An Improved Calculation of the Non-Gaussian Halo Mass Function

Guido D’Amico\textsuperscript{a,b}, Marcello Musso\textsuperscript{c}, Jorge Noreña\textsuperscript{a,b}, Aseem Paranjape\textsuperscript{c}

\textsuperscript{a} SISSA, via Bonomea 265, 34136 Trieste, Italy
\textsuperscript{b} INFN - Sezione di Trieste, via Bonomea 265, 34136 Trieste, Italy
\textsuperscript{c} Abdus Salam International Centre for Theoretical Physics
Strada Costiera 11, 34151, Trieste, Italy

Abstract
The abundance of collapsed objects in the universe, or halo mass function, is an important theoretical tool in studying the effects of primordially generated non-Gaussianities on the large scale structure. The non-Gaussian mass function has been calculated by several authors in different ways, typically by exploiting the smallness of certain parameters which naturally appear in the calculation, to set up a perturbative expansion. We improve upon the existing results for the mass function by combining path integral methods and saddle point techniques (which have been separately applied in previous approaches). Additionally, we carefully account for the various scale dependent combinations of small parameters which appear. Some of these combinations in fact become of order unity for large mass scales and at high redshifts, and must therefore be treated non-perturbatively. Our approach allows us to do this, and also to account for multi-scale density correlations which appear in the calculation. We thus derive an accurate expression for the mass function which is based on approximations that are valid over a larger range of mass scales and redshifts than those of other authors. By tracking the terms ignored in the analysis, we estimate theoretical errors for our result and also for the results of others. We also discuss the complications introduced by the choice of smoothing filter function, which we take to be a top-hat in real space, and which leads to the dominant errors in our expression. Finally, we present a detailed comparison between the various expressions for the mass functions, exploring the accuracy and range of validity of each.

1 Introduction
The primordial curvature inhomogeneities, generated by the inflationary mechanism, obey nearly Gaussian statistics. The deviations from Gaussianity, while expected to be small, provide a unique window into the physics of inflation. For example, single-field slow-roll models of inflation lead to a small level of non-Gaussianity (NG), so that an observation of a large NG would indicate a deviation from this paradigm.

Until a few years ago, the main tool to constrain NG was considered to be the statistics of the cosmic microwave background (CMB) temperature field, since inhomogeneities at the CMB epoch are small and their physics can be described by a perturbative treatment. In recent years, however, thanks to observations and developments in the theory, the large-scale structure (LSS) of the universe has emerged as a complementary probe to constrain primordial NG. While it is true that the \(n\)-point functions of the density field on small scales are dominated by the recent gravitational evolution, and do not reflect anymore the statistics of primordial perturbations, it turns out that the abundance of very massive objects, which form out of high peaks of the density perturbations, is a powerful probe of primordial NG. In this context, much attention has been given recently to three possible methods of constraining the magnitude and shape of the primordial NG with the LSS: the galaxy power spectrum, the galaxy bispectrum and the mass function. It was pointed out in Refs. [1, 2] that a NG of a local type induces a scale dependence on the galaxy power spectrum, thus making it a sensitive probe of the magnitude of local NG \(f_{NL}^{loc}\). From Ref. [3] one finds the following constraints: \(-29 < f_{NL}^{loc} < +69\).
already comparable with those obtained from CMB measurements in Ref. [4]: $-10 < f_{NL}^{loc} < +74$. The future is even more promising, with precisions of $\Delta f_{NL}^{loc} \sim 10$ [5] and $\Delta f_{NL}^{loc} \sim 1$ [6, 7, 8] being claimed for future surveys. The galaxy bispectrum is also a promising probe of NG as it could be more sensitive to other triangle configurations [9]. The mass function – which is the focus of this work, and which we discuss in detail below – has been used for example in Ref. [5] together with the scale dependent bias to produce forecasts for future surveys, and in Ref. [10] in an attempt to explain the presence of a very massive cluster at a large redshift as an indication of a large NG. For more references and information we refer the reader to reviews summarizing recent results on these topics [11, 12].

