1–129. ISSN: 0370-1573.
- [216] Rachel Mandelbaum et al. “Galaxy-galaxy lensing: dissipationless simulations versus the halo model”. In: *Monthly Notices of the Royal Astronomical Society* 362.4 (Oct. 2005), pp. 1451–1462. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/362/4/1451/3114897/362-4-1451.pdf>.
- [217] A. J. Mead et al. “An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models”. In: *Monthly Notices of the Royal Astronomical Society* 454.2 (Oct. 2015), pp. 1958–1975. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/454/2/1958/9381114/stv2036.pdf>.
- [218] Ramin A. Skibba and Ravi K. Sheth. “A halo model of galaxy colours and clustering in the Sloan Digital Sky Survey”. In: *Monthly Notices of the Royal Astronomical Society* 392.3 (Jan. 2009), pp. 1080–1091. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/392/3/1080/3647280/mnras0392-1080.pdf>.
- [219] Shadab Alam et al. “Multitracer extension of the halo model: probing quenching and conformity in eBOSS”. In: *Monthly Notices of the Royal Astronomical Society* 497.1 (July 2020), pp. 581–595. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/497/1/581/33528083/staa1956.pdf>.

- [220] Bradley Greig et al. “Exploring the role of the halo mass function for inferring astrophysical parameters during reionisation”. In: *arXiv e-prints*, arXiv:2403.14061 (Mar. 2024), arXiv:2403.14061. arXiv: 2403 . 14061 [astro-ph.CO].
- [221] Simon D. M. White and Carlos S. Frenk. “Galaxy Formation through Hierarchical Clustering”. In: *The Astrophysical Journal* 379 (Sept. 1991), p. 52.
- [222] Subhabrata Majumdar and Joseph J. Mohr. “Importance of Cluster Structural Evolution in Using X-Ray and Sunyaev-Zeldovich Effect Galaxy Cluster Surveys to Study Dark Energy”. In: *The Astrophysical Journal* 585.2 (Mar. 2003), pp. 603–610.
- [223] Steven W. Allen, August E. Evrard, and Adam B. Mantz. “Cosmological Parameters from Observations of Galaxy Clusters”. In: *Annual Review of Astronomy and Astrophysics* 49.1 (2011), pp. 409–470. eprint: <https://doi.org/10.1146/annurev-astro-081710-102514>.
- [224] C To et al. “Dark Energy Survey Year 1 Results: Cosmological Constraints from Cluster Abundances, Weak Lensing, and Galaxy Correlations”. In: *Physical review letters* 126.14 (2021), p. 141301.
- [225] J.D. Cohn, J.S. Bagla, and Martin White. “A comparison of simulated and analytic major merger counts”. In: *Monthly Notices of the Royal Astronomical Society* 325.3 (Aug. 2001), pp. 1053–1064. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/325/3/1053/2840461/325-3-1053.pdf>.
- [226] Carlo Giocoli, Lidia Pieri, and Giuseppe Tormen. “Analytical approach to subhalo population in dark matter haloes”. In: *Monthly Notices of the Royal Astronomical Society* 387.2 (May 2008), pp. 689–697. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/387/2/689/3373044/mnras0387-0689.pdf>.
- [227] Yacine Ali-Haïmoud, Ely D. Kovetz, and Marc Kamionkowski. “Merger rate of primordial black-hole binaries”. In: *Phys. Rev. D* 96 (12 Dec. 2017), p. 123523.
- [228] Saeed Fakhry, Javad T. Firouzjaee, and Mehrdad Farhoudi. “Primordial black hole merger rate in ellipsoidal-collapse dark matter halo models”. In: *Phys. Rev. D* 103 (12 June 2021), p. 123014.
- [229] Matthias Bartelmann et al. “Arc statistics with realistic cluster potentials. IV. Clusters in different cosmologies”. In: *Arxiv preprint astro-ph/9707167* (1997).
- [230] Matthias Bartelmann. “Gravitational lensing”. In: *Classical and Quantum Gravity* 27.23 (Nov. 2010), p. 233001.
- [231] Richard Massey, Thomas Kitching, and Johan Richard. “The dark matter of gravitational lensing”. In: *Reports on Progress in Physics* 73.8 (July 2010), p. 086901.

- [232] Michele Maggiore and Antonio Riotto. “THE HALO MASS FUNCTION FROM EXCURSION SET THEORY. I. GAUSSIAN FLUCTUATIONS WITH NON-MARKOVIAN DEPENDENCE ON THE SMOOTHING SCALE”. In: *The Astrophysical Journal* 711.2 (Feb. 2010), pp. 907–927.
- [233] Dionysios Karagiannis et al. “Constraining primordial non-Gaussianity with bispectrum and power spectrum from upcoming optical and radio surveys”. In: *Monthly Notices of the Royal Astronomical Society* 478.1 (Apr. 2018), pp. 1341–1376.
- [234] Masato Shirasaki et al. “Constraining primordial non-Gaussianity with postreconstructed galaxy bispectrum in redshift space”. In: *Phys. Rev. D* 103 (2 Jan. 2021), p. 023506.
