

# Effects of Local Non-Gaussianity on Numerical Measurements of Halo Bias

M.Sc. Thesis

by

**Antoine Beauchamp**

Supervisor: Dr. Daniel Green

*Canadian Institute for Theoretical Astrophysics*

*Department of Physics*

*University of Toronto*

*60 St. George Street, Toronto, ON*

August 23rd, 2015

---

## **Abstract**

We study the effects of primordial non-Gaussianities of the local type on the large-scale biasing of dark matter halos at late epochs. In the near future, state of the art surveys of large-scale structure promise to be an exciting new means of constraining the degree of non-Gaussianity in the primordial spectrum. An important feature to be measured by such surveys is the introduction of scale-dependence in the biasing of dark matter halos due to primordial non-Gaussianities. In preparation for future observational data, we perform numerical simulations of structure formation to examine this effect. In particular we make measurements of the linear and scale-dependent bias of dark matter halos. We find that the model used to describe scale-dependent bias is basically correct, with the scale-dependent correction due to non-Gaussianities obeying a linear relationship with  $f_{NL}$ .

---

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Theoretical Considerations</b>	<b>3</b>
2.1	The FLRW Universe . . . . .	4
2.2	Cosmological Perturbation Theory . . . . .	8
2.3	The Biasing of Dark Matter Halos . . . . .	14
2.4	The Power Spectrum . . . . .	16
2.5	Scale-Dependent Bias . . . . .	18
<b>3</b>	<b>Numerical Analysis</b>	<b>22</b>
3.1	$N$ -Body Simulations and GADGET-2 . . . . .	22
3.2	Generating the Initial Conditions . . . . .	25
3.3	The ROCKSTAR Halo Finder . . . . .	26
3.4	Computing the Power Spectrum . . . . .	27
<b>4</b>	<b>Results and Discussion</b>	<b>31</b>
4.1	Measuring the Linear Bias . . . . .	31
4.2	Measuring the Scale-Dependent Bias . . . . .	33
<b>5</b>	<b>Conclusion</b>	<b>37</b>
	<b>Appendix A Raw Measurements of Scale-Dependent Bias</b>	<b>40</b>
	<b>References</b>	<b>41</b>

# 1 Introduction

Over the past century, the field of cosmology has evolved from a speculative and predominantly philosophical area of study to one of rigorous scientific research. The technological advances in observational instrumentation achieved during this period have resulted in an abundance of data about the universe as a whole, its history and evolution. In addition to this, a rigorous theoretical framework has been developed following the pioneering work of Friedmann, Lemaitre, Robertson and Walker (FLRW) in the first half of the 20th century, allowing us to make real quantitative predictions that can be tested observationally. The most important measurements in modern cosmology have been made using the cosmic microwave background (CMB) and the distributions of galaxies, dark matter halos, and clusters on large scales, commonly referred to as large-scale structure (LSS). These areas of study have provided ample evidence that the universe is mostly homogeneous and isotropic on the largest scales. As an example, the anisotropies in the temperature field of the CMB have a relative amplitude on the order of  $10^{-5}$  compared to the mean isotropic temperature [1]. This observational fact, and other measurements of the kind [1], suggest that the universe as a whole can reasonably be approximated as being homogeneous and isotropic. It turns out however that these minute deviations from homogeneity and isotropy open the door to the most interesting physics in modern cosmology. It could be argued that the primary goal of contemporary cosmology is to study these inhomogeneities and anisotropies in detail. Following the success of the FLRW model, theorists developed the theoretical framework of cosmological perturbation theory to quantify the deviations from isotropy and homogeneity [2]. Within this framework, the anisotropies and inhomogeneities are treated as perturbations to the FLRW universe. Refined measurements of the CMB and LSS performed in recent decades have allowed us to study in detail the physics contained in these perturbations.

The anisotropies in the CMB, the oldest light in the universe, suggest that these perturbations were generated very early in the history of the cosmos and were present in the earliest distributions of the different forms of matter and radiation. Furthermore, these early, or *primordial*, perturbations are believed to have been generated stochastically, with statistics that are constrained by the CMB to be very close to Gaussian. This is one of the most important features of the cosmological perturbations. The nature of these statistics carries important information about the earliest epochs in the universe. A large number of theoretical scenarios have been proposed to explain how such perturbations could have been generated without recourse to fine tuning, the most common being the theory of inflation. Such theoretical models make predictions about the statistics of the primordial perturbations. For instance, the simplest inflationary scenarios predict perturbations that obey Gaussian statistics. The statistical properties of the primordial perturbations thus provide indispensable information about the physical processes that took place in the very early universe.

It has become a point of interest in recent years to make rigorous measurements of the degree to which the statistics of the primordial perturbations deviate from being Gaussian. Such ‘non-Gaussianities’ would rule out a number of theoretical models used to describe the initial perturbations, including the simplest forms of inflation. A number of different types of non-Gaussianities have been proposed. The most important type pertaining to our research is known as *local* non-Gaussianity. This

is commonly expressed as [3][4][5][6][7][8][9]

$$\Phi_p = \phi_p + f_{NL} (\phi_p^2 - \langle \phi_p^2 \rangle) \quad (1.0.1)$$

where  $\Phi_p$  is the primordial gravitational potential and  $\phi_p$  is an underlying potential with Gaussian statistics. These variables are discussed in detail in subsequent sections. The quantity  $f_{NL}$ , often expressed as  $f_{NL}^{local}$ , is a measure of the deviation from Gaussianity. Measurements of the CMB have already been used to constrain the degree of local non-Gaussianity in the primordial perturbations [9][10][11][12]. The current best estimate [12] obtained from measurements of the CMB temperature is  $f_{NL} = 2.5 \pm 5.7$ . Combining temperature and polarization data results in  $f_{NL} = 0.8 \pm 5.0$ . The CMB however is not the only way to acquire constraints on primordial non-Gaussianities. It has been known for some time that the degree of primordial non-Gaussianity can be constrained by various measurements of LSS [3][10][4][5][6][7][8][9]. Moreover, further measurements of the CMB are not expected to result in more precise measurements of non-Gaussianity. This is not the case with LSS however, where significant improvements could be made towards constraining primordial non-Gaussianity [10]. The most promising measurement performed on LSS is the scale-dependent bias of dark matter halos, which arises due to non-Gaussianities of the local type [3]. The scale-dependent bias on large scales is already capable of providing constraints competitive with the CMB using data from currently available galaxy surveys [10]. A number of LSS surveys planned for the coming decade have the potential to improve these constraints significantly [10][13][14][15][16]. An example of an ideal survey for constraining local non-Gaussianities is outlined in [10].

In preparation for the availability of future LSS data, it is important to have a solid understanding of the physics involved in the effects that we hope to measure. It is additionally important to understand how the data from such surveys could be used to constrain the degree of non-Gaussianity in the primordial perturbations. To this end, we have endeavoured to make measurements of the scale-dependent bias of dark matter halos due to local non-Gaussianities using data from numerical simulations. The results of these measurements are presented in Section 4 of this report. The content of this report is organized into three broad sections. We begin, in Section 2, with a discussion of the theoretical framework used to describe modern cosmological observations and the physics leading to the scale-dependent bias. In Section 3 we discuss the importance of using numerical tools to supplement analytical considerations and we elaborate on the tools used to perform the simulations and analyze the output. Finally, in Section 4, we present the results of the simulations and the measurements of linear and scale-dependent bias.

## 2 Theoretical Considerations

We begin our discussion with a review of some of the most important theoretical aspects of modern cosmology, with the aim of ultimately developing the scale-dependent bias of dark matter halos. Section 2.1 develops the physics of the FLRW universe. This is improved upon in Section 2.2 with development of cosmological perturbation theory. In Section 2.3 we discuss some of the physics of dark matter halos and the local bias ansatz. This is followed by a discussion of the statistical nature of cosmological predictions and an introduction of the power spectrum in Section 2.4. Finally, in Section 2.5 we develop the scale-dependent bias of dark matter halos due to local primordial non-Gaussianities.

## 2.1 The FLRW Universe

A discussion of the physics of the largest scales in the universe can be approached using the formalism of general relativity. In general relativity, the universe is described as a curved spacetime, with the line element between points, or events, in the spacetime being given as

$$ds^2 = g_{\alpha\beta}(x^\mu)dx^\alpha dx^\beta \quad (2.1.1)$$

where  $g_{\alpha\beta}$  is the metric tensor for the spacetime and  $dx^\alpha$  is a coordinate differential. The geometric properties of the spacetime, most importantly its symmetries, are encoded in the metric tensor. As mentioned in the introduction, on the largest scales the universe appears to be mostly homogeneous and isotropic. Furthermore, the universe is known to be expanding [1][17][18][19][20]. These properties together suggest that the universe can be described by a spacetime that consists of a foliation of maximally symmetric, i.e. isotropic and homogeneous, spacelike hypersurfaces that evolve in time [18][17][19]. The class of spacetimes satisfying this property are known as the FLRW spacetimes. The most general line element for such a spacetime is

$$ds^2 = -dt^2 + a^2(t) \left[ \frac{dr^2}{1 - \kappa r^2} + r^2 d\Omega^2 \right] \quad (2.1.2)$$

where  $t$  is the time coordinate,  $r$  is a radial coordinate and  $d\Omega$  is the line element for a 2-sphere. The speed of light has been set to  $c = 1$ . The function  $a(t)$ , called the *scale factor*, is a measure of the ‘size’ of the spacelike hypersurface at a given value of the time coordinate, and its evolution describes the expansion (or contraction) of the universe. We choose to normalize scale factor so that it is equal to  $a_0 \equiv a(t_0) = 1$  at the present time. This choice corresponds to a dimensionless scale factor. The variable  $\kappa$  is the *curvature index*, which is a normalized measure of the three-dimensional Ricci scalar, describing the curvature of the spacelike hypersurfaces [18]. Our convention is such that  $\kappa$  has dimensions of  $[\kappa] = L^{-2}$ . Since the hypersurfaces are maximally symmetric, they are also surfaces for which the Ricci scalar takes on a constant value [18]. Therefore, apart from the absolute value of Ricci, the most important quality of the curvature scalar is its sign. This allows us to characterize the spacelike hypersurfaces, and consequently the universe, in terms of their curvature as being closed ( $\kappa > 0$ ), flat ( $\kappa = 0$ ), or open ( $\kappa < 0$ ) <sup>‡</sup>. We take the spacelike sections to be flat for the remainder of the discussion since observation suggests that this is an appropriate description of our universe [17][21]. Moreover, the algebra is simpler with the choice of  $\kappa = 0$  since the spacelike sections are simply three-dimensional Euclidean spaces in this case. Thus, setting the curvature to zero and moving to rectangular coordinates  $(t, x, y, z)$  the line element takes the form

$$ds^2 = -dt^2 + a^2(t)(dx^2 + dy^2 + dz^2) . \quad (2.1.3)$$

Coordinates such as the ones used here and in equation (2.1.2), in which the spacelike surfaces are orthogonal to the timelike direction (made manifest by the fact that  $g_{0i} = 0$ ), are called *comoving coordinates*. An observer whose local frame coincides with these coordinates will observe a homogeneous and isotropic universe. The FLRW metric is also commonly written in terms of the *conformal time*, denoted  $\eta$ , instead of the standard time  $t$ . The transformation between the conformal time and the

<sup>‡</sup> Some conventions take  $r$  and  $\kappa$  to be dimensionless so that  $\kappa = 0, \pm 1$ .

coordinate time can be obtained by considering null paths through the spacetime, for which  $ds^2 = 0$ . Then, identifying  $d\eta^2 \equiv dx^2 + dy^2 + dz^2$ , we have

$$d\eta = \frac{dt}{a(t)} . \quad (2.1.4)$$

Integrating this differential from  $t = 0$  to some later time  $t$ , we obtain an expression for the conformal time

$$\eta(t) = \int_0^t \frac{dt'}{a(t')} . \quad (2.1.5)$$

Using conformal time as the time coordinate has a number of advantages, one of which is that the metric in equation (2.1.3) can be expressed as a conformal transformation of the Minkowski spacetime:

$$ds^2 = a^2(\eta) (-d\eta^2 + dx^2 + dy^2 + dz^2) . \quad (2.1.6)$$

Additionally, the conformal time has the nice physical interpretation that it is the maximum distance on the comoving coordinate grid that light could have travelled since the beginning of time [17][19]. In this way it serves as a causal horizon, called the *comoving horizon*, since regions separated by a comoving distance greater than  $\eta$  at a given time could *never* have been in causal contact. The comoving horizon will grow as the universe evolves, so that causally disconnected regions will eventually become connected. An important caveat is that this interpretation of the conformal time as a causal horizon is only correct when  $\eta = 0$  corresponds to the beginning of time. This ceases to be true when considering inflationary theories. The conformal time will be used preferentially over the standard time when we discuss the evolution of perturbations in Section 2.2.

The FLRW spacetime provides a description of the geometric structure of the universe, but the description of the universe is not complete without a specification of the matter that occupies it. In general relativity the matter content in the spacetime is characterised by the energy-momentum tensor,  $T_{\alpha\beta}$ . In order to make progress, we require a specific form of the energy-momentum tensor. This is done by describing the matter on the largest scales in the universe approximately as a fluid [18][17][19]. The conditions of homogeneity and isotropy require that this fluid be perfect in the comoving frame, with the energy density taking on a constant value across the spatial sections, and the pressure being isotropic about every point. Formally, the fluid can be described in terms of a congruence of timelike geodesics with 4-velocity field  $u^\alpha = \frac{dx^\alpha}{d\lambda}$ , where  $\lambda$  is an appropriate parameter along the geodesics. With our choice of metric signature, the 4-velocities are normalized so that  $u^\alpha u_\alpha = -1$ . The energy-momentum tensor for the perfect fluid can then be written as

$$T_{\alpha\beta} = [\rho(t) + P(t)] u_\alpha u_\beta + P(t) g_{\alpha\beta} \quad (2.1.7)$$

where  $\rho(t)$  is the energy density and  $P(t)$  is the pressure, as measured in the comoving frame. Further, in the comoving frame the 4-velocity takes on the simple form  $u^\alpha = (1, 0, 0, 0)$ . The energy density and pressure here are the total energy density and pressure for the matter in the universe. Therefore these will include contributions from all types, or species, of matter in the universe, i.e.

$$\rho(t) = \sum_i \rho_i(t) \quad (2.1.8)$$

$$P(t) = \sum_i P_i(t) \quad (2.1.9)$$

where the subscript  $i$  denotes a particular species.