The formation of bound dark matter halos from initially small density perturbations, as seen in numerical simulations, is a complicated and violent process. Some insight into the physics involved has been gained from the study of analytical models. The quantity of interest is the halo mass function, defined as the number density of dark matter halos with a mass between $M$ and $M + dM$,

$$\frac{dn}{dM} = \frac{\bar{\rho}}{M^2} f(\sigma) \left| \frac{d\ln \sigma}{d\ln M} \right| ,$$

where $\bar{\rho}$ is the average density of the universe, $\sigma(M)$ is the variance of the density contrast $\delta_R$ filtered on some comoving scale $R$ corresponding to the mass $M$, and the function $f(\sigma)$ is to be computed. Throughout this work, we will refer to $f(\sigma)$ itself as the mass function. A very useful tool in the analysis is the spherical collapse model [13], which predicts that the value of the linearly extrapolated density contrast of a spherical halo, at the time when the halo collapses, is $\delta_c \approx 1.686$, with a weak cosmology dependence. This value serves as a collapse threshold for determining which inhomogeneous regions will end up as collapsed objects. Using this idea, Press & Schechter [14] (PS) first computed the mass function $f(\sigma)$ in the case of Gaussian initial conditions. Their calculation however suffered from a problem of undercounting which affects the overall normalization – their approach does not count underdense regions embedded in larger overdense regions as eventually collapsed objects. To account for this discrepancy, PS introduced an ad-hoc factor of 2 by demanding that the mass function be correctly normalized, such that all the mass in the universe must be contained in collapsed objects. In the excursion set approach, Bond et al. [15] resolved this issue and derived a correctly normalized mass function, for Gaussian initial conditions. They argued that the filtered density contrast $\delta_R$ follows a random walk as a function of the filtering scale, and the problem of computing $f(\sigma)$ is translated into the problem of finding the rate of “first crossing” of the barrier $\delta_c$, whose solution is well-known. We will study this formalism in detail in section 3 for the more general non-Gaussian case.

Turning to non-Gaussianities, the most popular non-Gaussian mass functions are those due to Matarrese, Verde and Jimenez [16] (MVJ) and LoVerde et al. [17] (LMSV). Both groups used the PS approach, by modifying the probability density function for the (linearized) density contrast to describe non-Gaussian initial conditions. In their prescription, the relevant object is the ratio $R_{ng}$ of non-Gaussian to Gaussian mass functions. The full mass function is usually taken as the product of $R_{ng}$ and an appropriate Gaussian mass function as given by N-body simulations, e.g. the Sheth & Tormen mass function [18]. It is not clear however that this is the correct way to proceed. Indeed, in a series of papers [19, 20, 21], Maggiore & Riotto (MR) presented a rigorous approach to the first-passage problem in terms of path integrals, and in Ref. [21] they pointed out that a PS-like prescription in fact misses some important non-Gaussian effects stemming from 3-point correlations between different scales (so-called “unequal time” correlators).

On the other hand, MR treated non-Gaussian contributions to $f(\sigma)$ by simply linearizing in the 3-point function of $\delta_R$, i.e. by linearizing in the non-Gaussian parameter $f_{NL}$. Since the NG are assumed to be small, in the sense that the parameter $\epsilon = \langle \delta^3 \rangle / \sigma^3$ satisfies $\epsilon \ll 1$, one might expect that such a perturbative treatment is valid. However, another crucial ingredient in the problem is that the length scales of interest are large, which leads to a second small parameter $\nu^{-1}$ where $\nu = \delta_c / \sigma$. This is evident in the calculations of MR, who crucially use $\nu^{-2} \propto \sigma^2$ as a small parameter. Any perturbative treatment now depends not only on the smallness of $\epsilon$ and $\nu^{-1}$ individually, but also on the specific combinations of these parameters which appear in the calculation. It is known (and we will explicitly see below) that a natural combination that appears is $\epsilon \nu^3$, which can become of order unity on scales of interest. The mass functions given by LMSV and MR therefore break down as valid series expansions when this occurs. Interestingly, MVJ’s PS-like treatment on the other hand involved a saddle point approximation, allowing them to non-perturbatively account for the
\( \epsilon \nu^3 \) term (which appears in an exponential in their approach). For a discussion, see Ref. [22].

It appears to us therefore, that there is considerable room for improvement in the theoretical calculation of the mass function. The goal of our paper is twofold. Firstly, we present a rigorous calculation of the mass function in the following way: (a) we use the techniques developed by MR in Refs. [19, 20, 21], which allow us to track the complex multi-scale correlations involved in the calculation, and (b) we demonstrate that MR’s approach can be combined with saddle point techniques (used by MVJ), to non-perturbatively handle terms which can become of order unity. This leads to an expression for the mass function which is valid on much larger scales than those presented by MR and LMSV. Secondly, by keeping track of the terms ignored, we calculate theoretical error bars on the expressions for \( f(\sigma) \) resulting not only from our own calculations, but also for those of the other authors [16, 17, 21]. Since the terms ignored depend on \( \nu \) in general, these error bars are clearly scale dependent. This allows us to estimate the validity of each of the expressions for the mass function at different scales, but importantly it also allows us to analytically compare between different expressions. In this paper we will not explicitly account for effects of the ellipsoidal collapse model [23, 24], since these are expected to be negligible on the very large scales which are of interest to us. For a recent treatment of ellipsoidal collapse effects in the presence of non-Gaussianities on scales where \( \epsilon \nu^3 \ll 1 \), see Lam & Sheth [25]. For a different approach to computing the non-Gaussian halo mass function, see Ref. [1], where the authors proposed that this mass function can be approximated as a convolution of the Gaussian mass function with a probability distribution function that maps between halos identified in Gaussian and non-Gaussian N-body simulations. This probability distribution itself was approximated as a Gaussian, with mean and variance fit from simulations. While in practice this approach is easy to implement, it relies heavily on the output of N-body simulations. Our approach, on the other hand, allows us to compute a mass function almost entirely from first principles.