- [235] Shiming Gu et al. “On constraining cosmology and the halo mass function with weak gravitational lensing”. In: *Monthly Notices of the Royal Astronomical Society* 525.4 (Nov. 2023), pp. 4871–4886. arXiv: 2302.00780 [astro-ph.CO].
- [236] William H Press and Paul Schechter. “Formation of galaxies and clusters of galaxies by self-similar gravitational condensation”. In: *The Astrophysical Journal* 187 (1974), pp. 425–438.
- [237] George Efstathiou et al. “Gravitational clustering from scale-free initial conditions.” In: *Monthly Notices of the Royal Astronomical Society* 235 (Dec. 1988), pp. 715–748.
- [238] R. G. Carlberg and H. M. P. Couchman. “Mergers and Bias in a Cold Dark Matter Cosmology”. In: *The Astrophysical Journal* 340 (May 1989), p. 47.
- [239] Pierluigi Monaco. “A Lagrangian Dynamical Theory for the Mass Function of Cosmic Structures - I. Dynamics”. In: *Monthly Notices of the Royal Astronomical Society* 287.4 (June 1997), pp. 753–770. arXiv: astro-ph/9606027 [astro-ph].
- [240] Edouard Audit, Romain Teyssier, and Jean-Michel Alimi. “Non-linear dynamics and mass function of cosmic structures. II. Numerical results”. In: *Astronomy & Astrophysics* 333 (May 1998), pp. 779–789. arXiv: astro-ph/9802287 [astro-ph].
- [241] Jounghun Lee and Sergei F. Shandarin. “Comparison of Analytical Mass Functions with Numerical Simulations”. In: *The Astrophysical Journal* 517.1 (May 1999), pp. L5–L8.
- [242] Ravi K. Sheth and Giuseppe Tormen. “Large-scale bias and the peak background split”. In: *Monthly Notices of the Royal Astronomical Society* 308.1 (Sept. 1999), pp. 119–126. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/308/1/119/18409158/308-1-119.pdf>.
- [243] H. J. Mo and S. D. M. White. “An analytic model for the spatial clustering of dark matter haloes”. In: *Monthly Notices of the Royal Astronomical Society* 282.2 (Sept. 1996), pp. 347–361. arXiv: astro-ph/9512127 [astro-ph].

- [244] Aseem Paranjape and Ravi K. Sheth. “Peaks theory and the excursion set approach”. In: *Monthly Notices of the Royal Astronomical Society* 426.4 (Nov. 2012), pp. 2789–2796. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/426/4/2789/3300413/426-4-2789.pdf>.
- [245] Joung-hun Lee and Sergei F. Shandarin. “The Cosmological Mass Distribution Function in the Zeldovich Approximation”. In: *The Astrophysical Journal* 500.1 (June 1998), pp. 14–27.
- [246] A. Jenkins et al. “The mass function of dark matter haloes”. In: *Monthly Notices of the Royal Astronomical Society* 321.2 (Feb. 2001), pp. 372–384. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/321/2/372/3032036/321-2-372.pdf>.
- [247] Darren S. Reed et al. “The halo mass function from the dark ages through the present day”. In: *Monthly Notices of the Royal Astronomical Society* 374.1 (Nov. 2006), pp. 2–15. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/374/1/2/2835466/mnras0374-0002.pdf>.
- [248] J. Courtin et al. “Imprints of dark energy on cosmic structure formation – II. Non-universality of the halo mass function”. In: *Monthly Notices of the Royal Astronomical Society* 410.3 (Jan. 2011), pp. 1911–1931. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/410/3/1911/2864964/mnras0410-1911.pdf>.
- [249] Benedikt Diemer. “Universal at Last? The Splashback Mass Function of Dark Matter Halos”. In: *The Astrophysical Journal* 903.2 (Nov. 2020), p. 87.
- [250] Suman Bhattacharya et al. “MASS FUNCTION PREDICTIONS BEYOND Λ CDM”. In: *The Astrophysical Journal* 732.2 (Apr. 2011), p. 122.
- [251] Lurdes Ondaro-Mallea et al. “Non-universality of the mass function: dependence on the growth rate and power spectrum shape”. In: *Monthly Notices of the Royal Astronomical Society* 509.4 (Feb. 2022), pp. 6077–6090. arXiv: 2102.08958 [astro-ph.CO].
- [252] William A. Watson et al. “The halo mass function through the cosmic ages”. In: *Monthly Notices of the Royal Astronomical Society* 433.2 (June 2013), pp. 1230–1245. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/433/2/1230/4901588/stt791.pdf>.
- [253] Giulia Despali et al. “The universality of the virial halo mass function and models for non-universality of other halo definitions”. In: *Monthly Notices of the Royal Astronomical Society* 456.3 (Dec. 2015), pp. 2486–2504. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/456/3/2486/18469260/stv2842.pdf>.

- [254] M Sten Delos. “Accurate halo mass functions from the simplest excursion set theory”. In: *Monthly Notices of the Royal Astronomical Society* 528.2 (Jan. 2024), pp. 1372–1387. ISSN: 1365-2966.
- [255] John D. Barrow and Paul Saich. “Growth of large-scale structure with a cosmological constant”. In: *Monthly Notices of the Royal Astronomical Society* 262.3 (June 1993), pp. 717–725. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/262/3/717/3371391/mnras262-0717.pdf>.