We now have a reasonable model for the universe as a whole, where the background spacetime is given by the FLRW spacetime and the matter content is described approximately as a perfect fluid. We can now consider how the two interact to generate the expansion of the universe. From general relativity, we know that the dynamics of the spacetime in response to the matter content are governed by the Einstein field equations

$$G_{\alpha\beta} \equiv R_{\alpha\beta} - \frac{1}{2}Rg_{\alpha\beta} = 8\pi GT_{\alpha\beta} \quad (2.1.10)$$

where  $R_{\alpha\beta}$  and  $R$  are, respectively, the Ricci tensor and scalar for the spacetime. Furthermore, we can obtain an equation describing the evolution of the matter in the spacetime by considering the conservation law for the energy-momentum tensor

$$T^\alpha{}_{\beta||\alpha} = 0 \quad (2.1.11)$$

where  $||_\alpha$  denotes covariant differentiation with respect to the  $\alpha$ th coordinate. To begin, we consider the Einstein equations. Since the metric tensor for the FLRW spacetime is characterized by a single dynamical variable, i.e.  $a(t)$ , we need only consider one of the Einstein equations, since the others will either be redundant or null. Considering only the time-time component of the field equations,  $G_{00} = 8\pi GT_{00}$ , and using the FLRW metric to compute the Ricci tensor and scalar, one obtains the *Friedmann equation*

$$H^2(t) = \left(\frac{da/dt}{a}\right)^2 = \frac{8\pi G}{3}\rho(t) \quad (2.1.12)$$

where we have used the fact that  $T_{00} = \rho(t)$  and defined the *Hubble rate*

$$H(t) \equiv \frac{da/dt}{a} . \quad (2.1.13)$$

In terms of conformal time, the Hubble rate takes the form

$$H(\eta) = \frac{\dot{a}}{a^2} \quad (2.1.14)$$

with  $\dot{\phantom{x}} \equiv \frac{d}{d\eta}$ . The Friedmann equation describes how the scale factor (and consequently the spacetime) evolves in the presence of matter. It is also commonly written as

$$\left(\frac{H(t)}{H_0}\right)^2 = \frac{\rho(t)}{\rho_{cr}} \quad (2.1.15)$$

where  $H_0$  is the value of the Hubble rate today and  $\rho_{cr}$  is the critical density necessary to generate the value of  $H_0$  in a flat universe. In other words,  $H_0$  and  $\rho_{cr}$  together form a solution to equation (2.1.12). The value of Hubble today and the critical density are [1][17][19]

$$H_0 = 100h \text{ km} \cdot \text{s}^{-1} \cdot \text{Mpc}^{-1} \quad (2.1.16)$$

$$\rho_{cr} = 1.88h^2 \times 10^{-29} \text{ g} \cdot \text{cm}^{-3} \quad (2.1.17)$$

where  $h$  is a dimensionless numerical parameterization, with a current measured value [21] of

$$h = 0.6774 \pm 0.0046 . \quad (2.1.18)$$

An equation for the dynamics of the fluid can be obtained by considering the time ( $\beta = 0$ ) component of the conservation equation, in equation (2.1.11). This leads to the cosmological generalization of the continuity equation:

$$\frac{\partial \rho(t)}{\partial t} + 3H(t) [\rho(t) + P(t)] = 0 . \quad (2.1.19)$$

The Friedmann equation and the continuity equation together are enough to describe how the universe evolves. However, in order to solve these equations one requires an equation of state for the species at hand. For a given species  $i$ , this can be written as

$$P_i = w_i \rho_i \quad (2.1.20)$$

where  $w_i$  is some expression that can be defined in general using  $w \equiv P/\rho$ . The components of the universe are usually described in terms of matter, radiation, and dark energy. The term matter is used to describe non-relativistic species such as baryonic matter and cold dark matter, while radiation is used to describe relativistic species such as photons and neutrinos. Each of these species has an equation of state of the form in equation (2.1.20) with constant values of  $w_i$ . Since matter is non-relativistic, the pressure exerted by these species is negligible compared to their density. It is thus approximated as being null, resulting in  $w_m = 0$ . The equation of state for radiation takes on the standard form, with  $w_r = 1/3$ . Finally, dark energy is usually given an equation of state with  $w_\Lambda = -1$ , so that its density remains constant in time. If we restrict ourselves to equations of state for which  $w_i$  is a constant, such as the ones for these species, then equation (2.1.19) can be solved readily to yield

$$\rho_i(a) = \rho_{i,0} a^{-3(1+w_i)} . \quad (2.1.21)$$

where  $\rho_{i,0}$  is the density of the species today. In general the energy density and pressure will include contributions from all species, as in equations (2.1.8) and (2.1.9). It is not possible to solve the Friedmann equation for the time evolution of the scale factor if we consider all species. However, from equation (2.1.21), we note that the constituents of the universe scale differently with  $a(t)$  due to their respective equations of state. We can thus consider a particular species to be dominating the expansion at a given epoch in history. This allows us to approximate the universe as a single-component universe for the era in question. The history of the universe is thus organized into three broad eras based on the dominant form of energy. These are the radiation era, the matter era, and the dark energy era. The fact that  $\rho_r \propto a^{-4}$  means that radiation will dominate in the early universe, until it is replaced by matter, which scales as  $\rho_m \propto a^{-3}$ . The period when this shift occurs is known as the period of matter-radiation equality, or the matter-radiation transition. At very late times, when the densities of matter and radiation have



decayed, the universe will be dominated by dark energy. For a single-component universe, the Friedmann equation can be solved using the expression for  $\rho_i(a)$  to obtain the time dependence of the scale factor:

$$a(t) \propto \begin{cases} t^{\frac{2}{3(1+w_i)}} & , \quad w_i \neq -1 \\ e^{H(t)t} & , \quad w_i = -1 \end{cases} . \quad (2.1.22)$$

In terms of conformal time,

$$a(\eta) \propto \begin{cases} \eta^{\frac{2}{1+3w_i}} & , \quad w_i \neq -1 \\ -\frac{1}{\eta} & , \quad w_i = -1 \end{cases} . \quad (2.1.23)$$

The different ways in which the densities of species scale with  $a$  lead to different behaviours of the scale factor. The abundance of a given species is often described using the dimensionless parameter,  $\Omega_i$ , defined as

$$\Omega_i \equiv \frac{\rho_{i,0}}{\rho_{cr}} . \quad (2.1.24)$$

If the spatial sections of the universe are flat, then this is simply the fraction of the total density that a given species occupies at present. The current best estimates of these quantities are [21]

$$\Omega_m = 0.3089 \pm 0.0062 \quad (2.1.25)$$

$$\Omega_\Lambda = 0.6911 \pm 0.0062 . \quad (2.1.26)$$

Due to the negligible contribution from radiation at the present epoch, the abundance of radiation is usually expressed via the temperature of the CMB,

$$T_{CMB} = 2.718 \pm 0.021 \text{ K} . \quad (2.1.27)$$

The FLRW description of cosmology, in which the universe is treated as being perfectly spatially homogeneous and isotropic but evolving in time, is a reasonable first approximation to the physics of the largest observable scales. However, as mentioned in the introduction, the universe is not perfectly homogeneous and isotropic. The deviations from these symmetries provide an important window into the physics of large scales and the early universe. We now turn our attention to the study of these cosmological perturbations.

## 2.2 Cosmological Perturbation Theory

There is ample evidence of deviations from homogeneity and isotropy in both the CMB and LSS [1][17][21]. The theory of perturbations that we develop in this section can be used to study both of these regimes in extensive detail [17][19]. However, since the goal of this research project is to improve our understanding of the scale-dependent biasing of halos, we develop the theory of cosmological perturbations exclusively within the framework of LSS. The formation of LSS occurs predominantly in the matter era, and is driven primarily by the interaction between cold dark matter and gravity. We can therefore ignore radiation and baryonic matter in our treatment of the perturbations [1][17][19][22]. For the remainder of the report, we use the term ‘matter’ interchangeably with ‘cold dark matter’.

We begin our discussion of cosmological perturbations by noting that the absence of perfect symmetry is necessary in order for structure to form in the universe. This can be understood loosely in the following way: One can imagine that regions in space that are initially slightly more dense than the mean density of the FLRW cosmology will undergo a slightly slower expansion due to the effects local gravitation. Over time these initially overdense regions will condense to form the gravitationally bound systems that make up the LSS. Such a process would not be possible if the density was identically equal to the mean density at all points in space. This fact necessitates a reconsideration of the FLRW cosmologies. In its complete formulation, this is a complex problem involving a large number of dynamical variables characterizing the perturbations to the matter and radiation fields, as well as to the spacetime. Fortunately the deviations from homogeneity and isotropy on large scales are small enough to be treated perturbatively. This allows us to expand the dynamical equations around the zeroth order solution of the FLRW universe, and to truncate the perturbation theory at a suitable order. The standard procedure, and the one that we follow here, is to truncate at linear order in the variables characterizing the perturbations, though significant advances have been made recently in extending cosmological perturbation theory to higher orders [22][23][24][25]. The absence of radiation within the context of LSS results in further simplification, allowing us to restrict our attention to the variables used to characterize the perturbations to the matter and the gravitational field. In this section we set up the framework used to describe these perturbations, the resulting dynamical equations, and their solutions.

As we did in Section 2.1, the first thing to consider when developing the theory of cosmological perturbations is the spacetime. The existence of gravitational and density perturbations in the universe breaks the assumptions of perfect homogeneity and isotropy that are represented by the FLRW spacetime. The FLRW metric in equation (2.1.3) must therefore be modified to account for local variations in the gravitational field. In general, we can write the metric for a perturbed spacetime as

$$g_{\alpha\beta} = \bar{g}_{\alpha\beta} + \delta g_{\alpha\beta} \quad (2.2.1)$$

where  $\bar{g}_{\alpha\beta}$  is the zeroth order FLRW metric and  $\delta g_{\alpha\beta}$  is the linear perturbation. In the ensuing discussion we will denote the zeroth order, or mean, quantities of the FLRW spacetime using an overbar. The specific form of the perturbation  $\delta g_{\alpha\beta}$  can be expressed in a number of different ways. In particular one can decompose the metric perturbation into scalar, vector, and tensor perturbations, where the designation is based on how the perturbations transform with respect to a transformation of the *spatial* coordinates. Remarkably, the three types of perturbations decouple at linear order and evolve independently [17][19]. This allows us to focus on a single class of perturbations. In this case we consider only scalar perturbations, since these play the most important role in the formation of LSS [17]. We can write the metric with scalar perturbations as

$$g_{00}(\vec{x}, t) = -1 - 2A(\vec{x}, t) \quad (2.2.2)$$

$$g_{0i}(\vec{x}, t) = -a(t)B(\vec{x}, t)|_i \quad (2.2.3)$$

$$g_{ij}(\vec{x}, t) = a(t)^2 \left\{ \delta_{ij} [1 + 2C(\vec{x}, t)] - 2E(\vec{x}, t)|_{ij} \right\} \quad (2.2.4)$$

where  $A, B, C$  and  $E$  are the scalar perturbation variables and  $|_i$  denotes partial differentiation with

respect to the  $i$ th coordinate. Though we have written the perturbation in terms of these variables, they are not uniquely defined. This is due to our freedom in how we define the coordinates on the spacelike hypersurfaces in the spacetime, and how we choose to foliate the spacetime with these surfaces. We can exploit this coordinate, or gauge, freedom to set two of the perturbation variables to zero [17][19]. Of particular usefulness to the theory of LSS is the choice of *conformal Newtonian gauge*, for which  $B = E = 0$ , and for which we define  $A \equiv \Psi$  and  $C \equiv \Phi$ . The metric then takes the form

$$g_{00}(\vec{x}, t) = -1 - 2\Psi(\vec{x}, t) \quad (2.2.5)$$

$$g_{0i}(\vec{x}, t) = 0 \quad (2.2.6)$$

$$g_{ij}(\vec{x}, t) = a^2(t)\delta_{ij} [1 + 2\Phi(\vec{x}, t)] \quad (2.2.7)$$

This choice of gauge allows us to identify  $\Psi$  as the classical Newtonian gravitational potential. The variable  $\Phi$ , on the other hand, is the perturbation to the spatial curvature of the spacetime, in the sense that it causes the three-dimensional Ricci scalar on the hypersurfaces to deviate from its constant value.

Having set up the metric for the perturbed spacetime, we can now turn our attention to the matter perturbations. We can describe these perturbations by first decomposing the energy-momentum tensor into zeroth and first order parts:

$$T_{\alpha\beta} = \bar{T}_{\alpha\beta} + \delta T_{\alpha\beta} \quad (2.2.8)$$

Implicit in this description is the addition of perturbations to the energy density, the pressure, and the fluid 4-velocity:

$$\rho(\vec{x}, t) = \bar{\rho}(t) [1 + \delta(\vec{x}, t)] \quad (2.2.9)$$

$$P(\vec{x}, t) = \bar{P}(t) + \delta P(\vec{x}, t) \quad (2.2.10)$$

$$u^\alpha(\vec{x}, t) = \bar{u}^\alpha + \delta u^\alpha(\vec{x}, t) \quad (2.2.11)$$

where  $\delta$  is called the *fractional overdensity* or *density contrast*, describing the deviation of the density field from its mean value, and  $\delta P$  and  $\delta u^\alpha$  are the perturbations to the pressure and 4-velocity. In addition to these, the perturbed energy-momentum tensor could include a contribution from anisotropic stresses, but we choose to neglect this. At linear order, the contributions to anisotropic stress come from the quadrupoles of the radiation perturbations. In the matter era these quantities are negligible [17]. For the reasons discussed at the beginning of this section, we choose to describe the matter entirely in terms of cold dark matter, for which  $\bar{P}(t) = 0$ . In general the perturbations of a given species could induce perturbations in the pressure that cause deviations from the equation of state used in the FLRW universe. However, since the pressureless nature of cold dark matter in the FLRW universe is due to the fact that it is collisionless, it is reasonable to assume that this property is not influenced by the presence of perturbations. We can therefore set  $P(\vec{x}, t) = 0$ . The perturbations to the dark matter are then characterized in terms of the overdensity  $\delta$  and velocity perturbation  $\delta u^\alpha$ . Recalling the definition of the energy-momentum tensor of a perfect fluid from equation (2.1.7), we can write the perturbed

energy-momentum tensor for the matter as follows:

$$T_{\alpha\beta} = [\bar{\rho}(t) + \bar{\rho}(t)\delta(\vec{x}, t)] [\bar{u}_\alpha + \delta u_\alpha(\vec{x}, t)] [\bar{u}_\beta + \delta u_\beta(\vec{x}, t)] . \quad (2.2.12)$$

Next, expanding and keeping terms only to first order in the perturbation variables, we identify the zeroth and first order terms as

$$\bar{T}_{\alpha\beta} = \bar{\rho}(t) \bar{u}_\alpha \bar{u}_\beta \quad (2.2.13)$$

$$\delta T_{\alpha\beta} = \bar{\rho}(t) [\bar{u}_\alpha \delta u_\beta(\vec{x}, t) + \bar{u}_\beta \delta u_\alpha(\vec{x}, t) + \bar{u}_\alpha \bar{u}_\beta \delta(\vec{x}, t)] . \quad (2.2.14)$$

We can make this more precise by further determining the components of  $\delta u_\alpha$ . Using the following conditions

$$g_{\alpha\beta} u^\alpha u^\beta = -1 \quad (2.2.15)$$

$$\bar{g}_{\alpha\beta} \bar{u}^\alpha \bar{u}^\beta = -1 \quad (2.2.16)$$

$$\bar{u}^\alpha = \delta_0^\alpha \quad (2.2.17)$$

and

$$\bar{g}_{0i} = \delta g_{0i} = 0 , \quad (2.2.18)$$

one finds

$$\delta u^\alpha(\vec{x}, t) = (-\Psi(\vec{x}, t) , v^i(\vec{x}, t)) . \quad (2.2.19)$$

Here  $v^i$  is the *peculiar velocity* of the fluid, defined to be the spatial part of the perturbed 4-velocity. Since  $\Psi$  is a perturbation variable for the metric, we conclude that the perturbations to the cold dark matter can be described in terms of the density contrast  $\delta$  and the peculiar velocity  $v^i$ .

With these definitions of the perturbation variables for the matter and the metric tensor, we are now in a position to compute the equations that govern their evolution. This can be accomplished by calculating the first-order corrections to the Einstein equations and to the conservation equation for the energy-momentum tensor, written down in equations (2.1.10) and (2.1.11). Though we will not go through the details, this is a fairly algorithmic process that entails the computation of the first-order corrections to various geometric quantities, such as the Christoffel symbols, the Ricci tensor, and the Ricci scalar. In performing this calculation, we drop all terms higher than first-order in the perturbative variables  $\Psi$ ,  $\Phi$ ,  $\delta$ , and  $v^i$ . As expected, the zeroth order equations generated by this process reproduce the Friedmann equation and the FLRW continuity equation from equations (2.1.12) and (2.1.19). The first-order equations obtained from the conservation equation in (2.1.11) are as follows:

$$\frac{\partial \delta(\vec{x}, t)}{\partial t} + \frac{1}{a} \frac{\partial v^i(\vec{x}, t)}{\partial x^i} + 3 \frac{\partial \Phi(\vec{x}, t)}{\partial t} = 0 \quad (2.2.20)$$

and

$$\frac{\partial v^i(\vec{x}, t)}{\partial t} + \frac{da/dt}{a} v^i(\vec{x}, t) + \frac{1}{a} \frac{\partial \Psi(\vec{x}, t)}{\partial x^i} = 0 . \quad (2.2.21)$$

These equations describe the dynamics and evolution of the matter perturbations. By taking the Fourier transform of these equations, noting that  $\frac{\partial}{\partial x^i} \rightarrow ik_i$ , and using conformal time as the independent variable, these equations can be written as

$$\dot{\delta}(\vec{k}, \eta) + ikv(\vec{k}, \eta) + 3\dot{\Phi}(\vec{k}, \eta) = 0 \quad (2.2.22)$$

$$\dot{v}(\vec{k}, \eta) + \frac{\dot{a}}{a}v(\vec{k}, \eta) + ik\Psi(\vec{k}, \eta) = 0 \quad (2.2.23)$$

where  $k \equiv |\vec{k}|$ . We've additionally assumed that the velocity field is irrotational and can thus be described as the gradient of a scalar function,  $v^i = \frac{k^i}{k}v$  [17]. The advantage of transforming to Fourier space is that we've converted the set of partial differential equations into ordinary differential equations, which are a lot easier to solve. Additionally, equations (2.2.22) and (2.2.23) are decoupled in the Fourier modes, i.e.  $\delta(\vec{k})$ , allowing us to solve for the evolution of one mode without knowing anything about the others. Finally, expressing the quantities as Fourier modes allows us to discuss the behaviour of the perturbations on different length scales, which are characterized by  $k^{-1}$ . It is worth noting that one cannot describe the perturbations in the distributions of all species using fluid equations of this type. For example, one cannot describe the behaviour of photons in the matter era using such equations. The case of cold dark matter is special however, since these particles do not travel very far on cosmological time scales. In general, the equation for the velocity, such as the one in equation (2.2.23), is not guaranteed to close. It does for dark matter since the velocity perturbations are small. The fact that the peculiar velocity of dark matter is small means that the perturbations can be treated using a fluid approximation where the velocity of the fluid is given the Hubble expansion.