This paper is organized as follows. In section 2 we fix some notation and briefly introduce the two most popular shapes of primordial NG, i.e. the local and equilateral ones. In section 3 we present our calculation of the mass function. In section 4 we discuss certain subtleties regarding the truncation of the perturbative series, and also compare with the other expressions for \( f(\sigma) \) mentioned above. In section 5 we discuss the effects induced by some additional complications introduced in the problem due to the specific choice of the filter function [19], which we take to be a top-hat in real space, and due to the inclusion of stochasticity in the value of the collapse threshold \( \delta_c \) (which is also expected to partially account for effects of ellipsoidal collapse) [20]. In section 6 we compare our final result Eqn. (43) with those of other authors, including theoretical errors for each, and conclude with a brief discussion of the results and directions for future work. Some technical asides have been relegated to the Appendices.

## 2 Models of non-Gaussianity

We need to relate the linearly evolved density field to the primordial curvature perturbation, which carries the information of the non-linearities produced during and after inflation. We start from the Bardeen potential \( \Phi \) on subhorizon scales, given by

\[
\Phi(k, z) = -\frac{3}{5} T(k) \frac{D(z)}{a} R(k) ,
\]

where \( R(k) \) is the (comoving) curvature perturbation, which stays constant on superhorizon scales; \( T(k) \) is the transfer function of perturbations, normalized to unity as \( k \to 0 \), which describes the suppression of power for modes that entered the horizon before the matter-radiation equality; and \( D(z) \) is the linear growth factor of density fluctuations, normalized such that \( D(z) = (1 + z)^{-1} \) in the matter dominated era. Then, the density contrast field is related to the potential by the Poisson equation, which in Fourier space reads

\[
\delta(k, z) = -\frac{2ak^2}{3\Omega_m H_0^2} \Phi(k, z) = \frac{2k^2}{3\Omega_m H_0^2} T(k) D(z) R(k)
\]

\[
\equiv \mathcal{M}(k, z) R(k) ,
\]

where we substituted Eqn. (2). Here, \( \Omega_m \) is the present time fractional density of matter (cold dark matter and baryons), and \( H_0 = 100h \text{ km s}^{-1}\text{Mpc}^{-1} \) is the present time Hubble constant. The redshift dependence...
is trivially accounted for by the linear growth factor $D(z)$ and in the following, for notational simplicity, we will often suppress it. All our calculations will use a reference ΛCDM cosmology compatible with WMAP7 data [4], using parameters $h = 0.702$, $\Omega_m = 0.272$, present baryon density $\Omega_b = 0.0455$, scalar spectral index $n_s = 0.961$ and $\sigma_8 = 0.809$, where $\sigma_8^2$ is the variance of the density field smoothed on a length scale of $8h^{-1}\text{Mpc}$. For simplicity, for the transfer function $T(k)$ we use the BBKS form, proposed in Bardeen et al. [23]:

$$T_{\text{BBKS}}(x) = \frac{1}{2.34x} \ln (1 + 2.34x) (1 + 3.89x + (16.1x)^2 + (5.46x)^3 + (6.71x)^4)^{-1/4},$$

(4)

where $x \equiv k(h\text{Mpc}^{-1})/\Gamma$ with a shape parameter $\Gamma = \Omega_m h \exp [-\Omega_b(1 + \sqrt{2h}/\Omega_m)]$ that accounts for baryonic effects as described in Ref. [26]. For more accurate results, one could use a numerical transfer function, as obtained by codes like CMBFAST [27] or CAMB [28]; the results are not expected to be qualitatively different.