- [256] Darren Reed et al. “Evolution of the mass function of dark matter haloes”. In: *Monthly Notices of the Royal Astronomical Society* 346.2 (Dec. 2003), pp. 565–572. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/346/2/565/4294343/346-2-565.pdf>.
- [257] Michael S. Warren et al. “Precision Determination of the Mass Function of Dark Matter Halos”. In: *The Astrophysical Journal* 646.2 (Aug. 2006), pp. 881–885.
- [258] Jeremy Tinker et al. “Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality”. In: *The Astrophysical Journal* 688.2 (Dec. 2008), pp. 709–728.
- [259] Martín Crocce et al. “Simulating the Universe with MICE: the abundance of massive clusters”. In: *Monthly Notices of the Royal Astronomical Society* 403.3 (Apr. 2010), pp. 1353–1367. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/403/3/1353/6170753/mnras0403-1353.pdf>.
- [260] R. E. Angulo et al. “Scaling relations for galaxy clusters in the Millennium-XXL simulation”. In: *Monthly Notices of the Royal Astronomical Society* 426.3 (Nov. 2012), pp. 2046–2062. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/426/3/2046/3108030/426-3-2046.pdf>.
- [261] Sebastian Bocquet et al. “Halo mass function: baryon impact, fitting formulae, and implications for cluster cosmology”. In: *Monthly Notices of the Royal Astronomical Society* 456.3 (Dec. 2015), pp. 2361–2373. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/456/3/2361/18466324/stv2657.pdf>.
- [262] Johan Comparat et al. “Accurate mass and velocity functions of dark matter haloes”. In: *Monthly Notices of the Royal Astronomical Society* 469.4 (May 2017), pp. 4157–4174. ISSN: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/469/4/4157/17712000/stx1183.pdf>.
- [263] Thomas McClintock et al. “The Aemulus Project. II. Emulating the Halo Mass Function”. In: *The Astrophysical Journal* 872.1 (Feb. 2019), p. 53.
- [264] Takahiro Nishimichi et al. “Dark Quest. I. Fast and Accurate Emulation of Halo Clustering Statistics and Its Application to Galaxy Clustering”. In: *The Astrophysical Journal* 884.1 (Oct. 2019), p. 29.

- [265] Sebastian Bocquet et al. “The Mira-Titan Universe. III. Emulation of the Halo Mass Function”. In: *The Astrophysical Journal* 901.1 (Sept. 2020), p. 5.
- [266] Seppi, R. et al. “The mass function dependence on the dynamical state of dark matter haloes”. In: *A&A* 652 (2021), A155.
- [267] Nishikanta Khandai et al. “The MassiveBlack-II simulation: the evolution of haloes and galaxies to $z \approx 0$ ”. In: *Monthly Notices of the Royal Astronomical Society* 450.2 (Apr. 2015), pp. 1349–1374. issn: 0035-8711. eprint: <https://academic.oup.com/mnras/article-pdf/450/2/1349/3020639/stv627.pdf>.
- [268] Euclid Collaboration. “Euclid preparation. XXXIX. The effect of baryons on the halo mass function”. In: *Astronomy & Astrophysics* 685, A109 (May 2024), A109. arXiv: 2311.01465 [astro-ph.CO].
- [269] Ningyuan Guo et al. “Deep learning insights into non-universality in the halo mass function”. In: *Monthly Notices of the Royal Astronomical Society* 532.4 (Aug. 2024), pp. 4141–4156. arXiv: 2405.15850 [astro-ph.CO].
- [270] M. Manera, Ravi K. Sheth, and R. Scoccimarro. “Large-scale bias and the inaccuracy of the peak-background split: Large-scale bias”. In: *Monthly Notices of the Royal Astronomical Society* 402.1 (Dec. 2009), pp. 589–602. issn: 0035-8711.
- [271] J. S. Bagla, Nishikanta Khandai, and Girish Kulkarni. “Mass function of haloes: scale invariant models”. In: *arXiv e-prints*, arXiv:0908.2702 (Aug. 2009), arXiv:0908.2702. arXiv: 0908.2702 [astro-ph.CO].
- [272] Marcello Musso and Ravi K. Sheth. “Excursion set peaks in energy as a model for haloes”. In: *Monthly Notices of the Royal Astronomical Society* 508.3 (Dec. 2021), pp. 3634–3648. arXiv: 1907.09147 [astro-ph.CO].
- [273] Marcello Musso and Ravi K. Sheth. “Getting in shape with minimal energy: a variational principle for protohaloes”. In: *Monthly Notices of the Royal Astronomical Society* 523.1 (July 2023), pp. L4–L8. arXiv: 2303.02142 [astro-ph.GA].
- [274] Marcello Musso, Giulia Despali, and Ravi K. Sheth. “The energy shear of protohaloes”. In: *arXiv e-prints*, arXiv:2405.20207 (May 2024), arXiv:2405.20207. arXiv: 2405.20207 [astro-ph.CO].
- [275] Emanuele Castorina et al. “Excursion set peaks: the role of shear”. In: (Nov. 2016). arXiv: 1611.03619 [astro-ph.CO].