Moving on to the perturbations in the metric, the equations governing the evolution of  $\Phi$  and  $\Psi$  can be obtained by computing the first-order corrections to the Einstein equations. The resulting equations in Fourier space are

$$k^2\Phi(\vec{k}, \eta) + 3\frac{\dot{a}}{a}\left[\dot{\Phi}(\vec{k}, \eta) - \Psi(\vec{k}, \eta)\frac{\dot{a}}{a}\right] = 4\pi Ga^2\bar{\rho}(\eta)\delta(\vec{k}, \eta) \quad (2.2.24)$$

and

$$k^2\left(\Phi(\vec{k}, \eta) + \Psi(\vec{k}, \eta)\right) = 0. \quad (2.2.25)$$

In general the right-hand side of equation (2.2.25) receives a contribution from the anisotropic stress, but we have neglected this contribution, as mentioned earlier in this section. The absence of anisotropic stress allows us to characterize the scalar perturbations using a single variable, by writing  $\Psi = -\Phi$ . In addition to these two equations, a useful algebraic equation can be obtained from the Einstein equations:

$$k^2\Phi(\vec{k}, \eta) = 4\pi Ga^2\left[\bar{\rho}(\eta)\delta(\vec{k}, \eta) + i\frac{3aH(\eta)}{k}\bar{\rho}(\eta)v(\vec{k}, \eta)\right]. \quad (2.2.26)$$

In the large- $k$  limit, this equation reduces to the cosmological Poisson's equation in Fourier space.

Having developed equations (2.2.22) and (2.2.23) for the matter perturbations and equations (2.2.24) to (2.2.26) for the metric perturbations, we would like to solve these equations for the time evolution of  $\delta$  and  $\Phi$ . Without going into detail, it can be shown that the evolution of the overdensity and the

potential in the matter era is mode-independent, i.e. it depends only on the time coordinate [1][17][19]. In particular the potential evolves as

$$\Phi(\vec{k}, a) = \frac{D(a)}{a} \Phi(\vec{k}, a_i) \quad (2.2.27)$$

where we have used the scale factor as the time coordinate. Moreover,  $D(a)$  is the *growth function* and  $a_i$  is some time after the end of the matter-radiation transition, but suitably early in the matter era. It can be shown that  $D(a) = a$  in matter domination [1][17][19], so the potential remains constant in this epoch. Though this provides a description of how the potential evolves in the matter era, we have so far neglected the effects of the radiation era. In fact, the epoch of radiation domination and the matter-radiation transition have a significant effect on the potential. The most important effect is that the amplitude of all modes undergoes some decay from their primordial value during the transition from radiation to matter domination. However this effect is significantly more pronounced for the small-scale modes, which decay considerably. In comparison, the large-scale modes only decrease in amplitude by about 10% compared to their primordial value. This leads to a suppression of the potential on small scales in the matter era, where the potential is frozen on all scales.

These effects can be examined rigorously by considering the equations that govern the evolution of the potential on various scales in the radiation and matter eras. However the mode-dependent behaviour of the potential during the matter-radiation transition can be captured in terms of a transfer function,  $T(k)$ , which relates the amplitude of a given mode early in the matter era, i.e. before the effects of the growth function have taken place, to its primordial value. We can write this as

$$\Phi(\vec{k}, a_i) = \frac{9}{10} T(k) \Phi_p(\vec{k}) \quad (2.2.28)$$

where  $\Phi_p(\vec{k})$  is the primordial potential describing the initial conditions of the perturbations in the universe. The transfer function is normalized so that  $T(k) = 1$  on large scales. The prefactor of  $\frac{9}{10}$  thus describes the marginal decrease in amplitude of the large-scale modes. In practice the functional form of  $T(k)$  is obtained by solving the evolution equations during the matter-radiation transition, but such details are not necessary here. The growth function and the transfer function together relate the potential deep in the matter era to the primordial potential. The late time potential can then be written as

$$\Phi(\vec{k}, a) = \frac{9}{10} T(k) \frac{D(a)}{a} \Phi_p(\vec{k}) . \quad (2.2.29)$$

With this expression for the evolution of the potential from the initial conditions to the matter era, we can obtain an expression for the late-time matter overdensity by using the large- $k$  limit of equation (2.2.26):

$$\delta(\vec{k}, a) = \frac{k^2}{4\pi G \bar{\rho}(a) a^2} \Phi(\vec{k}, a) \quad (2.2.30)$$

$$= \frac{2k^2 a}{3\Omega_m H_0^2} \Phi(\vec{k}, a) \quad (2.2.31)$$

where we have used equations (2.1.12) and (2.1.15) as well as equations (2.1.21) and (2.1.24) to obtain the final expression. Using the expression for the late-time potential in equation (2.2.29), we obtain the final expression for the late-time density contrast:

$$\delta(\vec{k}, a) = \frac{3k^2}{5\Omega_m H_0^2} T(k) D(a) \Phi_p(\vec{k}) . \quad (2.2.32)$$

We therefore have analytic descriptions relating the late-time density contrast and potential to the primordial potential. Though we have performed this discussion within the context of the matter era, equations (2.2.29) and (2.2.32) still hold in the present dark energy era, as long as we modify the growth function accordingly. In general, the growth function can be written as

$$D(a) = \frac{5\Omega_m}{2} \frac{H(a)}{H_0} \int_0^a da' \left( \frac{H_0}{a' H(a')} \right)^3 . \quad (2.2.33)$$

The cosmological perturbation theory developed in this section is successful in describing the anisotropies and homogeneities observed in the LSS. Even so, it is limited. In particular one cannot hope to describe in detail the formation of the highly non-linear objects that make up the LSS, such as galaxies, halos, and clusters, using a perturbation theory of this type. Though it is possible to extend the perturbation theory beyond linear order, such extensions require substantial computation with low returns. An alternative and supplementary approach is thus necessary for an adequate description of LSS.

### 2.3 The Biasing of Dark Matter Halos

In order to understand the scale-dependent bias of dark matter halos, we must first introduce the concept of a dark matter halo and understand how to treat the physics of these objects. Dark matter halos can be understood in a simplistic manner as an end state of the gravitational collapse of a distribution of cold dark matter. In other words a dark matter halo is a bound gravitational system of dark matter. The formation of such objects cannot be described in detail using the perturbation theory developed in the previous section, due to the highly non-linear nature of the process. However perturbation theory provides insight into how this formation process begins, with halos being more likely to form in overdense regions. It is not possible to develop a completely rigorous analytical description of gravitational collapse, but approximations can be used to understand the most important features of this phenomenon [1][17]. A simple model used to describe the formations of halos is that of spherical collapse [1][17]. In this model, a perfectly spherical region of the universe is considered with a density  $\rho = \bar{\rho}(1 + \delta)$ . Due to the enhanced gravitational attraction, this spherical region will expand at a rate that is slightly slower than the universe's expansion rate. This will increase the density contrast, which will in turn slow the expansion even more. If the initial overdensity of the spherical region is large enough, the expansion of this region will eventually cease and the sphere will begin to recollapse. The initial overdensity of the region required for this collapse to happen can be approximated using the Friedmann equation. A number of properties of dark matter halos can be estimated using this model, such as the mass dispersion of the halos [1][17][26]. For our purposes the most important result of this model is the intuitive understanding of a dark matter halo as a region in space where the density contrast exceeds some threshold value. This is known as *barrier-crossing* [1][17][26][27][28]. The model of spherical collapse can be extended to account for lack of spherical symmetry. Estimates of the angular momentum of halos can also be made.

In addition to developing some understanding of the formation of dark matter halos, we are interested in understanding their spatial distribution in the universe. In analogy to the underlying dark matter distribution, the distribution of halos can be expressed in terms of a number density field,  $n_h(\vec{x})$ . This number density can be decomposed into a mean isotropic and homogeneous density,  $\bar{n}_h$ , and a fractional overdensity,  $\delta_h(\vec{x})$ , so that

$$\delta_h(\vec{x}) = \frac{n_h(\vec{x}) - \bar{n}_h}{\bar{n}_h} . \quad (2.3.1)$$

The abundance of dark matter halos in the universe is then described by this density contrast. We would like to understand how the distribution of halos is related to the underlying distribution of dark matter. Since halos are collapsed regions of dark matter, we expect that regions with a larger matter density will also have a large density of halos. We cannot assume however that the two densities are equal. In fact the two should be different. In the model of barrier crossing, we described halos as regions where the matter density in a small region of space exceeds some threshold value. If we decompose the matter density into small- and large-scale fluctuations, i.e.  $\delta = \delta_s + \delta_l$ , we can imagine that halos might form in regions of large  $\delta_l$ , even if  $\delta_s$  is less than the threshold value. Conversely, in regions of small  $\delta_l$ , halos aren't as likely to form even if  $\delta_s$  exceeds the threshold. This decomposition is known as *peak-background split* [1]. We can describe the dependence of the halo density on the matter density by noting that the formation of halos from the underlying dark matter is expected to be a primarily local process. This approximation is reasonable due to the fact that the spatial velocity of cold dark matter is mostly negligible. Perturbations in the dark matter in one region of space are thus not influenced by perturbations in far-off regions on any reasonable time scale. This suggests that we can write the halo density as an arbitrary local functional of the matter density [1][27][29][30][31][32]:

$$\delta_h(\vec{x}) = F[\delta(\vec{x})] . \quad (2.3.2)$$

This procedure is known as *local biasing* of the halo density. In practice we do not know the form of the functional  $F[\delta]$ . However we are able to perform a Taylor expansion around the matter density to obtain

$$\delta_h(\vec{x}) = b_1\delta(\vec{x}) + b_2\delta^2(\vec{x}) + b_3\delta^3(\vec{x}) + \dots \quad (2.3.3)$$

This does not provide any new information about the relation between the halo and matter densities, but it allows us to confine our ignorance to the bias parameters,  $b_i$ . Furthermore it allows us to treat the bias of halos perturbatively. In doing so, the order of perturbation must be consistent with that used in the development of the fluid equations for the matter density. In Section 2.2, we only developed these equations to linear order. We therefore restrict our attention to local linear biasing:

$$\delta_h(\vec{x}) = b_1\delta(\vec{x}) . \quad (2.3.4)$$

The absence of higher order terms allows us to express this straightforwardly in terms of the Fourier modes:

$$\delta_h(\vec{k}) = b_1\delta(\vec{k}) . \quad (2.3.5)$$



The local bias ansatz provides a simple model for relating the halo distribution to the underlying distribution of dark matter. The linear bias presented here can be improved upon by including contributions from higher orders. Considerable progress has been made to this end in recent years using perturbative techniques from quantum field theory and statistical mechanics [28][29][30][31][32]. It has also been shown that proper consideration of higher orders must include non-local terms to account for the history of the formation process [30][32]. For the purpose of our research, however, the local linear bias of equation (2.3.4) will suffice.

## 2.4 The Power Spectrum

The ultimate aim of a theory of LSS is to make contact with observation using quantitative predictions. The theory that we have developed in the previous sections, from the FLRW universe to halo bias, has been deterministic. Given the primordial potential,  $\Phi_p$ , the distribution of the perturbations at late times is obtained using the evolution equations. When attempting to connect theory with observation however, we cannot expect that the theory of LSS will provide a description of the *exact* distribution of galaxies and clusters in our universe [1][17]. We can however make statistical predictions about these distributions using a statistical description of the structure in our universe. It was mentioned in the introduction that the primordial perturbations characterized by  $\Phi_p$  are believed to have been generated stochastically. The random nature of the perturbations carries over to the present epoch by virtue of the deterministic nature of the evolution equations developed in Section 2.2. A statistical description of LSS can thus be obtained by considering the statistics of the matter density field. To this end we consider an ensemble of universes with a variety of distributions of matter but with the same statistical properties. This ensemble is called a *random field*. The specific form of  $\delta$ , such as the one describing our universe, is a *realization of the random field* [1]. Given two points in space,  $\vec{x}$  and  $\vec{y}$ , we can describe the correlation between the densities at these points via the quantity  $\langle \rho(\vec{x})\rho(\vec{y}) \rangle$ . The angular brackets denote an ensemble average over the random field. In the case of a homogeneous FLRW universe, this is equal to the square of the mean density,  $\bar{\rho}$ . However, in the context of linear perturbation theory, this must be modified in the following way:

$$\langle \rho(\vec{x})\rho(\vec{y}) \rangle = \bar{\rho}^2 \langle [1 + \delta(\vec{x})][1 + \delta(\vec{y})] \rangle \quad (2.4.1)$$

$$= \bar{\rho}^2 [1 + \langle \delta(\vec{x})\delta(\vec{y}) \rangle] \quad (2.4.2)$$

$$= \bar{\rho}^2 [1 + \xi(\vec{x}, \vec{y})] \quad (2.4.3)$$

with

$$\xi(\vec{x}, \vec{y}) \equiv \langle \delta(\vec{x})\delta(\vec{y}) \rangle . \quad (2.4.4)$$

We have used the fact that  $\langle \delta \rangle = 0$  from the definition of  $\delta$ . In the process we have defined the two-point correlation function  $\xi(\vec{x}, \vec{y})$ . The correlation function describes the excess correlation between the structure at  $\vec{x}$  and that at  $\vec{y}$  [1]. Since the universe is measured to be homogeneous, the correlation function between two points can only depend on the separation between those points,  $\vec{x} - \vec{y}$ , and not on their absolute positions [1]. Furthermore, due to the isotropy of the universe, it cannot depend on the

direction of this separation vector, and thus can only depend on the distance between the two points,  $|\vec{x} - \vec{y}|$  [1]. An important point of consideration is that the ensemble average over the random field cannot be measured using observational data due to the fact that we only experience one realization of the random field. However, this is salvaged by the fact that the random field is homogeneous. In this case the ensemble average can be replaced by a spatial average over pairs of points in our universe with the same separation. This fact, known as the *ergodic theorem*, is what allows us to connect the statistical description of structure in the universe with observation [1][33]. From the correlation function, we can obtain a statistical measure of the structure in the universe by taking the Fourier transform. This allows us to define the *power spectrum* of the density contrast,

$$\langle \delta(\vec{k}) \delta(\vec{k}') \rangle = (2\pi)^3 P_{mm}(k) \delta^{(3)}(\vec{k} + \vec{k}') \quad (2.4.5)$$

where  $\delta^{(3)}(\vec{k})$  is the three-dimensional Dirac delta function. The delta function is simply used to enforce the fact that  $\vec{k}' = -\vec{k}$ . In other words, for a Gaussian random field the individual modes are uncorrelated. The function  $P_{mm}(k)$  is the power spectrum. The subscript  $mm$  stands for ‘matter-matter’, to distinguish from other two-point correlators that will be discussed shortly. Again, due to the symmetries of the universe, the power spectrum depends only on the magnitude of the wave vector  $\vec{k}$ , and not on its orientation. Since the magnitude of the wave vector is a measure of scale in the universe, the power spectrum describes the degree of correlation in  $\delta$  at different scales, and thus the degree of structure at different scales. For a Gaussian random field, the power spectrum provides a complete characterization of the statistics of the field. This is not true if the field is non-Gaussian, in which case higher order correlation functions are required to characterize the statistics. We can relate the late time matter power spectrum to the primordial power spectrum using equation (2.2.32). This results in

$$P_{mm}(k) = \frac{9k^4}{25\Omega_m^2 H_0^4} T^2(k) D^2(a) P_{\Phi\Phi}(k) \quad (2.4.6)$$

where  $P_{\Phi\Phi}$  is the power spectrum of the primordial potential, which can be written as [17]

$$P_{\Phi\Phi}(k) = \frac{A_s}{k^3} k^{n_s-1} . \quad (2.4.7)$$

The quantity  $n_s$  is the *spectral scalar index* and  $A_s$  is the *scalar amplitude*. The best estimates of these quantities measured using the CMB are [21]

$$n_s = 0.968 \pm 0.006 \quad (2.4.8)$$

$$A_s = (2.142 \pm 0.049) \times 10^{-9} . \quad (2.4.9)$$

These values correspond to a power spectrum expressed as [21]

$$P_{\Phi\Phi}(k) = A_s \left( \frac{k}{k_0} \right)^{n_s-1} \quad (2.4.10)$$

where  $k_0 = 0.05 \text{Mpc}^{-1}$ .