In order to study halos, which form where an extended region of space has an average overdensity which is above threshold, it is useful to introduce a filter function $W_R(|\mathbf{x}|)$, and consider the smoothed density field (around one point, which we take as the origin),

$$\delta_R = \int \frac{d^3k}{(2\pi)^3} \tilde{W}(kR)\delta(\mathbf{k}),$$

(5)

where $\tilde{W}(kR)$ is the Fourier transform of the filter function. For all numerical calculations we will use the spherical top-hat filter in real space, whose Fourier transform $\tilde{W}(kR)$ is given by

$$\tilde{W}(y) = \frac{3}{y^3} (\sin y - y \cos y).$$

(6)

This choice allows us to have a well-defined relation between length scales and masses, namely $M = (4\pi/3)\Omega_m \rho_c R^3$ with $\rho_c = 3H_0^2/(8\pi G) = 2.75 \cdot 10^{11} h^{-1} M_{\odot}(h^{-1}\text{Mpc})^{-3}$, however it introduces some complexities in the analysis, which we will comment on later. By using Eqns. (5) and (3) we have, for the 3-point function,

$$\langle \delta_R, \delta_R, \delta_R \rangle_c = \int \frac{d^3k_1}{(2\pi)^3} \frac{d^3k_2}{(2\pi)^3} \frac{d^3k_3}{(2\pi)^3} \tilde{W}(k_1R_1)\tilde{W}(k_2R_2)\tilde{W}(k_3R_3)\mathcal{M}(k_1)\mathcal{M}(k_2)\mathcal{M}(k_3)\langle \mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\mathcal{R}(\mathbf{k}_3) \rangle_c,$$

(7)

where the subscript $c$ denotes the connected part, and analogous formulae are valid for the higher order correlations.

### 2.1 Shapes of non-Gaussianity

The function $\langle \mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\mathcal{R}(\mathbf{k}_3) \rangle_c$ encodes information about the physics of the inflationary epoch. By translational invariance, it is proportional to a momentum-conserving delta function:

$$\langle \mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\mathcal{R}(\mathbf{k}_3) \rangle_c = (2\pi)^3\delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)B_\mathcal{R}(k_1, k_2, k_3),$$

(8)

where the (reduced) bispectrum $B_\mathcal{R}(k_1, k_2, k_3)$ depends only on the magnitude of the $k$’s by rotational invariance. According to the particular model of inflation, the bispectrum will be peaked about a particular shape of the triangle. The two most common cases are the squeezed (or local) NG, peaked on squeezed triangles $k_1 \ll k_2 \approx k_3$, and the equilateral NG, peaked on equilateral triangles $k_1 \approx k_2 \approx k_3$. Indeed, one can define a scalar product of bispectra, which describes how sensitive one is to a NG of a given type if the analysis is performed using some template form for the bispectrum. As expected, the local and equilateral shapes are approximately orthogonal with respect to this scalar product [29]. We will now describe these two models in more detail.

**The local model:**
The local bispectrum is produced when the NG is generated outside the horizon, for instance in the curvaton model [30, 31] or in the inhomogeneous reheating scenario [32]. In these models, the curvature perturbation can be written in the following form,

$$\mathcal{R}(x) = \mathcal{R}_g(x) + \frac{3}{5} f^2_{\text{loc}} \left( \mathcal{R}^2_g(x) - \langle \mathcal{R}^2_g \rangle \right) + \frac{9}{25} g_{\text{NL}} \mathcal{R}^3_g(x),$$

(9)

where $\mathcal{R}_g$ is the linear, Gaussian field. We have included also a cubic term, which will generate the trispectrum at leading order. The bispectrum is given by

$$B_R(k_1, k_2, k_3) = \frac{6}{5} f^2_{\text{NL}} [P_R(k_1)P_R(k_2) + \text{cycl.}],$$

(10)

where “cycl.” denotes the 2 cyclic permutations of the wavenumbers, and $P_R(k)$ is the power spectrum given by $P_R(k) = Ak^{n_s - 4}$. The trispectrum is given by

$$\langle \mathcal{R}(k_1)\mathcal{R}(k_2)\mathcal{R}(k_3)\mathcal{R}(k_4) \rangle_c = (2\pi)^3 \delta_D(k_1 + k_2 + k_3 + k_4) \times \left[ \frac{36}{25} f^2_{\text{NL}} \sum_{a \neq b, c} P_R(|k_a + k_b|)P_R(k_b)P_R(k_c) + \frac{54}{25} g_{\text{NL}} \sum_{a < b < c} P_R(k_a)P_R(k_b)P_R(k_c) \right].$$

(11)

The equilateral model:

Models with derivative interactions of the inflaton field [33, 34, 35] give a bispectrum which is peaked around equilateral configurations, whose specific functional form is model dependent. Moreover, the form of the bispectrum is usually not convenient to use in numerical analyses. This is why, when dealing with equilateral NG, it is convenient to use the following parametrization, given in Ref. [36],

$$B_R(k_1, k_2, k_3) = \frac{18}{5} f^2_{\text{NL}} A^2 \left[ \frac{1}{2k_1^4 - n_s k_2^4 - n_s} + \frac{1}{3(k_1 k_2 k_3)^2(4 - n_s)/3} - \frac{1}{(k_1 k_2^4 k_3^2)^{4 - n_s}/3} + \text{5 perms.} \right].$$

(12)

This is peaked on equilateral configurations, and its scalar product with the bispectra produced by the realistic models cited above is very close to one. Therefore, being a sum of factorizable functions, it is the standard template used in data analyses.