- [276] Marcello Musso, Aseem Paranjape, and Ravi K. Sheth. “Scale-dependent halo bias in the excursion set approach”. In: *Monthly Notices of the Royal Astronomical Society* 427.4 (Dec. 2012), pp. 3145–3158. arXiv: 1205.3401 [astro-ph.CO].

- [277] Sujatha Ramakrishnan, Aseem Paranjape, and Ravi K. Sheth. “Mock halo catalogues: assigning unresolved halo properties using correlations with local halo environment”. In: *Monthly Notices of the Royal Astronomical Society* 503.2 (May 2021), pp. 2053–2064. arXiv: 2012 . 10170 [astro-ph.CO].
- [278] Wendy L. Freedman and Barry F. Madore. “The Hubble Constant”. In: *Annual Review of Astronomy and Astrophysics* 48.1 (Aug. 2010), pp. 673–710. ISSN: 1545-4282.
- [279] R. Brent Tully. *The Hubble Constant: A Historical Review*. 2023.
- [280] Licia Verde, Nils Schöneberg, and Héctor Gil-Marín. *A tale of many H_0* . 2023.
- [281] Samuel Brieden, Héctor Gil-Marín, and Licia Verde. “A tale of two (or more) h ’s”. In: *Journal of Cosmology and Astroparticle Physics* 2023.04 (Apr. 2023), p. 023. ISSN: 1475-7516.
- [282] Adam G. Riess et al. “Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with Λ CDM”. In: *The Astrophysical Journal Letters* 908.1 (Feb. 2021), p. L6.
- [283] Adam G. Riess et al. “A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km/s /Mpc Uncertainty from the Hubble Space Telescope and the SH0ES Team”. In: *The Astrophysical Journal Letters* 934.1 (July 2022), p. L7. ISSN: 2041-8213.
- [284] Dillon Brout et al. “The Pantheon+ Analysis: Cosmological Constraints”. In: *The Astrophysical Journal* 938.2 (Oct. 2022), p. 110. ISSN: 1538-4357.
- [285] Paul Shah, Pablo Lemos, and Ofer Lahav. “A buyers guide to the Hubble constant”. In: *The Astronomy and Astrophysics Review* 29.1 (Dec. 2021). ISSN: 1432-0754.
- [286] George Efstathiou. “To H_0 or not to H_0 ? ” In: *Monthly Notices of the Royal Astronomical Society* 505.3 (June 2021), pp. 3866–3872. ISSN: 1365-2966.
- [287] Sunny Vagnozzi. “Seven Hints That Early-Time New Physics Alone Is Not Sufficient to Solve the Hubble Tension”. In: *Universe* 9.9 (Aug. 2023), p. 393. ISSN: 2218-1997.
- [288] Edvard Mörtsell et al. “Sensitivity of the Hubble Constant Determination to Cepheid Calibration”. In: *The Astrophysical Journal* 933.2 (July 2022), p. 212. ISSN: 1538-4357.
- [289] Dragan Huterer and Hao-Yi Wu. *Not empty enough: a local void cannot solve the H_0 tension*. 2023.
- [290] Daniel Majaess. *The Concerning SH₀ES Hubble Constant*. 2024.
- [291] P. Shah et al. “The dark energy survey: detection of weak lensing magnification of supernovae and constraints on dark matter haloes”. In: *Monthly Notices of the Royal Astronomical Society* 532.1 (July 2024), pp. 932–944. arXiv: 2406 . 05047 [astro-ph.CO].

- [292] R. Camilleri et al. *The Dark Energy Survey Supernova Program: An updated measurement of the Hubble constant using the Inverse Distance Ladder*. 2024. arXiv: 2406 . 05049 [astro-ph.CO].
- [293] Hao-Yi Wu and Dragan Huterer. “Sample variance in the local measurements of the Hubble constant”. In: *Monthly Notices of the Royal Astronomical Society* 471.4 (Nov. 2017), pp. 4946–4955. arXiv: 1706 . 09723 [astro-ph.CO].
- [294] Pavan Kumar Aluri et al. “Is the observable Universe consistent with the cosmological principle?” In: *Classical and Quantum Gravity* 40.9, 094001 (May 2023), p. 094001. arXiv: 2207 . 05765 [astro-ph.CO].
- [295] Richard Watkins et al. “Analysing the large-scale bulk flow using cosmicflows4: increasing tension with the standard cosmological model”. In: *Monthly Notices of the Royal Astronomical Society* 524.2 (Sept. 2023), pp. 1885–1892. arXiv: 2302 . 02028 [astro-ph.CO].
- [296] Pilar Ruiz-Lapuente and Jonay I. González Hernández. “SNe Ia twins for life” towards a precise determination of H_0 . 2023.
- [297] Lu Huang, Shao-Jiang Wang, and Wang-Wei Yu. *No-go guide for the Hubble tension: late-time or local-scale new physics*. 2024.
- [298] Basheer Kalbouneh, Christian Marinoni, and Roy Maartens. *Cosmography of the Local Universe by Multipole Analysis of the Expansion Rate Fluctuation Field*. 2024.