In addition to computing the power spectrum for the matter density, we can also compute the power spectrum for the halo density,

$$\langle \delta_h(\vec{k}) \delta_h(\vec{k}') \rangle = (2\pi)^3 P_{hh}(k) \delta^{(3)}(\vec{k} + \vec{k}') \quad (2.4.11)$$

as well as the matter-halo two-point correlation function, or cross spectrum:

$$\langle \delta(\vec{k}) \delta_h(\vec{k}') \rangle = (2\pi)^3 P_{mh}(k) \delta^{(3)}(\vec{k} + \vec{k}') . \quad (2.4.12)$$

The subscripts  $_{hh}$  and  $_{mh}$  stand for ‘halo-halo’ and ‘matter-halo’. By applying the linear bias model developed in Section 2.3, we can relate the halo-halo and matter-halo spectra to the underlying matter power spectrum in the following way:

$$P_{mh}(k) = b_1 P_{mm}(k) \quad (2.4.13)$$

$$P_{hh}(k) = (b_1)^2 P_{mm}(k) . \quad (2.4.14)$$

The power spectrum is perhaps the most important observable of LSS. In particular the shape of the matter power spectrum in equation (2.4.6) is dependent on the details of the transfer function and thus contains a wealth of information about the history of the universe. Additionally, the halo power spectrum and the matter-halo cross spectrum provide ways of testing the validity of the local bias ansatz, and testing our understanding of dark matter halos. As we will see in the next section, the spectra can also be used to constrain the degree of non-Gaussianity in the primordial perturbations, using the scale-dependent bias of dark matter halos.

## 2.5 Scale-Dependent Bias

The physics discussed so far has been predicated on the idea that the primordial perturbations obey Gaussian statistics. This is not unreasonable since the statistics of the perturbations have been constrained by measurements of the CMB to be very close to Gaussian [12]. Furthermore, the simplest theoretical scenarios for generating primordial perturbations that aren’t fine-tuned predict Gaussian statistics [17][19]. However, a number of alternative models that have been proposed to describe the generation of primordial fluctuations result in varying degrees of local, equilateral, and orthogonal non-Gaussianities [3]. We restrict our attention to non-Gaussianities of the local type. In order to account for a wide range of models, the non-Gaussian primordial potential that arises due to local non-Gaussianities is expressed in the following way [3][4][5][6][7][8][9] :

$$\Phi_p(\vec{x}) = \phi_p(\vec{x}) + f_{NL}(\phi_p^2(\vec{x}) - \langle \phi_p^2 \rangle) \quad (2.5.1)$$

where  $\phi_p$  is an underlying potential obeying Gaussian statistics. The parameter  $f_{NL}$  serves as a measure of the degree of non-Gaussianity in the model. As mentioned in the introduction, the current best estimate of  $f_{NL}$  obtained from measurements of the CMB temperature is  $f_{NL} = 2.5 \pm 5.7$ . When the temperature data is combined with polarization data, the estimate is  $f_{NL} = 0.8 \pm 5.0$  [12] .

In addition to producing effects in the CMB, non-Gaussianities are expected to have a measurable effect on a number of features of LSS. In particular, it has been shown that non-Gaussianities should

affect the biasing of dark matter halos on large scales [3]. The details of this effect, known as scale-dependent bias, are developed in this section. To begin, we write the primordial potential from equation (2.5.1) as

$$\Phi_p(\vec{x}) = \phi_p(\vec{x}) + f_{NL}\phi_p^2(\vec{x}) \quad (2.5.2)$$

where the variance of the Gaussian potential has been omitted for simplicity. We now apply the method of peak-background split discussed in Section 2.3, and decompose the Gaussian potential into small- and large-scale terms:

$$\phi_p(\vec{x}) = \phi_{p,s}(\vec{x}) + \phi_{p,l}(\vec{x}) . \quad (2.5.3)$$

The subscripts ‘s’ and ‘l’ denote ‘small’ and ‘large’. Applying this decomposition to equation (2.5.2) gives

$$\Phi_p(\vec{x}) = \phi_{p,s}(\vec{x}) + \phi_{p,l}(\vec{x}) + f_{NL} [\phi_{p,s}^2(\vec{x}) + 2\phi_{p,s}(\vec{x})\phi_{p,l}(\vec{x}) + \phi_{p,l}^2(\vec{x})] . \quad (2.5.4)$$

Next we define the small-scale non-Gaussian potential to be

$$\Phi_{p,s}(\vec{x}) = \phi_{p,s}(\vec{x}) + 2f_{NL}\phi_{p,s}(\vec{x})\phi_{p,l}(\vec{x}) \quad (2.5.5)$$

where we have dropped the quadratic contribution from the small-scale Gaussian potential. Our primary interest in using the peak-background split is to examine how the small and large scales are coupled by non-Gaussianities. The quadratic small-scale Gaussian potential only results in a correction to  $\phi_{p,s}$  and does not provide additional information about this coupling. We now express equation (2.5.5) in Fourier space as

$$\Phi_{p,s}(\vec{k}) = \phi_{p,s}(\vec{k}) + 2f_{NL}\phi_{p,s}(\vec{k})\phi_{p,l}(\vec{x}) \quad (2.5.6)$$

$$= \phi_{p,s}(\vec{k}) [1 + 2f_{NL}\phi_{p,l}(\vec{x})] . \quad (2.5.7)$$

The large-scale potential has not been transformed to Fourier space since we are effectively treating it as a constant in comparison to the the small-scale potential. On small scales, the large-scale potential serves primarily to modulate the small-scale potential. The position dependence of  $\phi_{p,l}$  is retained however, since the ‘constant’ value of the potential still changes from region to region. In other words, within the small-scale region under consideration, described by  $\Phi_{p,s}$ , the large-scale potential is approximately constant. If we then consider another small region somewhere else in space, the large-scale potential will also be constant in that region, but that constant might have a different value. To make it clear that the large-scale potential only varies on large scales, we will write the position argument as  $\vec{x}_l$ .

We can now apply the solution from equation (2.2.32) to  $\Phi_{p,s}$  to obtain the late time small-scale non-Gaussian matter density, which we denote  $\Delta_s$ . This gives

$$\Delta_s(\vec{k}, a) = \frac{3k^2}{5\Omega_m H_0^2} T(k) D(a) \phi_{p,s}(\vec{k}) [1 + 2f_{NL}\phi_{p,l}(\vec{x}_l)] . \quad (2.5.8)$$

Next, using equation (2.2.29) to evolve  $\phi_{p,s}$  forward in time, we obtain

$$\Delta_s(\vec{k}, a) = \frac{2k^2 a}{3\Omega_m H_0^2} \phi_s(\vec{k}, a) [1 + 2f_{NL}\phi_{p,l}(\vec{x}_l)] \quad (2.5.9)$$

$$= \delta_s(\vec{k}, a) [1 + 2f_{NL}\phi_{p,l}(\vec{x}_l)] \quad (2.5.10)$$

where we have used equation (2.2.31) to obtain the second line. Here  $\delta_s$  is the small-scale Gaussian overdensity associated with the underlying small-scale Gaussian potential  $\phi_s$ . We can now compute the non-Gaussian matter power spectrum to find

$$P_{NG,s}(k) = P_s(k) [1 + 2f_{NL}\phi_{p,l}(\vec{x}_l)]^2 . \quad (2.5.11)$$

$P_s(k)$  is the small-scale contribution from the Gaussian matter power spectrum, as defined in equation (2.4.5). This result demonstrates that in the case of non-Gaussian initial conditions, the matter spectrum on small scales at late times receives a contribution from the large-scale primordial fluctuations. Another way to characterize the power on a small scales, i.e. within the region of space that we are considering, is via the quantity  $\sigma_8$ , which evaluates the power spectrum smoothed over spherical regions with radii equal to  $8h^{-1}\text{Mpc}$ . This can be expressed as

$$\sigma_8^2 = \int \frac{d^3k}{(2\pi)^3} P(k) |W_8(\vec{k})|^2 \quad (2.5.12)$$

where  $W_8(\vec{k})$  is the Fourier transform of an appropriate window function used to smooth the density field. Instead of using the power spectrum in equation 2.5.11, we can write it in terms of  $\sigma_8$ :

$$\sigma_{8,NG}^2 = \sigma_8^2 [1 + 2f_{NL}\phi_{p,l}(\vec{x}_l)]^2 \quad (2.5.13)$$

so

$$\sigma_{8,NG} = \sigma_8 [1 + 2f_{NL}\phi_{p,l}(\vec{x}_l)] . \quad (2.5.14)$$

Treating the non-Gaussian contribution as a small deviation from Gaussianity, we can write this as

$$\sigma_{8,NG} = \sigma_8 + \delta\sigma_8 \quad (2.5.15)$$

with

$$\delta\sigma_8 = 2f_{NL}\phi_{p,l}(\vec{x}_l)\sigma_8 . \quad (2.5.16)$$

In the case of Gaussian statistics, the value of  $\sigma_8$  is expected to be the same throughout the universe. However from equation (2.5.16), we see that the small-scale power characterized by  $\sigma_8$  varies from region to region when non-Gaussianities are introduced. We can express this large-scale functional dependence as  $\sigma_{8,NG}(\vec{x}_l)$ .

We would now like to use these results to examine how the density of halos is affected by the non-Gaussian initial conditions. In particular, we would like to consider how  $n_h(\vec{x}_l)$  behaves on large

scales. Instead of using  $\vec{x}_l$  as the argument, we will parameterize the large-scale position dependence using  $\sigma_{8,NG}$ . To distinguish from the Gaussian halo density, we write the non-Gaussian counterpart as  $n_{h,NG}(\sigma_{8,NG})$ . This quantity can be Taylor expanded around the Gaussian limit, yielding

$$n_{h,NG}(\sigma_{8,NG}) = n_{h,NG}(\sigma_8 + \delta\sigma_8) \quad (2.5.17)$$

$$\sim n_h(\sigma_8) + \left. \frac{\partial n_h}{\partial \sigma_8} \right|_G \delta\sigma_8 \quad (2.5.18)$$

where the partial derivative in the second line is evaluated in the Gaussian limit, denoted by  $G$ . Then using equation (2.5.16) we have

$$n_{h,NG}(\sigma_{8,NG}) \sim n_h(\sigma_8) + 2f_{NL}\phi_{p,l}(\vec{x}_l) \left. \frac{\partial n_h}{\partial \sigma_8} \right|_{n_h=\bar{n}_h} \sigma_8 . \quad (2.5.19)$$

Now using the definition of the halo density contrast in equation (2.3.1), we can write

$$\bar{n}_h + \bar{n}_h \delta_{h,NG} = \bar{n}_h + \bar{n}_h \delta_h + 2f_{NL}\phi_{p,l}(\vec{x}_l) \left. \frac{\partial \log n_h}{\partial \log \sigma_8} \right|_{n_h=\bar{n}_h} \bar{n}_h . \quad (2.5.20)$$

Therefore

$$\delta_{h,NG} = \delta_h + 2f_{NL}\phi_{p,l}(\vec{x}_l) \frac{\partial \log n_h}{\partial \log \sigma_8} . \quad (2.5.21)$$

Finally, by using the local linear bias ansatz on the Gaussian halo density, as in equation (2.3.4), transforming to Fourier space, and applying the linear solution to the matter density from equation (2.2.32), we have

$$\delta_h(\vec{k}) = \left[ b_1 + \frac{10\Omega_m H_0^2}{3k^2 T(k) D(a)} f_{NL} \frac{\partial \log n_h}{\partial \log \sigma_8} \right] \delta(\vec{k}) \quad (2.5.22)$$

This result is the scale-dependent bias of the halo density. In the case of Gaussian initial conditions it reduces to the linear bias approximation. The associated halo-halo power spectrum and matter-halo cross spectrum can be computed, yielding

$$P_{hh}(k) = \left[ b_1 + \frac{10\Omega_m H_0^2}{3k^2 T(k) D(a)} f_{NL} \frac{\partial \log n_h}{\partial \log \sigma_8} \right]^2 P_{mm}(k) \quad (2.5.23)$$

and

$$P_{mh}(k) = \left[ b_1 + \frac{10\Omega_m H_0^2}{3k^2 T(k) D(a)} f_{NL} \frac{\partial \log n_h}{\partial \log \sigma_8} \right] P_{mm}(k) \quad (2.5.24)$$

as expected. This concludes our discussion of the central theoretical aspects of LSS. We now turn our attention to a discussion of the tools required in order to make measurements of the scale-dependent bias using data from numerical simulations.

### 3 Numerical Analysis

In the previous section we developed some the most important theoretical aspects of modern cosmology pertaining to the study of LSS. Though it does capture the broad concepts and provides us with an important degree of intuition, this analytic description is limited in its scope, particularly when one considers the highly non-linear processes involved in structure formation. This was especially apparent in Section 2.3, where we had to rely heavily on approximations to describe the formation of dark matter halos. In general one cannot aspire to develop a completely analytical formulation of the processes involved in the formation of LSS. Fortunately one can supplement analytical considerations with numerical analysis and simulations. Such numerical considerations, especially the use of cosmological simulations, have become an indispensable tool in the contemporary study of LSS, allowing us to examine in detail the complex gravitational dynamics involved in a way that isn't accessible using only analytical considerations [1][22][34][35]. In addition to this, numerical analysis allows us to prepare for the experimental data that will become available in the next decade or so due to surveys of LSS such as, for example, the Dark Energy Survey (DES) [13], the Large Synoptic Survey Telescope (LSST) [14], the Euclid mission [16], the Dark Energy Spectroscopic Instrument (DESI) [15], and so on [10]. In this section we discuss the numerical tools that were used over the course of this research project. We relied heavily on the use of cosmological  $N$ -body simulations to obtain our data. These simulations were run using the GADGET-2 code [34][35], with the initial conditions generated using the 2LPT initial conditions code [36]. Furthermore, in order to measure the biasing of dark matter halos, it was necessary to be able to identify halos in the output of the GADGET-2 simulations. For this purpose we used the ROCKSTAR phase-space halo finder [37]. Finally, it was necessary to write in-house programs in order to compute the matter and halo power spectra from the GADGET-2 and ROCKSTAR data, and to measure the linear and scale-dependent halo bias.

#### 3.1 $N$ -Body Simulations and GADGET-2

As mentioned in Section 2.2, the formation of structure in the universe is driven predominantly by gravitation. Furthermore, the most significant processes involved in structure formation take place during the matter era and are influenced primarily by the dark matter distribution due to its large abundance compared to baryonic matter. The purpose of cosmological simulations then, such as those performed by GADGET-2, is to track the gravitational time-evolution of the dark matter distribution using numerical methods [1]. Within the framework of cosmological perturbation theory, we described the dark matter distribution as a collisionless self-gravitating fluid. The fluid approximation was used in lieu of the underlying  $N$ -body interactions between particles due to the limitations of analytic techniques. When working with simulations however it is possible to express the situation in terms of the gravitational  $N$ -body problem. In this case the matter is described in terms of particles instead of as a fluid [34]. Studying the gravitational dynamics of dark matter as an  $N$ -body problem is advantageous since the fluid approximation ceases to be valid in the non-linear regimes for which we would like to use simulations as a complementary source of information. In practice however, one cannot hope to perform a reasonable simulation of the universe by computing the dynamics of each individual elementary dark matter particle. The number of particles involved in such a computation would be too great to simulate the dynamics on

any reasonably large scale. Consequently,  $N$ -body simulations approximate large numbers of particles using representative macroscopic particles with significantly larger masses [1][34][35]. This is a reasonable approximation on large scales where the internal structure of the representative particles cannot be resolved.