3 Random walks and the halo mass function

We now turn to the main calculation of the paper. The non-Gaussian halo mass function can be obtained by calculating the barrier first crossing rate $\mathcal{F}$ of a random walk with non-Gaussian noise, in the presence of an absorbing barrier. This can be done perturbatively, starting from a path integral approach as prescribed by MR [19, 21] and the mass function can be shown to be $f(\sigma) = 2\sigma^2 \mathcal{F}(\sigma)$. As discussed by MR, the calculation of $f$ involves certain assumptions regarding the type of filter used and also the location of the barrier. In particular, the formalism is simplest for a sharp filter in $k$-space, and using the spherical top-hat of Eqn. (6) introduces complications in the form of non-Markovian effects. Further, in order to make the spherical collapse ansatz more realistic and obtain better agreement with $N$-body simulations, MR show that it is useful to treat the location of the barrier $\delta_c$ as a stochastic variable itself, and allow it to diffuse. For the time being, we will ignore these complications, and will return to their effects in section 5.

To make the paper self-contained, we begin with a brief review of the path integral approach to the calculation of the mass function. The reader is referred to Ref. [19] for a more pedagogical introduction. In the path integral approach, one treats the variance $\sigma^2_{\delta_R} \equiv \langle \delta^2_R \rangle$ as a “time” parameter, $t \equiv \sigma^2_{\delta_R}$, and considers the random walk followed by the smoothed density field $\tilde{\delta}_R$ as this “time” is increased in discrete steps starting from small values (equivalently, as $R$ is decreased from very large values). Here $\tilde{\delta}_R(\vec{x})$ is a stochastic quantity in real space due to the stochasticity inherent in the initial conditions. We use the notation $\tilde{\delta}_R$ to distinguish
the stochastic variable from the values it takes, which will be noted by $\delta_i$ below. We probe this stochasticity by changing the smoothing scale at a fixed location $\vec{x} = 0$, thus making the variable perform a random walk, which obeys a Langevin equation
\[ \frac{\partial \hat{\delta}}{\partial t} = \hat{\eta}, \]
with a stochastic noise $\hat{\eta}$ whose statistical properties depend on the choice of filter used. In particular, for a top hat filter in $k$-space, the noise is white, i.e. its 2-point function is a Dirac delta [15],
\[ \langle \hat{\eta}(t_1)\hat{\eta}(t_2) \rangle = \delta_D(t_1 - t_2) \]
The random walk can be described as a trajectory $\{\delta_0, \delta_1, \ldots, \delta_n\}$ which starts with $\hat{\delta}$ taking the value $\delta_0 = 0$ at $t = 0$ (or $R \to \infty$ which is the homogeneous limit), then taking values $\delta_i$ at times $t_i$, finally arriving at $\delta_n$ at time $t_n$, with a discrete timestep $\Delta t = t_{k+1} - t_k = t_n/n$. The probability $P(t)$ that the trajectory crosses the barrier at $\delta_c$ at a time larger than some $t$ (i.e. at scales smaller than the corresponding $R$ or $M$), is the same as the probability that the trajectory did not cross the barrier at any time smaller than $t$, so that
\[ P(t) = \int_{-\infty}^{\delta_c} d\delta_1 \ldots d\delta_n W(\{\delta_j\}; t), \]
where the probability density over the space of trajectories, $W(\{\delta_j\}; t)$ is defined as
\[ W(\{\delta_j\}; t) \equiv \langle \delta_D(\hat{\delta}(t_1) - \delta_1) \ldots \delta_D(\hat{\delta}(t_n) - \delta_n) \rangle, \]
where $\delta_D$ is the Dirac delta distribution. The first crossing rate is given by the negative time derivative of $P$, $\mathcal{F} = -\partial_t P$, and the mass function is then $f = 2t\mathcal{F}(t)$. In Eqn. (16) one can write the Dirac deltas using the integral representation $\delta_D(x) = \int_{-\infty}^{\infty} d\lambda e^{-\lambda x}/2\pi$, to obtain
\[ W(\{\delta_j\}; t) = \int_{-\infty}^{\infty} \frac{d\lambda_1}{2\pi} \ldots \frac{d\lambda_n}{2\pi} \langle e^{-i\sum_j \lambda_j \hat{\delta}(t_j)} \rangle e^{i\sum_j \lambda_j \delta_j}. \]
The object $\langle e^{-i\sum_j \lambda_j \hat{\delta}_j} \rangle$ is the exponential of the generating functional of the connected Green’s functions, and can be shown to reduce to [37]
\[ \langle e^{-i\sum_j \lambda_j \hat{\delta}_j} \rangle = \exp \left[ \sum_{p=2}^{\infty} \frac{(-i)^p}{p!} \sum_{j_1, \ldots, j_p=1}^n \lambda_{j_1} \ldots \lambda_{j_p} \langle \hat{\delta}_{j_1} \ldots \hat{\delta}_{j_p} \rangle_c \right], \]
where $\langle \hat{\delta}_{j_1} \ldots \hat{\delta}_{j_p} \rangle_c$ is the connected $p$-point function of $\hat{\delta}$, with the short-hand notation $\hat{\delta}_j = \hat{\delta}(t_j)$.