- [299] B. Carreres et al. “ZTF SN Ia DR2: Peculiar velocities impact on the Hubble diagram”. In: *arXiv e-prints*, arXiv:2405.20409 (May 2024), arXiv:2405.20409. arXiv: 2405 . 20409 [astro-ph.CO].
- [300] Leonardo Giani et al. “An effective description of Laniakea: impact on cosmology and the local determination of the Hubble constant”. In: *Journal of Cosmology and Astroparticle Physics* 2024.1, 071 (Jan. 2024), p. 071. arXiv: 2311 . 00215 [astro-ph.CO].
- [301] Roy Maartens et al. *Covariant cosmography: the observer-dependence of the Hubble parameter*. 2023.
- [302] Gagandeep S. Anand et al. “Tip of the Red Giant Branch Distances with JWST: An Absolute Calibration in NGC 4258 and First Applications to Type Ia Supernova Hosts”. In: *The Astrophysical Journal* 966.1, 89 (May 2024), p. 89. arXiv: 2401 . 04776 [astro-ph.CO].
- [303] Adam G. Riess et al. “JWST Observations Reject Unrecognized Crowding of Cepheid Photometry as an Explanation for the Hubble Tension at 8σ Confidence”. In: *The Astrophysical Journal Letters* 962.1 (Feb. 2024), p. L17. ISSN: 2041-8213.

- [304] Edwin L. Turner, Renyue Cen, and Jeremiah P. Ostriker. “The relation of local measures of Hubble’s constant to its global value”. In: *The Astronomical Journal* 103 (May 1992), p. 1427.
- [305] Io Odderskov, Steen Hannestad, and Troels Haugbølle. “On the local variation of the Hubble constant”. In: *Journal of Cosmology and Astroparticle Physics* 2014.10 (Oct. 2014), pp. 028–028. arXiv: 1407.7364 [astro-ph.CO].
- [306] Radosław Wojtak et al. “Cosmic variance of the local Hubble flow in large-scale cosmological simulations”. In: *Monthly Notices of the Royal Astronomical Society* 438.2 (Feb. 2014), pp. 1805–1812. arXiv: 1312.0276 [astro-ph.CO].
- [307] Io Odderskov, Steen Hannestad, and Jacob Brandbyge. “The variance of the locally measured Hubble parameter explained with different estimators”. In: *Journal of Cosmology and Astroparticle Physics* 2017.3, 022 (Mar. 2017), p. 022. arXiv: 1701.05391 [astro-ph.CO].
- [308] Ofer Lahav et al. “Dynamical effects of the cosmological constant.” In: *Monthly Notices of the Royal Astronomical Society* 251 (July 1991), pp. 128–136.
- [309] K. Dolag and R. Sunyaev. “Relative velocity of dark matter and baryons in clusters of galaxies and measurements of their peculiar velocities”. In: *Monthly Notices of the Royal Astronomical Society* 432.2 (May 2013), pp. 1600–1615. ISSN: 1365-2966.
- [310] Wojciech A. Hellwing et al. “The effect of baryons on redshift space distortions and cosmic density and velocity fields in the EAGLE simulation”. In: *Monthly Notices of the Royal Astronomical Society: Letters* 461.1 (Apr. 2016), pp. L11–L15. ISSN: 1745-3933.
- [311] Joseph Kuruvilla, Nabila Aghanim, and Ian G. McCarthy. “Imprint of baryons and massive neutrinos on velocity statistics”. In: *Astronomy and Astrophysics* 644 (Dec. 2020), A170. ISSN: 1432-0746.
- [312] Abbé M. Whitford, Cullan Howlett, and Tamara M. Davis. “Evaluating bulk flow estimators for CosmicFlows-4 measurements”. In: *Monthly Notices of the Royal Astronomical Society* 526.2 (Dec. 2023), pp. 3051–3071. arXiv: 2306.11269 [astro-ph.CO].
- [313] Yehuda Hoffman et al. “The large-scale velocity field from the Cosmicflows-4 data”. In: *Monthly Notices of the Royal Astronomical Society* 527.2 (Jan. 2024), pp. 3788–3805. arXiv: 2311.01340 [astro-ph.CO].
- [314] Maria Lopes et al. “Bulk Flow Motion Detection in the Local Universe with Pantheon+ Type Ia Supernovae”. In: *Astrophysical Journal* 967.1, 47 (May 2024), p. 47. arXiv: 2405.11077 [astro-ph.CO].
- [315] Virginia Trimble. “Existence and Nature of Dark Matter in the Universe”. In: *Annual Review of Astronomy and Astrophysics* 25.1 (Sept. 1987), pp. 425–472.

- [316] D. N. Spergel et al. “First-Year Wilkinson Microwave Anisotropy Probe Observations: Determination of Cosmological Parameters”. In: *The Astrophysical Journal Supplement Series* 148.1 (Sept. 2003), pp. 175–194.
- [317] F. Hoyle. “On the Fragmentation of Gas Clouds Into Galaxies and Stars.” In: *The Astrophysical Journal* 118 (Nov. 1953), p. 513.
- [318] M. J. Rees and J. P. Ostriker. “Cooling, dynamics and fragmentation of massive gas clouds: clues to the masses and radii of galaxies and clusters”. In: *Monthly Notices of the Royal Astronomical Society* 179.4 (Aug. 1977), pp. 541–559.