The dynamics of the dark matter distribution can be studied in simulations using the numerical time integration of Newton's second law for  $N$  particles interacting gravitationally. This can be accomplished by using a reasonable integration scheme [34][35]. However, for every time step in the integration, one is required to compute the total gravitational interaction between every particle pair in the simulation [1]. In terms of force this interaction can be represented as

$$\vec{F}_i = \sum_{j \neq i} \frac{m^2 (\vec{r}_j - \vec{r}_i)}{|\vec{r}_j - \vec{r}_i|^3} \quad (3.1.1)$$

where  $\vec{F}_i$  and  $\vec{r}_i$  are, respectively, the force on and the position of the  $i$ th particle, and  $m$  is the mass of the representative particles. The summation extends over all particles in the simulation, and each particle will be associated with an expression of this form. Conceptually, equation (3.1.1) is a straightforward statement, but in practice this requires extensive computational power, especially in the case of large simulations. For each time step, the complete computation of the force involves a double sum over the total number of particles. Thus one cannot hope to compute the gravitational interaction in this way. The inefficiency of the method of direct summation has lead to development of other methods of computing the force, or alternatively the gravitational potential, in the simulation volume [1][22][34]. In GADGET-2, the method used is that of a TreePM algorithm, in which the computation of the gravitational interaction across the entire simulation volume is broken down into two parts based on scale: a particle-mesh (PM) method is used to compute the interaction between particles with large separations, while a hierarchical Tree algorithm is used to compute the interaction at short ranges. The two methods are described in greater detail below.

The general aim of the PM method is to compute the gravitational field using Fourier techniques [1][34]. To implement this method, one begins by defining a discrete grid, or mesh, on the simulation volume. A particle distribution scheme, such as the cloud-in-cell algorithm [35][38] (discussed in detail in Section 3.4), is then used to convert the discrete particle distribution in the volume to a mass density field on the grid. One can then solve Poisson's equation for the gravitational potential on the grid using a Fast Fourier Transform [1][22][35]. The corresponding force field can be obtained by applying a finite difference technique to the potential [35]. Once the force field on the grid is obtained, it can be interpolated back to the particles [22][35]. The PM method is a quick and efficient means of computing the force. However, it suffers from the fact that spatial resolution is lost when moving from the particles to the mesh [1]. The spatial resolution of the resulting force field is limited by the mesh size. However since the force between two particles is proportional to their separation, the loss of resolution has a less significant effect when the particles are widely separated. This suggests that one can mitigate the effects of the loss of resolution while still making use of the efficiency of the PM method by considering separately the long range and short range interactions [1][22][34][35]. This is the motivation behind the TreePM algorithm used in GADGET-2.

Though a number of different methods can be used to compute the short range force, including the direct summation discussed above, the hierarchical Tree method has the advantage of a relatively low



computational cost [22][34][35]. In such an algorithm, one organizes the particles into a hierarchical, or tree, structure by dividing the entire integration volume into consecutive nested octants called nodes [34][35][39]. To be precise, one begins with the entire volume as the so-called ‘root’ node. This root node is then subdivided into eight cubes of equal side length, the daughter nodes. Each of these nodes is subdivided into eight more nodes. This process continues until each node encompasses either a single particle or none [34][35][39]. A schematic representation of this decomposition is presented in Figure 1. Furthermore, for each individual node in the tree, one computes a multipole expansion of the potential using the particles contained within that node. Thus each node has an associated potential, expressed as a multipole expansion [34][35]. For practical purposes this multipole expansion is truncated at some order. In the case of GADGET-2, only the monopole is retained [34][35]. Once the particles have been sorted into the tree, the force on a specific particle is obtained by a process known as ‘walking the tree’. This entails hierarchically going through the tree starting at the root (which contains the entire simulation volume) and, for each node, deciding whether the force approximated by the monopole in that particular node is accurate enough [34][35]. This decision is made based on a condition known as an ‘opening criterion’ [34][35]. The specific form of the opening criterion is dependent on the implementation. If the monopole in that node is accurate enough, it is used as the force contribution from that region. If it is not accurate enough, the node is ‘opened’ and its daughter nodes are considered in turn. This process is repeated until all particles are accounted for [34][35]. In the end the force obtained by walking the tree will only be an approximation of the true force on a given particle. However, the degree of accuracy of this force can be modified by changing the opening criterion and the order at which to truncate the multipole expansion for each node.

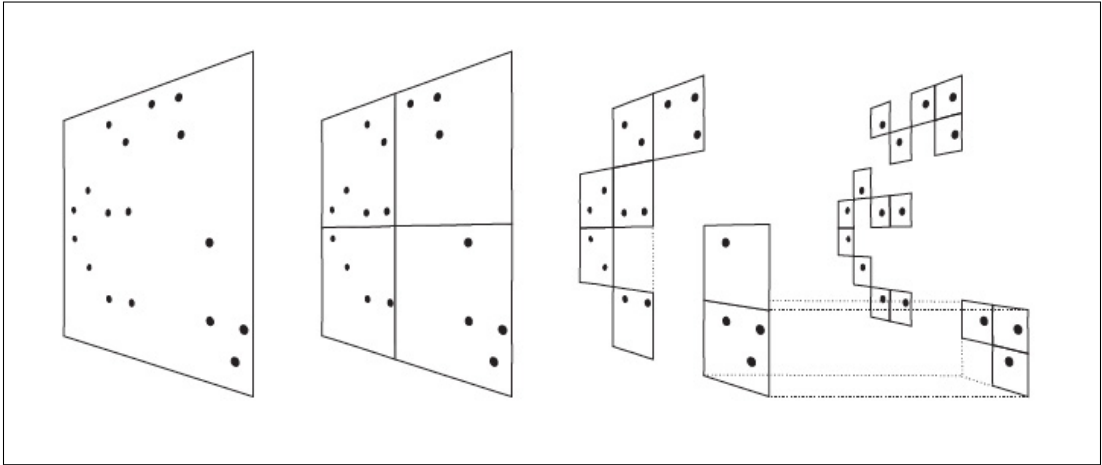


Figure 1: A schematic representation of the hierarchical tree decomposition for a two-dimensional simulation [34].

By using the PM method to compute the long-range force, and the Tree method to compute the short-range force, GADGET-2’s TreePM algorithm provides an efficient and cost-effective means of computing the force field in the simulation volume. These algorithms, in addition to GADGET-2’s integration schemes and parallelization strategies, which we won’t discuss here, allow GADGET-2 to perform large cosmological simulations in reasonable amounts of time. The code returns a number

of outputs pertaining to the simulation. For our purposes the most important output is the position information of all the particles in the simulation volume. In order to run simulations at all, however, GADGET-2 requires a file containing the initial distribution of the particles, i.e. an initial conditions file.

### 3.2 Generating the Initial Conditions

In the previous section, we discussed the way in which cosmological simulations can be used to obtain valuable information about the formation of LSS. We discussed the basic ideas implemented in numerical simulations, more specifically in the case of the GADGET-2 code. An important part of cosmological simulations that was not mentioned in the previous section is the generation of initial conditions. Provided with the initial perturbed positions of the particles in the simulation, GADGET-2 is capable of evolving the perturbations in time using the methods discussed above. However these initial conditions are not given and must be prepared separately. In order to generate initial conditions, we chose to use the 2LPTIC code §. The primary benefit of using this code for this research project is that it is capable of generating non-Gaussian as well as Gaussian initial conditions. In addition to this, the code computes the initial conditions using second-order Lagrangian Perturbation Theory (2LPT), instead of first-order Lagrangian Perturbation Theory, otherwise known as the Zel'dovich approximation. One of the advantages of using 2LPT over the Zel'dovich approximation is that the former significantly reduces the effect of transients on  $N$ -body simulations [36][40][41]. Moreover, the use of 2LPT incorporates more of the non-linear effects of gravity and is thus expected to be more accurate.

In order to generate initial conditions for a simulation, one begins with a homogeneous distribution of particles throughout the simulation volume. The aim of the initial conditions code is then to generate an initial distribution of perturbations in which to organize these particles. This perturbed distribution serves as the initial conditions for the  $N$ -body simulation. In broad terms the process of generating such a distribution can be described as follows. One begins by generating the primordial Gaussian or non-Gaussian potential using a prescribed technique. This can be accomplished in a number of different ways, some more complex than others, especially in the case of non-local non-Gaussian potentials [3][36][41][42][43]. Instead of going into the details of such techniques, we will use the case of local non-Gaussianities as an example. In this case the primordial potential is of the form presented in equation (2.5.1). One can generate the primordial potential by first generating a realization of the Gaussian random field  $\phi_p$  using a power law spectrum of the form in equation (2.4.7) or equation (2.4.10) [3][43]. This Gaussian potential can then be squared and its mean subtracted, to obtain the non-Gaussian part of  $\Phi_p$ . Finally, the quadratic contribution is added to the Gaussian potential to generate the full non-Gaussian potential [3][43]. Once equipped with the primordial potential, one can compute the *displacement field*,  $\vec{\Psi}$  (not to be confused with the metric perturbation from Section 2.2), which describes how the final perturbed positions of the particles, denoted  $\vec{x}$ , are related to their initial unperturbed homogeneous positions, denoted  $\vec{q}$ . This can be expressed as follows [22][41]:

$$\vec{x} = \vec{q} + \vec{\Psi}(\vec{q}) . \quad (3.2.1)$$

Thus the displacement field is a map from the initial to final positions. The relation of the displacement

§ <http://cosmo.nyu.edu/roman/2LPT/>

field to the primordial potential is computed using either the Zel’dovich approximation or 2LPT, the latter giving more accurate results [22][41]. After the displacement field is obtained from the primordial potential, the particles are moved to their final perturbed positions. The particles are also assigned the corresponding appropriate velocities, computed from the perturbation theory. These distributions of perturbations and velocities serve as the initial conditions for the  $N$ -body simulation.

### 3.3 The ROCKSTAR Halo Finder

In order to use the data from numerical simulations to measure the scale-dependent bias of dark matter halos, as described in Section 2.5, we need some way of measuring the halo data in our simulations. More specifically we need to be able to identify halos in the simulations, as well as measure some of their properties. In general, this can be done by using a halo finder code. In our case we chose to use the ROCKSTAR halo finder [37]. In order to identify halos in the simulation and measure their properties accurately, one first needs to define what is meant by a dark matter halo. As we saw in Section 2.3, it isn’t immediately clear how to do this rigorously, since the definition of a dark matter halo is ambiguous [37]. This ambiguity in what constitutes a halo can give rise to a number of problems. For example, how does one identify the centre of a halo, or its position, based on the underlying distribution of dark matter particles? One of the basic algorithms used to identify halos in simulations is known as the *friends-of-friends* (FOF) algorithm. In this algorithm, one groups all particles separated by a distance smaller than a specified ‘linking length’, usually chosen as a fraction of the mean particle separation. The particles in this group are then considered to be part of a halo [1][37]. Three-dimensional, i.e. position space, halo finders, such as the FOF algorithm, do a reasonable job of identifying dark matter halos in simulations [37], but they do suffer from some weaknesses. An example of such a weakness is in identifying sub-halos that are close to the centres of their host halos [37]. That being said, one can improve a halo finder’s ability to identify halos by supplementing the position space information with information about the particles’ velocities, thus effectively working in a six-dimensional phase space [37]. The ROCKSTAR halo finder code uses such a phase space algorithm to identify halos [37].

The steps used in ROCKSTAR to measure halos are as follows [37]. The code begins by performing a rapid version of the FOF algorithm in position space, described above, in order to identify general clustered overdense regions in the simulation data. For each FOF group obtained in position space, the code then determines a hierarchy of FOF subgroups in six-dimensional phase space. One of the main challenges pertaining to this step is defining a phase space metric that combines the position and velocity metrics. On their own the position and velocity metrics are well defined as the standard three-dimensional norms, but there is no one established way to combine these two into a single phase space metric. In ROCKSTAR the phase space metric is defined adaptively for each group by scaling the position and velocity metrics by the standard deviations of the position and velocity distributions of the particles within the group. For two particles  $i$  and  $j$  in a given group, the phase space distance is

$$d = \left[ \frac{|\vec{x}_i - \vec{x}_j|^2}{\sigma_x^2} + \frac{|\vec{v}_i - \vec{v}_j|^2}{\sigma_v^2} \right]^{\frac{1}{2}} \quad (3.3.1)$$

where  $\vec{x}_i$  and  $\vec{v}_i$  are the position and velocity, respectively. Further,  $\sigma_x$  and  $\sigma_v$  are the standard deviations in position and velocity space for that particular group. The phase space linking length is chosen based

on this metric so that a constant fraction of the particles in the host group are linked together in subgroups. This process is repeated for each subgroup, recalculating the phase space metric accordingly, until the deepest level of subgroups is found, consisting of a predetermined minimum number of particles. ROCKSTAR then associates each of these lowest subgroups with a ‘seed’ halo. Once the seed halos are generated, the algorithm moves recursively through consecutively higher levels of the subgroup hierarchy, assigning particles to the seed halos based on which halo is closest in terms of phase space distance, until all particles in the original FOF group have been assigned to a halo. In addition to calculating the particle membership of halos, ROCKSTAR is also capable of identifying which particles belong to sub-halos within host halos [37]. Using this algorithm, ROCKSTAR is able to identify halos in the simulation data and accurately measure a number of their properties, such as centre positions, masses at different radii, circular velocities and angular momenta, and so on [37].

### 3.4 Computing the Power Spectrum

In order to measure the linear and scale-dependent biasing of dark matter halos, we must first be able to measure the power spectra from the GADGET-2 and ROCKSTAR data. In this case no external codes were used. Instead it was necessary to write in-house programs to compute the spectra from the simulation data. I wrote these programs over the course of the research project using the Python computing language. Overall, computing the spectra from the simulation data is a fairly algorithmic procedure. To begin, recall the expressions for the spectra, introduced in Section 2.4. Specifically, the expressions for the matter and halo power spectra, as well the expression for the matter-halo cross spectrum are written down in equations (2.4.5), (2.4.11), and (2.4.12). In order to compute these quantities, we must take an ensemble average of the second moments of the density contrasts in Fourier space. The procedure for doing this is identical in the case of the power spectrum and the cross spectrum, so we will frame the discussion in terms of the former.

As discussed in the previous sections, the data generated by GADGET-2 and ROCKSTAR are expressed using the positions of individual particles (or halos) in the volume, rather than a density field. The first step in computing the power spectrum is thus to convert the particle data into a density field. Moreover, we want to define this density on a grid in order to be able to convert to Fourier space using a Fast Fourier Transform. This can be done by means of a particle distribution scheme, as mentioned briefly in Section 3.1. A number of such schemes exist to distribute the particles on to the grid [38]. I used the cloud-in-cell (CIC) algorithm. This algorithm can be described as follows [38]. The number density of  $N$  particles (or halos) in the simulation is

$$n(\vec{x}) = \sum_{i=1}^N \delta^{(3)}(\vec{x} - \vec{x}_i) \quad (3.4.1)$$

where  $\vec{x}$  denotes the coordinates in the simulation volume and  $\vec{x}_i$  is the position of a given particle. The process of defining a density field on the grid using the CIC algorithm can be expressed as a convolution of the number density in the simulation with a window function:

$$n_g(\vec{x}_g) = \int d^3y \, n(\vec{y}) W_{CIC}(\vec{x}_g - \vec{y}) . \quad (3.4.2)$$

The integral is performed over the simulation volume. Further,  $n_g$  is the density defined on the grid,

expressed in terms of its discrete positions  $\vec{x}_g$ .  $W_{CIC}$  is the window function describing the CIC method. Using the expression for  $n(\vec{x})$  in equation (3.4.1) and integrating out the Dirac delta function, we can write

$$n_g(\vec{x}_g) = \sum_{i=1}^N W_{CIC}(\vec{x}_g - \vec{x}_i) . \quad (3.4.3)$$

The functional form of the window function varies depending on the distribution scheme. For the CIC algorithm on a one-dimensional grid, i.e. a discretized line, with spacing  $d$  between grid points it can be written as

$$W_{CIC}(x) \equiv \frac{1}{d} \begin{cases} 1 - \frac{1}{d} |x| & \text{if } |x| < d \\ 0 & \text{otherwise} \end{cases} . \quad (3.4.4)$$

The three-dimensional window function is then defined as the product of three one-dimensional window functions, i.e.

$$W(\vec{x}) \equiv W(x)W(y)W(z) \quad (3.4.5)$$

where the coordinates have been expressed as  $\vec{x} = (x, y, z)$ . It can be verified that the three-dimensional window function for the CIC algorithm generates the density field by distributing each particle to the vertices of the cubic cell that it occupies. Alternatively, the density field at a specific point on the grid,  $\vec{x}_g$ , only receives contributions from the particles that occupy the eight cubic cells adjacent to the grid point. In practice the CIC density field can be implemented numerically by iterating through the particle positions and distributing each particle to the vertices of its cubic cell. The density contrast can then be obtained from the density field using equation (2.2.9).