### 3.1 Halo mass function: Gaussian case, sharp-$k$ filter

In the Gaussian case, all connected $n$-point correlators vanish except for $n = 2$, and in the Markovian (sharp-$k$ filter) case which we are considering, the 2-point function becomes $\langle \hat{\delta}\hat{\delta}_k \rangle = \min(t_j, t_k)$, where $\min(t_j, t_k)$ is the minimum of $t_j$ and $t_k$. The resulting $n$-dimensional Gaussian integral can be handled in a straightforward way to obtain
\[ W_{gm} = \prod_{k=0}^{n-1} \Psi_{\Delta t}(\delta_{k+1} - \delta_k); \quad \Psi_{\Delta t}(x) = (2\pi\Delta t)^{-1/2}e^{-x^2/(2\Delta t)} , \]
where we follow MR’s notation and use the superscript “gm” to denote “Gaussian Markovian”. As MR have shown [19], the resulting expression for $P_{gauss}(t)$ in the continuum limit $\Delta t \to 0$ is simply
\[ P_{gauss} = \int_{-\infty}^{\delta_c} d\delta_1 \ldots d\delta_n W_{gm} = \text{erf} \left( \frac{\nu}{\sqrt{2}} \right), \]
where we use the notation $\nu \equiv \delta_c/\sigma$. (This in principle also includes the redshift dependence of the collapse threshold $\delta_c$, see below.) This expression for the continuum limit probability $P_{gauss}$ is of course a well-known
result going back to Chandrasekhar [38]. This leads to the standard excursion set result for the Gaussian mass function $f_{PS} = -2t \partial_t |_{\delta_c} P_{gauss}$,

$$f_{PS}(\nu) = \sqrt{\frac{T}{\pi}} \nu e^{-\nu^2/2},$$

where we use the subscript PS (for Press-Schechter), to conform with the conventional notation for this object.

### 3.2 Halo mass function: non-Gaussian case, sharp-$k$ filter

In the non-Gaussian case (but still retaining the sharp-$k$ filter), the probability density $W(\{\delta_j\}; t)$ also gets contributions from connected $n$-point correlators with $n \geq 3$, since these in general do not vanish. These can be handled by using the relation $\lambda_k e^{i \sum \lambda_j \delta_j} = -i \partial_k e^{i \sum \lambda_j \delta_j}$, with $\partial_j \equiv \partial / \partial \delta_j$. A straightforward calculation then shows the mass function to be

$$f = -2t \frac{\partial}{\partial t} |_{\delta_c} \int_{-\infty}^{\delta_c} d\delta_1 \ldots d\delta_n \exp \left[ -\frac{1}{3!} \sum_{j,k,l=1}^{n} \langle \delta_j \delta_k \delta_l \rangle \partial_j \partial_k \partial_l + \frac{1}{4!} \sum_{j,k,l,m=1}^{n} \langle \delta_j \delta_k \delta_l \delta_m \rangle \partial_j \partial_k \partial_l \partial_m + \ldots \right] W_{gm},$$

where it is understood that one takes the continuum limit $\Delta t \to 0$ before computing the overall derivative with respect to $t$. We will find it useful to change variables from $(\delta_c, t)$ to $(\nu, t)$, in which case the partial derivative becomes

$$-2t (\partial / \partial t) |_{\delta_c} = \nu (\partial / \partial \nu) |_{t} - 2t (\partial / \partial t) |_{\nu} \equiv \nu \partial_\nu - 2t \partial_t.$$