- [319] J. Silk. “On the fragmentation of cosmic gas clouds. I - The formation of galaxies and the first generation of stars”. In: *The Astrophysical Journal* 211 (Feb. 1977), p. 638.
- [320] J. Binney. “The physics of dissipational galaxy formation”. In: *The Astrophysical Journal* 215 (July 1977), p. 483.
- [321] P. J. E. Peebles. “The Effect of a Lumpy Matter Distribution on the Growth of Irregularities in an Expanding Universe”. In: *Astronomy & Astrophysics* 32 (June 1974), p. 391.
- [322] Martin Crocce and Roman Scoccimarro. “Renormalized cosmological perturbation theory”. In: *Physical Review D* 73.6 (Mar. 2006).
- [323] Francisco-Shu Kitaura and Steffen Heb. “Cosmological structure formation with augmented Lagrangian perturbation theory”. In: *Monthly Notices of the Royal Astronomical Society: Letters* 435.1 (Aug. 2013), pp. L78–L82.
- [324] C. Pixius, S. Celik, and M. Bartelmann. “Kinetic field theory: perturbation theory beyond first order”. In: *Journal of Cosmology and Astroparticle Physics* 2022.12 (Dec. 2022), p. 030.
- [325] Atsuhisa Ota et al. “Lagrangian displacement field estimators in cosmology”. In: *Physical Review D* 107.12 (June 2023). ISSN: 2470-0029.
- [326] Mathias Garny, Dominik Laxhuber, and Román Scoccimarro. “Perturbation theory with dispersion and higher cumulants: Nonlinear regime”. In: *Physical Review D* 107.6 (Mar. 2023).
- [327] Mathias Garny, Dominik Laxhuber, and Román Scoccimarro. “Perturbation theory with dispersion and higher cumulants: Framework and linear theory”. In: *Physical Review D* 107.6 (Mar. 2023).
- [328] Cornelius Rampf and Oliver Hahn. “Renormalization group and UV completion of cosmological perturbations: Gravitational collapse as a critical phenomenon”. In: *Physical Review D* 107.2 (Jan. 2023).
- [329] B. Jain, H. J. Mo, and S. D. M. White. “The evolution of correlation functions and power spectra in gravitational clustering”. In: *Monthly Notices of the Royal Astronomical Society* 276.1 (Sept. 1995), pp. L25–L29. arXiv: astro-ph/9501047 [astro-ph].

- [330] Nissim Kanekar. “A Critique of Scaling Behavior in Nonlinear Structure Formation Scenarios”. In: *The Astrophysical Journal* 531.1 (Mar. 2000), pp. 17–21.
- [331] Chung-Pei Ma. “Analytical Approximation to the Nonlinear Power Spectrum of Gravitational Clustering”. In: *The Astrophysical Journal* 508.1 (Nov. 1998), pp. L5–L8.
- [332] J. A. Peacock and S. J. Dodds. “Non-linear evolution of cosmological power spectra”. In: *Monthly Notices of the Royal Astronomical Society* 280.3 (June 1996), pp. L19–L26. ISSN: 1365-2966.
- [333] John Joseph M. Carrasco, Mark P. Hertzberg, and Leonardo Senatore. “The effective field theory of cosmological large scale structures”. In: *Journal of High Energy Physics* 2012, 82 (Sept. 2012), p. 82. arXiv: 1206.2926 [astro-ph.CO].
- [334] Daniel Baumann et al. “Cosmological non-linearities as an effective fluid”. In: *Journal of Cosmology and Astroparticle Physics* 2012.07 (July 2012), pp. 051–051.
- [335] J. S. Bagla. “Cosmological N-Body simulation: Techniques, Scope and Status”. In: *Current Science* (2004).
- [336] James M. Gelb and Edmund Bertschinger. “Cold dark matter. 1: The formation of dark halos”. In: *The Astrophysical Journal* 436 (Dec. 1994), p. 467.
- [337] James M. Gelb and Edmund Bertschinger. “Cold dark matter. 2: Spatial and velocity statistics”. In: *The Astrophysical Journal* 436 (Dec. 1994), p. 491.
- [338] Giuseppe Tormen and Edmund Bertschinger. “Adding Long-Wavelength Modes to an N-Body Simulation”. In: *The Astrophysical Journal* 472.1 (Nov. 1996), pp. 14–24.
- [339] S. Cole. “Adding long-wavelength power to N-body simulations”. In: *Monthly Notices of the Royal Astronomical Society* 286.1 (Mar. 1997), pp. 38–47.
- [340] C. Power and A. Knebe. “The impact of box size on the properties of dark matter haloes in cosmological simulations”. In: *Monthly Notices of the Royal Astronomical Society* 370.2 (Aug. 2006), pp. 691–701.
- [341] Ryuichi Takahashi et al. “Simulations of baryon acoustic oscillations - I. Growth of large-scale density fluctuations”. In: *Monthly Notices of the Royal Astronomical Society* 389.4 (Sept. 2008), pp. 1675–1682.
- [342] Raul E. Angulo and Andrew Pontzen. “Cosmological N-body simulations with suppressed variance”. In: *Monthly Notices of the Royal Astronomical Society: Letters* 462.1 (May 2016), pp. L1–L5.