Once the discretized density contrast is generated, it can be transformed to Fourier space using a Fast Fourier Transform routine. In principle this gives us all that we need to compute the power spectrum. The numerical implementation itself however requires a few more steps, the first of which is to obtain the physical overdensity in Fourier space by removing the effects of CIC distribution scheme. It follows from equation (3.4.2) that the density contrast on the grid,  $\delta_g$ , is the convolution of the physical density contrast in the simulation,  $\delta$ , with the window function. Its Fourier transform will thus be expressed as the product

$$\delta_g(\vec{k}) = \delta(\vec{k})W(\vec{k}) . \quad (3.4.6)$$

The physical density contrast can therefore be obtained by dividing the discretized density contrast in Fourier space by the Fourier transform of the window function. The Fourier transform of the CIC window function has the specific form

$$W(\vec{k}) = \left[ \text{sinc} \left( \frac{\pi k_x}{2k_N} \right) \text{sinc} \left( \frac{\pi k_y}{2k_N} \right) \text{sinc} \left( \frac{\pi k_z}{2k_N} \right) \right]^2 \quad (3.4.7)$$

where  $k_N$  is the Nyquist frequency, defined as  $k_N \equiv \pi/d$ , and we have used the vector components  $\vec{k} = (k_x, k_y, k_z)$ . After the effects of the cloud-in-cell method are removed, we can take the average of the second of moment of the distribution to obtain the power spectrum. In principle it would be

impossible to perform such an ensemble average using data from a single simulation, but the ergodic theorem discussed in Section 2.4 allows us to replace this ensemble average with an average over modes with the same magnitude. In practice it is advantageous to compute this average over modes with a spread in  $k$ . This is due to the fact that the modes are defined on a discrete grid and the number of modes with identical magnitudes are limited. Performing the average using bins of a finite bin width allows us to obtain a better estimate for the power. For a bin containing  $N_b$  modes,  $\delta(\vec{k}_i)$ , with  $i = 1$  to  $N_b$ , the power value for the bin is

$$P_b(k_b) = \langle |\delta(\vec{k})|^2 \rangle \quad (3.4.8)$$

$$= \frac{1}{N_b} \sum_{i=1}^{N_b} |\delta(\vec{k}_i)|^2 \quad (3.4.9)$$

with

$$k_b = \frac{1}{N_b} \sum_{i=1}^{N_b} |\vec{k}_i| . \quad (3.4.10)$$

Further, for a particular bin, the error in the power,  $\Delta P_b$ , is computed from the scatter in the bin using the standard error in the mean:

$$\Delta P_b = \sqrt{\frac{2}{N_b}} \left[ \left\langle \left( |\delta(\vec{k})|^2 - P_b \right)^2 \right\rangle \right]^{\frac{1}{2}} \quad (3.4.11)$$

$$= \sqrt{\frac{2}{N_b}} \left[ \frac{1}{N_b} \sum_{i=1}^{N_b} \left( |\delta(\vec{k}_i)|^2 - P_b \right)^2 \right]^{\frac{1}{2}} . \quad (3.4.12)$$

The prefactor of  $\sqrt{2}$  is to account for the fact that we are double counting the modes that contribute to the power due to the reality condition of the density contrast, i.e.  $\delta^*(\vec{k}) = \delta(-\vec{k})$  so  $|\delta(\vec{k})|^2 = |\delta(-\vec{k})|^2$ . Therefore to get an accurate estimate of the power in the bin, we have to reduce  $N_b$  by half. For the majority of the bins, expressing the error in terms of the scatter in the bin is a reasonable estimate. However this ceases to be true for bins at the lowest values of  $k_b$ . This is due to the fact that the individual modes are defined on a finite grid. There are therefore very few modes with the lowest values of  $k$ . For instance, the smallest value of  $k$  will only be associated with six modes  $\parallel$ . Furthermore, the number of these modes contributing to the scatter is halved, due to the degeneracy in  $|\delta|^2$ . A very large bin width would therefore be necessary to accurately approximate the error using the scatter in small- $k$  modes. On the other hand the bin width need not be this large for the majority of the data. Therefore, in order to make the most of the abundance of data at large values of  $k$ , while still coming up with a reasonable error for the bins at small  $k$ , we approximate the error in the power for the lowest values of  $k_b$  as

$$\Delta P_b = \sqrt{\frac{2}{N_b}} P_b . \quad (3.4.13)$$

$\parallel$  The  $k = 0$  mode is excluded since it has been removed by the definition of the density contrast.

To close off this section, we present a plot of the matter-matter, halo-halo, and matter-halo spectra generated via the method outlined in this section. This is shown in Figure 2. The spectra were computed using matter data from a Gaussian GADGET-2 simulation and the associated ROCKSTAR halo data. The details of the simulation are discussed in Section 4 below. The analytic matter power spectrum used for comparison was obtained by using the CAMB Web Interface [44][45]. Both the CAMB linear and non-linear power spectra are included. The GADGET-2 power spectrum agrees well with the non-linear analytic spectrum on all scales, apart from a few discrepancies on large scales. The importance of using numerical simulations to study the dynamics of dark matter is made apparent when comparing the GADGET-2 data to the linear CAMB spectrum on small scales. This deviation also demonstrates the limitations of the linear perturbation theory developed in Section 2.2. However the plot also shows that the large scales are well represented by the linear approximation. In addition to the matter power spectrum, the halo power spectrum and the cross spectrum have been included, from which we can already see the effects of halo bias. From equations (2.4.13) and (2.4.14), we expect that the cross spectrum will lie somewhere between the matter and halo power spectra, and it does. Moreover, the halo power spectrum and the cross spectrum appear to be a constant multiple of the matter spectrum on large scales. Of course, we would like to be more precise than this, which brings us to a discussion of the measurement of halo bias, both linear and scale-dependent, using matter and halo data from Gaussian and non-Gaussian simulations.

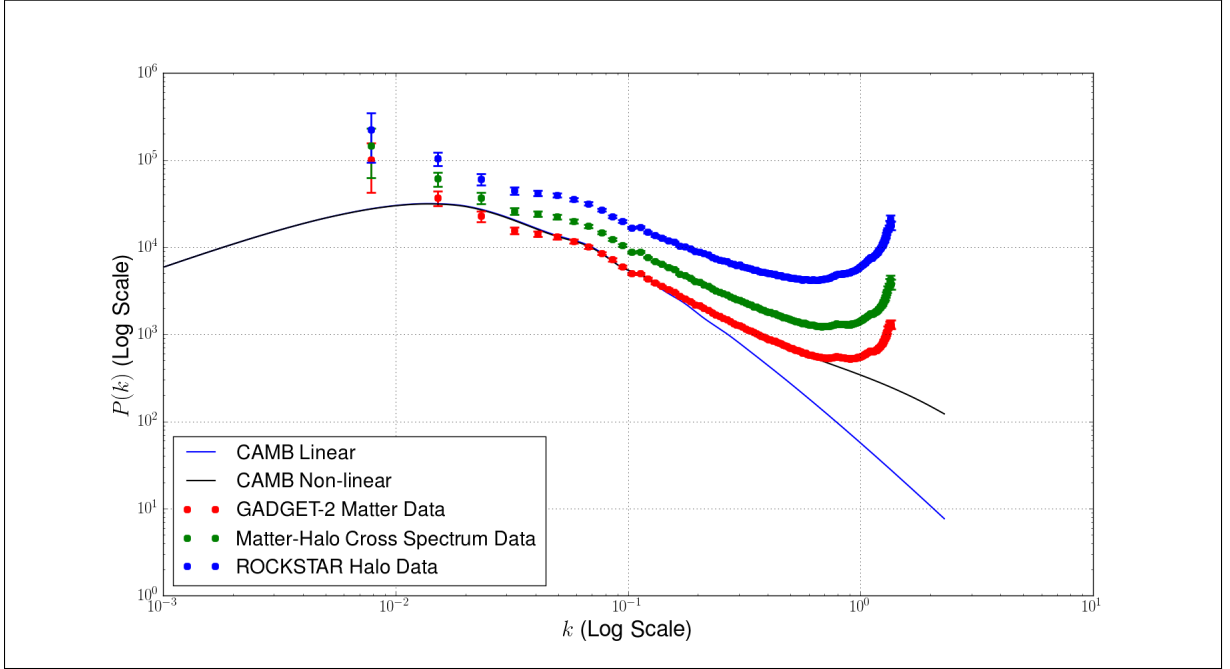


Figure 2: The matter-matter, matter-halo, and halo-halo spectra measured using data from a Gaussian simulation run with a box size of  $800h^{-1}$  Mpc and  $410^3$  particles. Additional parameters are presented in Table 1.

## 4 Results and Discussion

Over the course of the previous section, we discussed in some depths the tools that allowed us to be able to make measurements similar to those that will soon be possible using data from LSS surveys [10][13][14][15][16]. Our primary goal is to use LSS to measure and constrain the degree of non-Gaussianity in the initial conditions. As discussed in Section 2.5, we expect that non-Gaussianities of the local type would have a non-negligible effect on the biasing of dark matter halos, causing deviations from the linear bias approximation on large scales. In this section we present and discuss the results of measurements of linear and scale-dependent bias made using data from Gaussian and non-Gaussian simulations. To this effect, a number of simulations were performed using GADGET-2, with varying degrees of local non-Gaussianity in the initial conditions. As mentioned in Section 3, the initial conditions were generated using 2LPT, and ROCKSTAR was used to identify halos in the simulation outputs. We ran all simulations on the Sunnyvale cluster at CITA¶, using a box size of  $800 h^{-1}\text{Mpc}$  with  $410^3$  particles. Apart from the different values of  $f_{NL}$  used to generate the non-Gaussian initial conditions, the simulations were identical. A list of the simulation parameters is given in Table 1. Before presenting the measurements of scale-dependent bias, we begin with a discussion of the measurements of linear bias in the Gaussian ( $f_{NL} = 0$ ) simulation.

Parameter	Value(s)
Box Size ( $h^{-1}\text{Mpc}$ )	800
Number of Particles	$410^3$
$f_{NL}$	-800, -500, -200, 0, 200, 500, 800
$a(t)$	1.0
$\Omega_m$	0.25
$\Omega_\Lambda$	0.75
$H_0$	70
$n_s$	1.0
Softening Length ( $h^{-1}\text{Mpc}$ )	0.6

Table 1: A list of parameters used to perform the Gaussian and non-Gaussian simulations.

### 4.1 Measuring the Linear Bias

At the end of Section 3.4, we noted that the effects of halo bias could already be seen by eye in the the ROCKSTAR halo power spectrum data and the matter-halo cross spectrum data, presented in Figure 2. Here we endeavour to quantify this statement and make a measurement of the linear bias parameter. In Section 2.4, we introduced the idea of linear local biasing between the halo and matter

¶ <http://wiki.cita.utoronto.ca/mediawiki/index.php/Sunnyvale>



overdensities, expressed in Fourier space as

$$\delta_h(\vec{k}) = b_1 \delta(\vec{k}) . \quad (4.1.1)$$

Additionally, it was shown that at linear order, the local bias ansatz leads to the following relations between the Gaussian spectra:

$$P_{mh}(k) = b_1 P_{mm}(k) \quad (4.1.2)$$

$$P_{hh}(k) = (b_1)^2 P_{mm}(k) . \quad (4.1.3)$$

In practice the matter data is given by the data from GADGET-2, and the halo data is given by the ROCKSTAR data. The bias parameter  $b_1$  can thus be measured from the data by fitting the GADGET-2 data according to these equations. More specifically, we are provided we three ways in which to measure the bias parameter. In order to do this, I wrote in-house programs that performed these fits using a non-linear least squares routine. This was accomplished using the SciPy ‘optimize.curve\_fit’ routine. When fitting at the level of the density contrast, both real and imaginary parts of the fields were fit simultaneously. Before fitting, however, the data had to be modified slightly. As discussed in Section 2.3, the linear bias approximation is only expected to be valid in the regime where linear perturbation theory holds true. Furthermore, we saw in Figure 2 that the matter data only traces the expected linear power spectrum above a certain scale. This is the regime in which we expect the linear bias approximation to hold. Consequently, in order to make a reasonable measurement of the bias parameter, the data was truncated at some scale,  $k_{NL}$ , marking the beginning of the non-linear regime. This cutoff was identified approximately by noting where the CAMB nonlinear spectrum began to deviate from the linear spectrum. In practice the cutoff value was taken to be  $k_{NL} = 0.12$ . It is important to note that by restricting our attention to the linear regime in this way, we are introducing systematic error into the model. This error must be taken into account when fitting the matter data. An estimate for the systematic error was obtained by considering the quadratic contribution that was dropped from the local bias ansatz, since this is the most important correction after linear order. For the mode-by-mode fit performed according to equation (4.1.1), this systematic error is the only source of error. However when fitting the data at the level of spectra, according to equations (4.1.2) and (4.1.3), we must additionally account for the random error due to the scatter in the bins, as discussed in Section 3.4.

Having taken these considerations into account, the data sets were fit for the linear bias parameter. The results are presented in Table 2. The error in the bias parameters was taken to be the standard deviation returned by the fitting routine. We compute a mean value of  $1.67 \pm 0.02$ , where the errors were added in quadrature. As can be seen, the three values measured for the bias are in good agreement, but they do not all agree within error. The two values obtained from fitting the spectra are in agreement within error, but the value obtained using the mode by mode fit is somewhat discrepant. Perhaps the simplest explanation for this discrepancy is that we have underestimated the error in the model. It is expected that fitting the data at the level of the density contrast would result in smaller error compared to a fit at the level of the spectra, but the estimated values of the bias should be in agreement regardless. As mentioned above, the error in the model was estimated solely from the quadratic contribution in the local bias ansatz. It is possible that consideration of higher orders in the expansion would reduce the

discrepancy between the value obtained from the fit of the density contrast and the values obtained from the fits of the spectra. It isn't entirely clear that this would resolve the issue however, since the inclusion of orders cubic or higher would only provide a small correction. Another possibility regarding why the value obtained from the mode by mode fit is discrepant from the other two values is that some effects related to the phase of the field might be causing the bias to be overestimated. Such phase effects would be multiplied out in the case of the spectra. It isn't immediately clear how such an overestimation of the bias would happen however. Due to the lack of statistics it is difficult to say if the discrepancies observed here are inherent to the measurement and modelling procedures, or if they are simply an artefact of the particular seed we used to run this simulation. An investigation into this query could be undertaken in future work by running a number of simulations with identical parameters, but with different primordial seeds. The resulting improvement in statistics would surely provide additional information about the behaviour of the bias, as well as allowing for a more accurate calculation of the mean. Moreover this could be supplemented with an estimation of the cubic contribution to the systematic error, as well as examination of the possibility of phase effects on the bias, in hopes of arriving at some understanding of the discrepancy between the results of the mode by mode fit and the fit of the spectra.

Fitting Method	Linear Bias Parameter ( $b_1$ )
$\delta_h(\vec{k})$	$1.695 \pm 0.002$
$P_{hh}(k)$	$1.65 \pm 0.01$
$P_{mh}(k)$	$1.67 \pm 0.02$

Table 2: Measured linear bias parameters using data from the Gaussian ( $f_{NL} = 0$ ) simulation performed in a box size of  $800h^{-1}$  Mpc and  $410^3$  particles.

Having verified that the linear halo bias is a reasonable approximation on large scales, and having made measurements of the linear bias parameter, we now turn our attention to the simulations with non-Gaussian initial conditions in order to measure the effects of scale-dependent bias.