It is also useful at this stage to take a small detour and introduce some notation which we will use throughout the rest of the paper. We define the scale dependent “equal time” functions

$$\epsilon_{n-2} = \frac{\langle \delta^2 \rangle_{\nu}}{\sigma^2_R}; \quad n \geq 3,$$

which as we will see, remain approximately constant over the scales of interest. We assume the ordering $\epsilon_{n-2} \sim O(\epsilon^{n-2})$ with $\epsilon \ll 1$, which can be motivated from their origin in inflationary physics, where one finds $\epsilon_1 \sim f_{NL} A^{1/2}$, $\epsilon_2 \sim g_{NL} A$, etc. Typically we expect $\epsilon \lesssim 10^{-2}$ for $f_{NL} \lesssim 100$. Fig. 1 shows the behaviour of $\epsilon_1$ and $\epsilon_2$ in the local and equilateral models, as a function of $t = \sigma_R^2$. We see e.g. that $\epsilon_2$ in the local model is comparable to $\epsilon_1^2$. In the literature one usually encounters the reduced cumulants $S_n$, which are related to the $\epsilon_{n-2}$ by $\epsilon_1 = \sigma S_3$, $\epsilon_2 = \sigma^2 S_4$ and so on. The motivation for using the $S_n$ comes from the study of NG induced by nonlinear gravitational effects. However, as we see from Fig. 1, when studying primordial NG it is more meaningful to consider the $\epsilon_n$ which are approximately scale-independent and perturbatively ordered.

We will soon see that a natural expansion parameter that arises in the calculation has the form $\sim \nu$, and we therefore require that the mass scales under scrutiny are not large enough to spoil the relation $\epsilon \nu \ll 1$. It turns out that observationally interesting mass scales can nevertheless be large enough to satisfy $\epsilon \nu^3 \sim O(1)$. Fig. 2 shows the behaviour of $\epsilon_1 \nu^3$ and $\epsilon_1 \nu$ at different redshifts, as a function of mass, in our reference ΛCDM model for local type NG, with $f_{NL} = 100$. The behaviour for the equilateral NG is qualitatively similar. The redshift dependence of these quantities comes from the definition of $\nu$,

$$\nu(M, z) = \sqrt{\frac{\sigma}{\sigma(M)}} \frac{D(0)}{D(z)} = \frac{\delta_c(z)}{\sigma(M)},$$

where we denote the usual spherical collapse threshold as $\delta_{c0} = (3/5)(3\pi/2)^{2/3} \simeq 1.686$, reserving $\delta_c$ for the full, redshift dependent quantity, and $a$ is a parameter accounting for deviations from the simplest collapse model. In the standard spherical collapse picture we have $a = 1$. A value of $a$ different from unity (specifically

---

1Notationally we distinguish the order parameter $\epsilon$ from the specific NG functions $\epsilon_1$ and $\epsilon_2$. 

---
\( \sqrt{a} \approx 0.89 \) can be motivated by allowing the collapse threshold to vary stochastically [20], as we will discuss in section 5. We will soon see that the object \( \epsilon \nu^3 \) appears naturally in the calculation, and to be definite we will assume \( \epsilon \nu^3 \sim O(1) \) for now. In section 4 we will discuss the effects of relaxing this condition and probing smaller length scales.

We now turn to the “unequal time” correlators appearing in Eqn. (22). Since we are concerned with large scales, we are in the small \( t \) limit, and following MR we expand the \( n \)-point correlators around the “final time” \( t \). We can define the Taylor coefficients

\[
G_3^{(p,q,r)}(t) = \left[ \frac{dp}{dt_j} \frac{dq}{dt_k} \frac{dr}{dt_l} \right] \langle \delta(t_j) \delta(t_k) \delta(t_l) \rangle_c, \tag{26}
\]

and then expand

\[
\langle \delta_j \delta_k \delta_l \rangle_c = \sum_{p,q,r=0}^{\infty} \frac{(-1)^{p+q+r}}{p!q!r!} G_3^{(p,q,r)}(t)(t-t_j)^p(t-t_k)^q(t-t_l)^r. \tag{27}
\]

For the 4-point function we will have an analogous expression involving coefficients \( G_4^{(p,q,r,s)} \).

Since calculations involving a general set of coefficients \( G_3, G_4, \) etc. are algebraically rather involved, we find it useful to first consider an example in which these coefficients take simple forms. In this toy model we assume that the \( \epsilon_n \) are exactly constant, and moreover that the \( n \)-point correlators take the form\(^2\)

\[
\langle \delta_j \delta_k \delta_l \delta_m \rangle_c = \frac{\epsilon_1}{2}(t_jk_l)^{1/2}; \langle \delta_j \delta_k \delta_l \delta_m \rangle_c = \epsilon_2(t_jk_lk_m)^{1/2}. \tag{28}
\]

For clarity, we will display details of the calculation only for this model. We have relegated most of the technical details of our calculation to Appendix A. In Appendix A.1 we show that the mass function for this model can be brought to the form

\[
f = \left( \frac{2}{\pi} \right)^{1/2} e^{-(\epsilon_1/3)\nu^3 + (\epsilon_2/4)\nu^4} - \frac{1}{4} \nu e^{-\nu^2/2} + \frac{5}{16} \varepsilon_1 \left( \frac{\pi}{2} \right)^{1/2} \text{erfc} \left( \frac{\nu}{\sqrt{2}} \right) + \frac{1}{8} \left( \frac{\varepsilon_1^2 - 2\varepsilon_2}{3} \right) e^{-\nu^2/2} \left( \nu^2 - 1 \right) + O(e^3\nu^4), \tag{29}
\]

\(^2\)Throughout the paper we will consider at most 4-point correlators. This truncation is justified given our assumptions, as we will see later.