- [343] P. J. E. Peebles. “Renormalization group computation of the mass distribution in an expanding universe. I - Method”. In: *The Astrophysical Journal* 297 (Oct. 1985), p. 350.
- [344] Blane Little, David H. Weinberg, and Changbom Park. “Primordial fluctuations and non-linear structure”. In: *Monthly Notices of the Royal Astronomical Society* 253.2 (Nov. 1991), pp. 295–306.

- [345] H. M. P. Couchman and P. J. E. Peebles. “A Numerical Renormalization Solution for Self-similar Cosmic Structure Formation”. In: *The Astrophysical Journal* 497.2 (Apr. 1998), pp. 499–511.
- [346] J. S. Bagla, Jayanti Prasad, and Suryadeep Ray. “Gravitational collapse in an expanding background and the role of substructure - I. Planar collapse”. In: *Monthly Notices of the Royal Astronomical Society* 360.1 (June 2005), pp. 194–202.
- [347] D. Lynden-Bell. “Statistical Mechanics of Violent Relaxation in Stellar Systems”. In: *Monthly Notices of the Royal Astronomical Society* 136.1 (May 1967), pp. 101–121.
- [348] Martin D. Weinberg. “Noise-driven evolution in stellar systems - I. Theory”. In: *Monthly Notices of the Royal Astronomical Society* 328.1 (Nov. 2001), pp. 311–320.
- [349] P. J. E. Peebles. “Previrialization”. In: *The Astrophysical Journal* 365 (Dec. 1990), p. 27.
- [350] Kandaswamy Subramanian. “Self-Similar Collapse and the Structure of Dark Matter Halos: A Fluid Approach”. In: *The Astrophysical Journal* 538.2 (Aug. 2000), pp. 517–527.
- [351] Chung-Pei Ma and Michael Boylan-Kolchin. “Are Halos of Collisionless Cold Dark Matter Collisionless?” In: *Physical Review Letters* 93.2 (July 2004).
- [352] Chung-Pei Ma and Edmund Bertschinger. “A Cosmological Kinetic Theory for the Evolution of Cold Dark Matter Halos with Substructure: Quasi-Linear Theory”. In: *The Astrophysical Journal* 612.1 (Sept. 2004), pp. 28–49. arXiv: astro-ph/0311049 [astro-ph].
- [353] Anatoly A. Klypin and Adrian L. Melott. “Filamentary and hierarchical pictures - Kinetic energy criterion”. In: *The Astrophysical Journal* 399 (Nov. 1992), p. 397.
- [354] YA. B. Zel'dovich. “Survey of Modern Cosmology”. In: *Advances in Astronomy and Astrophysics*. Elsevier, 1965, pp. 241–379.
- [355] Pierluigi Monaco, Tom Theuns, and Giuliano Taffoni. “The pinocchio algorithm: pinpointing orbit-crossing collapsed hierarchical objects in a linear density field”. In: *Monthly Notices of the Royal Astronomical Society* 331.3 (Apr. 2002), pp. 587–608.
- [356] Pierluigi Monaco et al. “Predicting the Number, Spatial Distribution, and Merging History of Dark Matter Halos”. In: *The Astrophysical Journal* 564.1 (Jan. 2002), pp. 8–14.
- [357] T. Baertschiger et al. “Gravitational dynamics of an infinite shuffled lattice of particles”. In: *Physical Review E* 75.2 (Feb. 2007).
- [358] T. Baertschiger et al. “Erratum: Gravitational dynamics of an infinite shuffled lattice of particles”. In: *Physical Review E* 75.5 (May 2007).
- [359] T. Baertschiger et al. “Gravitational dynamics of an infinite shuffled lattice: Particle coarse-graining, nonlinear clustering, and the continuum limit”. In: *Physical Review E* 76.1 (July 2007).

- [360] Andrea Gabrielli et al. “Force distribution in a randomly perturbed lattice of identical particles with $1/r^2$ pair interaction”. In: *Physical Review E* 74.2 (Aug. 2006).
- [361] M. Joyce and B. Marcos. “Quantification of discreteness effects in cosmological N-body simulations. II. Evolution up to shell crossing”. In: *Physical Review D* 76.10 (Nov. 2007).
- [362] M. Joyce and B. Marcos. “Quantification of discreteness effects in cosmological N-body simulations: Initial conditions”. In: *Physical Review D* 75.6 (Mar. 2007).
- [363] M. Joyce et al. “Gravitational Evolution of a Perturbed Lattice and its Fluid Limit”. In: *Physical Review Letters* 95.1 (June 2005).
- [364] B. Marcos et al. “Linear perturbative theory of the discrete cosmological N-body problem”. In: *Physical Review D* 73.10 (May 2006).
- [365] Benjamin L’Huillier, Changbom Park, and Juhan Kim. “Effects of the initial conditions on cosmological N-body simulations”. In: *New Astronomy* 30 (July 2014), pp. 79–88.
- [366] Tianchi Zhang et al. “Numerical convergence of pre-initial conditions on dark matter halo properties”. In: *Monthly Notices of the Royal Astronomical Society* 507.4 (Sept. 2021), pp. 6161–6176.