## 4.2 Measuring the Scale-Dependent Bias

In Section 2.5 we demonstrated that the presence of local non-Gaussianities in the primordial potential leads to quantifiable effects in the late-time biasing of dark matter halos on large scales. Furthermore, it was shown that these corrections to the linear bias relation were scale-dependent. The modified biasing relation between the halo density contrast and the matter density contrast is written as

$$\delta_h(\vec{k}) = [b_1 + \Delta b_1(k)] \delta(\vec{k}) \quad (4.2.1)$$

where

$$\Delta b_1(k) \equiv \frac{10\Omega_m H_0^2}{3k^2 T(k) D(a)} f_{NL} \frac{\partial \log n_h}{\partial \log \sigma_8} . \quad (4.2.2)$$

Thus, the details of the scale-dependent bias are contingent on the history of the matter density contrast, as captured by the transfer function,  $T(k)$ , and the growth function,  $D(a)$ . The corresponding expressions for the spectra are, naturally,

$$P_{mh}(k) = [b_1 + \Delta b_1(k)] P_{mm}(k) \quad (4.2.3)$$

and

$$P_{hh}(k) = [b_1 + \Delta b_1(k)]^2 P_{mm}(k) . \quad (4.2.4)$$

As with linear biasing, we are thus equipped with three ways of making measurements the scale-dependent bias. In practice, when fitting for this effect, it isn't necessary to include all the numerical factors from equation (4.2.2) explicitly. The scale-dependent bias,  $\Delta b_1(k)$ , can simply be written as

$$\Delta b_1(k) = \frac{c_1}{k^2 T(k)} \quad (4.2.5)$$

where  $c_1$  is a numerical parameter. Since we are measuring the bias at a specific epoch, the growth factor takes on a numerical value. The only functional dependence is therefore the scale dependence. Using this expression, the aim of measuring the scale-dependent bias is to measure the parameters  $b_1$  and  $c_1$ . As in the case of linear biasing, these bias parameters can be measured by fitting the matter data according to equations (4.2.1), (4.2.3) and (4.2.4). The procedure for accomplishing this is almost identical to the one described in Section 4.1. As before, the data is truncated at a scale of  $k_{NL} = 0.12$ , and the errors are estimated in the same way. The one difference is that additional data is required in order to include the transfer function in the scale-dependent bias. This was done by using the numerical transfer function from the CAMB Web Interface used to generate the analytic power spectra [44][45]. The measured values of  $b_1$  and  $c_1$  for the different fits are presented in Figure 3 and Figure 5, respectively. The raw data for these measurements is available in Appendix A.

From Figure 3 we see that a wide range of values were measured for the linear bias parameter,  $b_1$ . These values are compared to the mean value of  $b_1 = 1.67 \pm 0.02$  from the linear bias fit described in the previous section. The values of  $b_1$  measured using the spectra agree with this mean value for the most part, the exception being the value obtained from the cross spectrum fit for  $f_{NL} = 200$ . The values measured from the overdensities are highly discrepant however, though there seems to be a trend towards agreement for large positive values of  $f_{NL}$ . Additionally, the hierarchy of values in terms of fitting methods is consistent with that observed in the linear bias fit, with the overdensity fits resulting in the largest values, and the halo spectrum fits resulting in the smallest. An important feature to notice is that  $b_1$  appears to have some functional dependence on  $f_{NL}$ , which is not expected theoretically. Moreover, the functional dependence appears to be qualitatively different for the values obtained by fitting the density contrast and the values obtained by fitting the spectra. In the former case, the values of  $b_1$  follow a decreasing trend as  $f_{NL}$  increases. In the case of the spectra however, the fitted values of the linear bias parameter appear to peak around  $f_{NL} = 0$  and decrease as  $|f_{NL}|$  increases. The point of symmetry however is not exactly at  $f_{NL} = 0$ , as one might expect, but rather is shifted towards  $f_{NL} > 0$ . This discrepancy in the  $f_{NL}$ -dependence of the data obtained from the fit of the overdensity and the fit of the spectra potentially arises due to the binning in the latter case. In particular the first

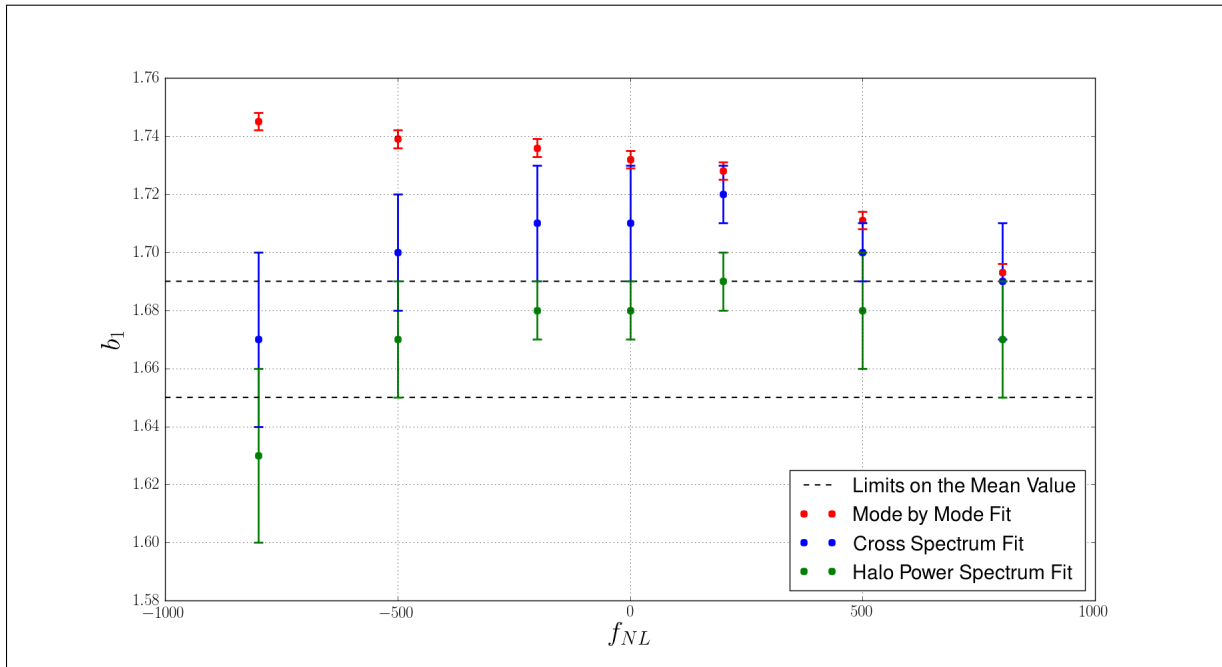


Figure 3: Measured values of the bias parameter  $b_1$  using data from simulations performed in a box size of  $800h^{-1}$  Mpc with  $410^3$  particles. The dashed lines are the limits on the mean value of  $b_1 = 1.67$  obtained from the Gaussian linear bias measurement in Section 4.1.

bin of the spectrum data, with the lowest value of  $k$ , has a strong susceptibility to the value of  $f_{NL}$ . This is shown for the halo power spectrum data in Figure 4. For each simulation this first bin contains the least amount of modes, as discussed in Section 4.1. This means that a small sample of the data is having a large influence on the fitting procedure. This is especially true when the power in that first bin is considerably different than the power in the next lowest bin, as in the case of  $f_{NL} = 800$  and  $f_{NL} = -800$ . This issue does not occur in the case of the density contrast, where all modes are fit simultaneously. However, even though the first bin of the spectra is associated with a limited number of modes, this should be taken into account in the error estimates, both in the systematic error of the model and in the expected random error for that bin. This bin should therefore be devalued considerably in the fit. If the hypothesis presented here is true, then it points to an underestimation of the error. Though these considerations might explain the functional dependence of the bias measured from the spectra, they do not explain the trend in the values of the bias measured from the overdensity. A potential explanation presents itself however, in the discussion of the  $c_1$  data.

The estimated values of the scale-dependent bias parameter,  $c_1$ , are presented in Figure 5, where they are plotted against the different values of  $f_{NL}$ . From this plot, we see that the data is described quite well by a linear relation between  $c_1$  and  $f_{NL}$ . This is the expected relation based on equation (4.2.2). The plot includes the line of best fit through the data, with an estimated slope of  $1.83 \pm 0.03$  and an intercept of  $-280 \pm 20$ . Though the linear relation is correct, the value of the intercept is highly discrepant. From equations (4.2.1) and (4.2.2), the intercept is expected to be identically equal to zero, since the bias must reduce to the Gaussian case when  $f_{NL} = 0$ . This suggests that the bias is being

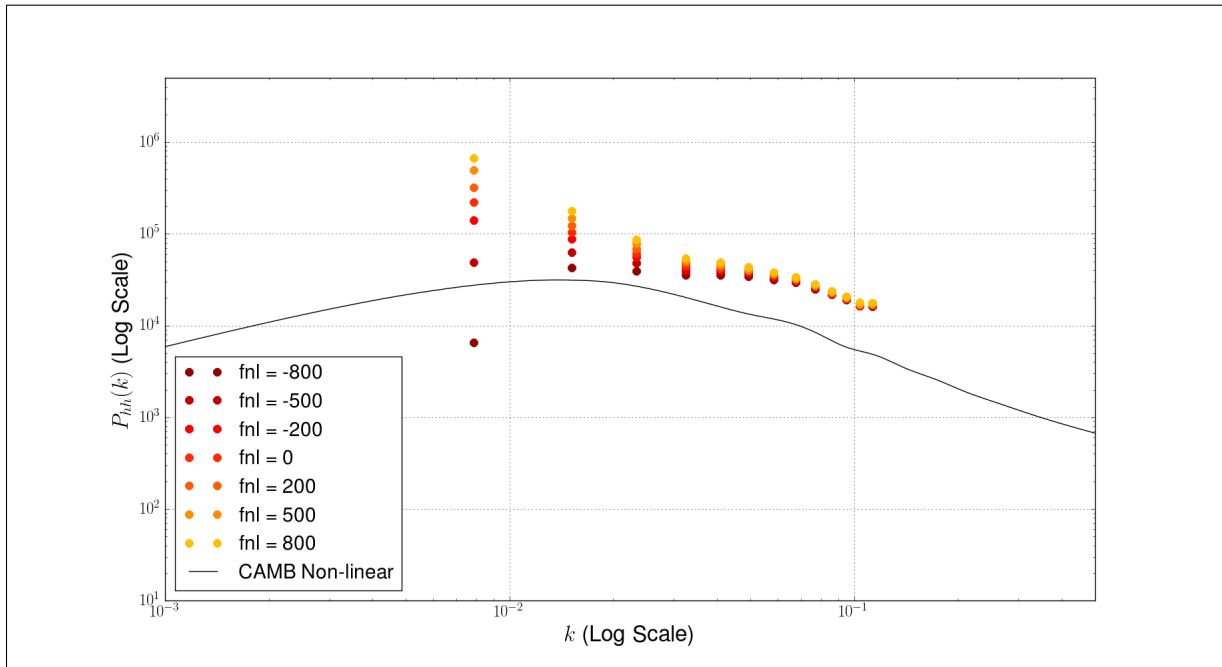


Figure 4: The ROCKSTAR halo power spectrum for various values of  $f_{NL}$ . The data has been truncated at  $k_{NL} = 0.12$ . The error bars have been omitted for clarity.

systematically underestimated in some way. We can express this as follows. The halo bias is

$$\frac{\delta_h}{\delta} = b_1 + \frac{c_1}{k^2 T(k)} . \quad (4.2.6)$$

If we write the scale-dependent parameter as  $c_1 = A f_{NL} + B$ , where  $A$  and  $B$  are the measured slope and intercept, we have

$$\frac{\delta_h}{\delta} = b_1 + \frac{A f_{NL}}{k^2 T(k)} + \frac{B}{k^2 T(k)} . \quad (4.2.7)$$

From the model that we developed in Section 2.5 we expect  $B = 0$ . The fact that we measure  $B \neq 0$  suggests that the assumptions we made when developing the expression for scale-dependent bias were not entirely justified. These assumptions are thus introducing error into the data and this error hasn't been accounted for. An important assumption that we made in Section 2.5 when developing the model of scale-dependent bias was the validity of dropping the quadratic contribution of the small-scale potential,  $\phi_{p,s}$ , in equation (2.5.5). This was done because we were primarily interested in how small- and large-scale modes are coupled by non-Gaussianity. It is reasonable to expect that this would introduce error into the model, especially since the systematic error in the linear bias was estimated by considering the quadratic contribution to the local bias ansatz in equation (2.3.3). An accurate estimate of the systematics could therefore be obtained by taking into account this quadratic contribution to the scale-dependent bias. In addition to throwing off the measured values of  $c_1$ , such systematic errors in the model of scale-dependent bias could also be inducing errors in the measured values of  $b_1$ . Since the values of  $c_1$  appear to have been systematically underestimated, it is possible that they have been compensated

for by  $b_1$ , resulting in the discrepantly large values seen in Figure 3 for the mode by mode fits. It isn't obvious how a better estimation of the error would specifically affect the measured values of the slope  $A$  and the intercept  $B$ , but it is reasonable to assume that it would have an effect.

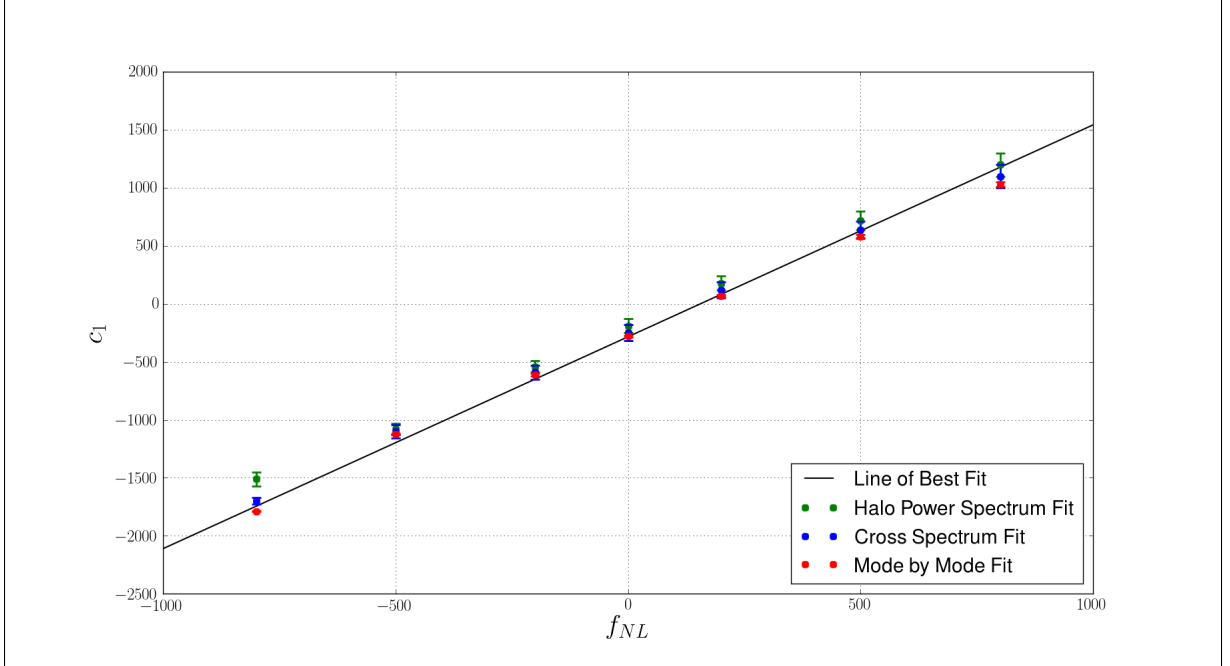


Figure 5: Measured values of the scale-dependent bias parameter  $c_1$  using data from simulations performed in a box size of  $800h^{-1}$  Mpc and  $410^3$  particles.

Overall, the measurement of scale-dependent bias has resulted in fairly reasonable results. The estimated linear bias parameters,  $b_1$ , for the various degrees of non-Gaussianity are in agreement with the mean value obtained in Section 4.1, at least for the fits performed on the spectra. Of course this is not entirely rigorous since the data used to compute the mean suffer from their own discrepancies, but it works as a present estimate. The most encouraging result presented here is the linear trend observed in the scale-dependent bias,  $c_1$ , when plotted with respect to  $f_{NL}$ . This suggests that the model of scale-dependent bias developed in Section 2.5 is an accurate description of how primordial non-Gaussianities induce coupling between small and large scales. Some refinements of the model are necessary, as described in this section, but the model itself is fundamentally correct. An accurate estimation of the systematics in the model has the potential of resulting in considerable improvement. This is a clear next step to be undertaken in future work.

## 5 Conclusion

Over the course of this report we have developed the tools required to understand and make numerical measurements of the scale-dependent bias of dark matter halos. This is an important phenomenon to study since it allows us to use LSS as a means of constraining the degree of non-Gaussianity in the primordial power spectrum [3]. With a number of LSS studies planned for the coming

decade [10][13][14][15][16], scale-dependent bias has the potential of drastically improving the constraints on primordial non-Gaussianities compared to those obtained from measurements of the CMB [10]. It is thus beneficial to have an understanding of the physics underpinning this effect. It is equally important to have an understanding of the methodology used to make measurements of this phenomenon.

We developed the theoretical aspects of scale-dependent bias by beginning with a review of modern cosmology (Section 2). We described the main results of the homogeneous and isotropic FLRW universe, as well as the linear cosmological perturbation theory for cold dark matter and gravitation. We then introduced the idea of halo bias, as an alternate means of describing the non-linear aspects of LSS. The statistical nature of cosmological observations was then discussed, with an introduction to the power spectrum. This was followed by a brief discussion of the initial conditions of the universe and the formulation of non-Gaussianity. Finally, we developed the model describing the scale-dependent bias. Once the main theoretical considerations were covered, we presented the details of the numerical tools used to make measurements of the effect (Section 3). These measurements were made using data from cosmological  $N$ -body simulations performed with the GADGET-2 code, using initial conditions generated by the 2LPT initial conditions code. The dark matter halos were identified in the output of the simulations using the ROCKSTAR halo finder. Furthermore, the power spectra were computed from the simulation data with in-house programs written in Python.

The results were presented in two parts (Section 4). In the first part we presented measurements of the linear halo bias using data from a Gaussian simulation. The bias was measured by fitting the matter data to the halo data. Three measurements were made using three different fits: a fit of the matter overdensity to the halo overdensity, a fit of the matter power to the matter-halo cross spectrum, and a fit of the matter power to the halo power. The mean value of the bias obtained from these fits was  $b_1 = 1.67 \pm 0.02$ . The individual measurements were in good agreement, though the measurement obtained from the overdensity fit was somewhat discrepant. This discrepancy could simply be an artefact of the given seed used to generate the primordial power, or it could indicate an underestimation of the systematic error in the model used. Additionally it is possible that it is due to some phase effects that are not present in the spectrum fits. The second part of the results presented and analyzed the measurements of scale-dependent bias. It was found that the scale-dependent bias parameter and the values of  $f_{NL}$  obey a linear relationship with an estimated slope of  $1.83 \pm 0.03$  and an intercept of  $-280 \pm 20$ . Though the value of the intercept is incorrect, the linear relationship suggests that the underlying model used to describe scale-dependent bias is correct. Furthermore, the discrepancy in the intercept can be explained as the result of an underestimation of the systematic error in the model. This additional error could potentially be due to fact that the quadratic contribution to the non-Gaussian potential was neglected when developing the model. This remains to be seen, however, and it provides a natural line of inquiry for future work. Properly accounting for the systematics in the model also has the potential of resolving some discrepancies in the linear bias parameters measured using the non-Gaussian data.

The obvious next step in the research is to resolve the discrepancies in the data presented here. The various lines of inquiry that we have mentioned could be pursued readily. An additional step is to improve the statistics of the data by running a number of simulations with different primordial seeds. The large amount of data resulting from this would provide important information about the behaviour of the bias parameters. One could even endeavour to measure the bias parameters by fitting data from multiple simulations simultaneously. If the computational and temporal resources are available, it would

also be beneficial to run even larger simulations so that more data can be acquired in the linear regime of large scales where the scale-dependent bias is expected to have its most prominent effect. Finally, higher-order correlation functions of the density fields, such as the bispectrum, could be measured in addition to the power spectrum, since these are also expected to be susceptible to non-Gaussianities [5][6][10][46][47].

In closing, the scale-dependent bias of dark matter halos is a promising method of detecting primordial non-Gaussianities and it will serve as one of the most important measurements used to further our understanding of the early universe in the coming era of precision LSS surveys.



## Appendix A Raw Measurements of Scale-Dependent Bias

$f_{NL}$	Fitting Method	Bias Parameter: $b_1$	Bias Parameter: $c_1$
-800	$\delta_h(\vec{k})$	$1.745 \pm 0.003$	$-1788 \pm 5$
	$P_{hh}(k)$	$1.63 \pm 0.03$	$-1510 \pm 60$
	$P_{mh}(k)$	$1.67 \pm 0.03$	$-1700 \pm 30$
-500	$\delta_h(\vec{k})$	$1.739 \pm 0.003$	$-1123 \pm 7$
	$P_{hh}(k)$	$1.67 \pm 0.02$	$-1080 \pm 50$
	$P_{mh}(k)$	$1.70 \pm 0.02$	$-1100 \pm 60$
-200	$\delta_h(\vec{k})$	$1.736 \pm 0.003$	$-610 \pm 10$
	$P_{hh}(k)$	$1.68 \pm 0.01$	$-540 \pm 50$
	$P_{mh}(k)$	$1.71 \pm 0.02$	$-590 \pm 60$
0	$\delta_h(\vec{k})$	$1.732 \pm 0.003$	$-280 \pm 10$
	$P_{hh}(k)$	$1.68 \pm 0.01$	$-190 \pm 60$
	$P_{mh}(k)$	$1.71 \pm 0.02$	$-250 \pm 70$
200	$\delta_h(\vec{k})$	$1.728 \pm 0.003$	$70 \pm 10$
	$P_{hh}(k)$	$1.69 \pm 0.01$	$180 \pm 60$
	$P_{mh}(k)$	$1.72 \pm 0.01$	$120 \pm 70$
500	$\delta_h(\vec{k})$	$1.711 \pm 0.003$	$580 \pm 20$
	$P_{hh}(k)$	$1.68 \pm 0.02$	$720 \pm 80$
	$P_{mh}(k)$	$1.70 \pm 0.02$	$640 \pm 70$
800	$\delta_h(\vec{k})$	$1.693 \pm 0.003$	$1030 \pm 20$
	$P_{hh}(k)$	$1.67 \pm 0.02$	$1200 \pm 100$
	$P_{mh}(k)$	$1.69 \pm 0.02$	$1100 \pm 100$

Table A1: Measurements of the scale-dependent bias parameters using data from simulations performed with a range of  $f_{NL}$  values in a box size of  $800h^{-1}$  Mpc and  $410^3$  particles.

## References

- [1] P. Schneider, *Extragalactic Astronomy and Cosmology: An Introduction*. Springer, second ed., 2015.
- [2] P. Peebles, *Principles of Physical Cosmology*. Princeton University Press, 1971.
- [3] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, “The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects,” *Phys. Rev.* **D77** (2008) 123514, [arXiv:0710.4560 \[astro-ph\]](#).
- [4] L. Verde, R. Jimenez, M. Kamionkowski, and S. Matarrese, “Tests for primordial nonGaussianity,” *Mon. Not. Roy. Astron. Soc.* **325** (2001) 412, [arXiv:astro-ph/0011180 \[astro-ph\]](#).
- [5] R. Scoccimarro, E. Sefusatti, and M. Zaldarriaga, “Probing primordial non-Gaussianity with large - scale structure,” *Phys. Rev.* **D69** (2004) 103513, [arXiv:astro-ph/0312286 \[astro-ph\]](#).
- [6] E. Sefusatti and E. Komatsu, “The bispectrum of galaxies from high-redshift galaxy surveys: Primordial non-Gaussianity and non-linear galaxy bias,” *Phys. Rev.* **D76** (2007) 083004, [arXiv:0705.0343 \[astro-ph\]](#).
- [7] J. Robinson and J. E. Baker, “Evolution of the cluster abundance in nonGaussian models,” *Mon. Not. Roy. Astron. Soc.* **311** (2000) 781, [arXiv:astro-ph/9905098 \[astro-ph\]](#).
- [8] S. Matarrese, L. Verde, and R. Jimenez, “The Abundance of high-redshift objects as a probe of non-Gaussian initial conditions,” *Astrophys. J.* **541** (2000) 10, [arXiv:astro-ph/0001366 \[astro-ph\]](#).
- [9] **WMAP** Collaboration, E. Komatsu *et al.*, “First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: tests of gaussianity,” *Astrophys. J. Suppl.* **148** (2003) 119–134, [arXiv:astro-ph/0302223 \[astro-ph\]](#).
- [10] M. Alvarez *et al.*, “Testing Inflation with Large Scale Structure: Connecting Hopes with Reality,” [arXiv:1412.4671 \[astro-ph.CO\]](#).
- [11] **Planck Collaboration** Collaboration, P. Ade *et al.*, “Planck 2013 Results. XXIV. Constraints on primordial non-Gaussianity,” [arXiv:1303.5084 \[astro-ph.CO\]](#).
- [12] **Planck** Collaboration, P. A. R. Ade *et al.*, “Planck 2015 results. XVII. Constraints on primordial non-Gaussianity,” [arXiv:1502.01592 \[astro-ph.CO\]](#).
- [13] “The Dark Energy Survey.” <http://www.darkenergysurvey.org/>.
- [14] Ansonika and Sowaleim, “The Large Synoptic Survey Telescope.” <http://www.lsst.org/>.
- [15] “Dark Energy Spectroscopic Instrument.” <http://desi.lbl.gov/>.
- [16] “ESA Science and Technology: Euclid.” <http://sci.esa.int/euclid/>.
- [17] S. Dodelson, *Modern Cosmology*. Academic Press, 2003.
- [18] S. Carroll, *Spacetime and Geometry*. Addison Wesley, 2004.
- [19] D. Baumann, “Cosmology.” [www.damtp.cam.ac.uk/user/db275/Cosmology/Lectures.pdf](http://www.damtp.cam.ac.uk/user/db275/Cosmology/Lectures.pdf).
- [20] E. Hubble, “A Relation Between Distance and Radial Velocity Among Extra-Galactic Nebulae,” *Proceedings of the National Academy of Sciences* **15** (1929) 168–173. <http://www.pnas.org/content/15/3/168.full>.
- [21] **Planck** Collaboration, P. A. R. Ade *et al.*, “Planck 2015 results. XIII. Cosmological parameters,” [arXiv:1502.01589 \[astro-ph.CO\]](#).
- [22] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, “Large scale structure of the universe and cosmological perturbation theory,” *Phys. Rept.* **367** (2002) 1–248, [arXiv:astro-ph/0112551 \[astro-ph\]](#).
- [23] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, “Cosmological Non-Linearities as an Effective Fluid,” *JCAP* **1207** (2012) 051, [arXiv:1004.2488 \[astro-ph.CO\]](#).
- [24] F. Bernardeau, “The evolution of the large-scale structure of the universe: beyond the linear regime,” in *100e Ecole d’Ete de Physique: Post-Planck Cosmology Les Houches, France, July 8-August 2, 2013*. 2013. [arXiv:1311.2724 \[astro-ph.CO\]](#). <http://inspirehep.net/record/1264888/files/arXiv:1311.2724.pdf>.
- [25] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, “The Effective Field Theory of Cosmological Large Scale Structures,” *JHEP* **09** (2012) 082, [arXiv:1206.2926 \[astro-ph.CO\]](#).
- [26] W. Press and P. Schechter, “Formation of galaxies and clusters of galaxies by self-similar gravitational condensation,” *Astrophys. J.* **187** (1974) 425.
- [27] S. Ferraro, K. M. Smith, D. Green, and D. Baumann, “On the Equivalence of Barrier Crossing, Peak-Background Split, and Local Biasing,” *Mon. Not. Roy. Astron. Soc.* **435** (2013) 934–942, [arXiv:1209.2175 \[astro-ph.CO\]](#).
- [28] D. Baumann, S. Ferraro, D. Green, and K. M. Smith, “Stochastic Bias from Non-Gaussian Initial Conditions,” *JCAP* **1305** (2013) 001, [arXiv:1209.2173 \[astro-ph.CO\]](#).

- [29] V. Assassi, D. Baumann, D. Green, and M. Zaldarriaga, “Renormalized Halo Bias,” *JCAP* **1408** (2014) 056, [arXiv:1402.5916 \[astro-ph.CO\]](#).
- [30] T. Baldauf, U. Seljak, V. Desjacques, and P. McDonald, “Evidence for Quadratic Tidal Tensor Bias from the Halo Bispectrum,” *Phys. Rev.* **D86** (2012) 083540, [arXiv:1201.4827 \[astro-ph.CO\]](#).
- [31] P. McDonald, “Clustering of dark matter tracers: Renormalizing the bias parameters,” *Phys. Rev.* **D74** (2006) 103512, [arXiv:astro-ph/0609413 \[astro-ph\]](#). [Erratum: *Phys. Rev.* **D74**, 129901 (2006)].
- [32] P. McDonald and A. Roy, “Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS,” *JCAP* **0908** (2009) 020, [arXiv:0902.0991 \[astro-ph.CO\]](#).
- [33] S. Weinberg, *Cosmology*. Oxford University Press, 2008.
- [34] V. Springel, N. Yoshida, and S. D. M. White, “GADGET: A Code for collisionless and gasdynamical cosmological simulations,” *New Astron.* **6** (2001) 79, [arXiv:astro-ph/0003162 \[astro-ph\]](#).
- [35] V. Springel, “The Cosmological simulation code GADGET-2,” *Mon. Not. Roy. Astron. Soc.* **364** (2005) 1105–1134, [arXiv:astro-ph/0505010 \[astro-ph\]](#).
- [36] R. Scoccimarro, L. Hui, M. Manera, and K. C. Chan, “Large-scale Bias and Efficient Generation of Initial Conditions for Non-Local Primordial Non-Gaussianity,” *Phys. Rev.* **D85** (2012) 083002, [arXiv:1108.5512 \[astro-ph.CO\]](#).
- [37] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, “The Rockstar Phase-Space Temporal Halo Finder and the Velocity Offsets of Cluster Cores,” *Astrophys. J.* **762** (2013) 109, [arXiv:1110.4372 \[astro-ph.CO\]](#).
- [38] D. Jeong, *Cosmology with high ( $z \gtrsim 1$ ) redshift galaxy surveys*. PhD thesis, The University of Texas at Austin, August, 2010. [www.personal.psu.edu/duj13/](http://www.personal.psu.edu/duj13/).
- [39] J. Barnes and P. Hut, “A Hierarchical O Nlogn Force Calculation Algorithm,” *Nature* **324** (1986) 446–449.
- [40] M. Crocce, S. Pueblas, and R. Scoccimarro, “Transients from Initial Conditions in Cosmological Simulations,” *Mon. Not. Roy. Astron. Soc.* **373** (2006) 369–381, [arXiv:astro-ph/0606505 \[astro-ph\]](#).
- [41] R. Scoccimarro, “Transients from initial conditions: a perturbative analysis,” *Mon. Not. Roy. Astron. Soc.* **299** (1998) 1097, [arXiv:astro-ph/9711187 \[astro-ph\]](#).
- [42] C. Wagner and L. Verde, “N-body simulations with generic non-Gaussian initial conditions II: Halo bias,” *JCAP* **1203** (2012) 002, [arXiv:1102.3229 \[astro-ph.CO\]](#).
- [43] S. Shandera, N. Dalal, and D. Huterer, “A generalized local ansatz and its effect on halo bias,” *JCAP* **1103** (2011) 017, [arXiv:1010.3722 \[astro-ph.CO\]](#).
- [44] “LAMBDA - CAMB Web Interface.” [http://lambda.gsfc.nasa.gov/toolbox/tb\\_camb\\_form.cfm](http://lambda.gsfc.nasa.gov/toolbox/tb_camb_form.cfm).
- [45] A. Lewis, “CAMB Notes.” <http://cosmologist.info/notes/CAMB.pdf>.
- [46] T. Baldauf, U. Seljak, and L. Senatore, “Primordial non-Gaussianity in the Bispectrum of the Halo Density Field,” *JCAP* **1104** (2011) 006, [arXiv:1011.1513 \[astro-ph.CO\]](#).
- [47] L. Verde, L.-M. Wang, A. Heavens, and M. Kamionkowski, “Large scale structure, the cosmic microwave background, and primordial non-gaussianity,” *Mon. Not. Roy. Astron. Soc.* **313** (2000) L141–L147, [arXiv:astro-ph/9906301 \[astro-ph\]](#).