- [367] James Binney and Alexander Knebe. “Two-body relaxation in cosmological simulations”. In: *Monthly Notices of the Royal Astronomical Society* 333.2 (June 2002), pp. 378–382.
- [368] Juerg Diemand et al. “Two-body relaxation in cold dark matter simulations”. In: *Monthly Notices of the Royal Astronomical Society* 348.3 (Mar. 2004), pp. 977–986.
- [369] James Binney. “Discreteness effects in cosmological N-body simulations”. In: *Monthly Notices of the Royal Astronomical Society* 350.3 (May 2004), pp. 939–948.
- [370] Amr A. El-Zant. “Two-body relaxation in simulated cosmological haloes”. In: *Monthly Notices of the Royal Astronomical Society* 370.3 (June 2006), pp. 1247–1256.
- [371] Alessandro B. Romeo et al. “Discreteness Effects in LCDM Simulations: A Wavelet-Statistical View”. In: *The Astrophysical Journal* 686.1 (Oct. 2008), pp. 1–12.
- [372] Philip Mansfield and Camille Avestruz. “How biased are halo properties in cosmological simulations?” In: *Monthly Notices of the Royal Astronomical Society* 500.3 (Nov. 2020), pp. 3309–3328.
- [373] J. Thiebaut et al. “On the onset of stochasticity in Lambda cold dark matter cosmological simulations”. In: *Monthly Notices of the Royal Astronomical Society* 387.1 (June 2008), pp. 397–406.
- [374] Takahiro Nishimichi, Francis Bernardeau, and Atsushi Taruya. “Response function of the large-scale structure of the universe to the small scale inhomogeneities”. In: *Physics Letters B* 762 (Nov. 2016), pp. 247–252.

- [375] F. R. Bouchet, J. -C. Adam, and R. Pellat. “On the way of understanding particle-in-cell simulations of gravitational clustering”. In: *Astronomy and Astrophysics* 144.2 (Mar. 1985), pp. 413–426.
- [376] R. Scoccimarro. “Transients from initial conditions: a perturbative analysis”. In: *Monthly Notices of the Royal Astronomical Society* 299.4 (Oct. 1998), pp. 1097–1118. ISSN: 1365-2966.
- [377] Somnath Bharadwaj. “The Evolution of Correlation Functions in the Zeldovich Approximation and Its Implications for the Validity of Perturbation Theory”. In: *Astrophysical Journal* 472 (Nov. 1996), p. 1. arXiv: astro-ph/9606121 [astro-ph].
- [378] C. Angrick and M. Bartelmann. “Triaxial collapse and virialisation of dark-matter haloes”. In: *Astronomy & Astrophysics* 518, A38 (July 2010), A38. arXiv: 1001 . 4984 [astro-ph.CO].
- [379] Ivo Labbé et al. “A population of red candidate massive galaxies 600 Myr after the Big Bang”. In: *Nature* 616.7956 (Feb. 2023), pp. 266–269. ISSN: 1476-4687.
- [380] Steven L. Finkelstein et al. “A Long Time Ago in a Galaxy Far, Far Away: A Candidate $z \sim 12$ Galaxy in Early JWST CEERS Imaging”. In: *The Astrophysical Journal Letters* 940.2, L55 (Dec. 2022), p. L55. arXiv: 2207 . 12474 [astro-ph.GA].
- [381] Rohan P. Naidu et al. *Schrodinger’s Galaxy Candidate: Puzzlingly Luminous at $z \approx 17$, or Dusty/Quenched at $z \approx 5$?* 2022. arXiv: 2208 . 02794 [astro-ph.GA].
- [382] Nick Kaiser. “Clustering in real space and in redshift space”. In: *Monthly Notices of the Royal Astronomical Society* 227.1 (July 1987), pp. 1–21. ISSN: 1365-2966.
- [383] John A. Peacock et al. “A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey”. In: *Nature* 410.6825 (Mar. 2001), pp. 169–173. ISSN: 1476-4687.
- [384] E. Hawkins et al. “The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe”. In: *Monthly Notices of the Royal Astronomical Society* 346.1 (Nov. 2003), pp. 78–96. ISSN: 1365-2966.
- [385] L. Guzzo et al. “A test of the nature of cosmic acceleration using galaxy redshift distortions”. In: *Nature* 451.7178 (Jan. 2008), pp. 541–544. ISSN: 1476-4687.
- [386] Isabel M E Santos-Santos et al. “Satellite mass functions and the faint end of the galaxy mass–halo mass relation in LCDM”. In: *Monthly Notices of the Royal Astronomical Society* 515.3 (July 2022), pp. 3685–3697. ISSN: 1365-2966.
- [387] L. Gao et al. “The statistics of the subhalo abundance of dark matter haloes: LCDM subhaloes”. In: *Monthly Notices of the Royal Astronomical Society* 410.4 (Oct. 2010), pp. 2309–2314. ISSN: 0035-8711.

- [388] V. Springel et al. “The Aquarius Project: the subhaloes of galactic haloes”. In: *Monthly Notices of the Royal Astronomical Society* 391.4 (Dec. 2008), pp. 1685–1711. ISSN: 1365-2966.