Figure 1: Scale dependence of the \( \epsilon_n \). Panel (a): Behaviour of \( \epsilon_1 \) vs. \( \sigma^2 \) in the local and equilateral models, for \( f_{NL} = 100 \) in each case. Panel (b): Behaviour of 2nd order \( (\sim \epsilon^2) \) terms. We show \( \epsilon_2 \) for the local model with \( f_{NL} = 100 \) and \( g_{NL} = 10^4 \). The terms proportional to \( f_{NL}^2 \) and \( g_{NL} \) are shown separately. Also shown is \( \epsilon_1^2 \) for the same model.
structure of the correlations. Appendix A.3 shows how to do this, and the result is by Eqn. (28). It is straightforward to relax these assumptions and perform the calculation with the exact scales, and $\epsilon \sim 1$ will require changes in our expressions for $f$ in terms of which cosmological NG are typically defined. We should then ask whether this difference in variances is of order $10^{-4}$. The remaining exponentiated derivative can be handled using a saddle point approximation. We show how to do this in Appendix A.2. The expression for the mass function $f(\nu)$ works out to

$$f(\nu) = \left(\frac{2}{\pi}\right)^{1/2} \nu \exp \left[ -\frac{1}{2} \nu^2 \left( 1 - \frac{\nu}{3} \right) + \frac{1}{4} \left( \frac{\nu}{3} - \frac{\nu}{2} \right) \nu^2 + O(\nu^3) \right]$$

which superficially at least, is comprised of two series expansions, one in the exponential and one as a polynomial, both based on the small parameter $\epsilon \nu$ (see however the next section).

This derivation assumed that $\epsilon_1$ and $\epsilon_2$ are constant, and that the unequal time correlations are given by Eqn. (28). It is straightforward to relax these assumptions and perform the calculation with the exact structure of the correlations. Appendix A.3 shows how to do this, and the result is

$$f(\nu, t) = \left(\frac{2}{\pi}\right)^{1/2} \nu \exp \left[ -\frac{1}{2} \nu^2 \left( 1 - \frac{\nu}{3} \right) + \frac{1}{4} \left( \frac{\nu}{3} - \frac{\nu}{2} \right) \nu^2 + O(\nu^3) \right]$$

$$\times \left\{ 1 - \frac{1}{4} \epsilon_1 \nu \left( 4 - c_1 \right) + \frac{1}{\nu^2} \left( c_1 - \frac{1}{4} c_2 - 2 \right) \right\}$$

$$+ \frac{1}{8} \nu^2 \left( \epsilon_1^2 (11 - 3c_1) - 2\epsilon_2 \left( 1 + \frac{1}{3} c_4 - \frac{1}{3} \frac{d \ln \epsilon_2}{\ln t} \right) \right) + O(\epsilon^3) \right\},$$

where the functions $c_n(t)$ are defined in Eqn. (A.17) and characterize the behaviour of the unequal time correlations. In our toy model above, the $c_n$ reduce to unity. Indeed, as a check we see that the expression in (31) reduces to Eqn. (30) if we take $\epsilon_1, \epsilon_2$ to be constant and set the $c_n$ to unity.

One issue which we have ignored so far, is that the definition of $\nu$ involves the variance $t = \sigma^2$ of the non-Gaussian field. Computationally it is more convenient to work with the variance $\sigma^2$ of the Gaussian field in terms of which cosmological NG are typically defined. We should then ask whether this difference will require changes in our expressions for $f$. We start by noting that this difference in variances is of order $\sim \epsilon^2$. For example, in the local model one has $\sigma^2(R) = Ad_1(R) + A(Ad^2_{NL})d_2(R)$ where $A \sim 10^{-9}$ is an overall normalization constant, $d_1$ and $d_2$ are scale dependent functions of comparable magnitude on all relevant scales, and $\epsilon$ is estimated as $\epsilon \sim f_{NL}^{1/2}$. We therefore have $\nu = \delta_\epsilon / \sigma = (\delta_\epsilon / \sigma_\delta)(1 + O(\epsilon^2))$. However, with our assumption that $\epsilon^2 \nu^3 \sim \mathcal{O}(1)$, we see that this correction is actually of order $\sim (\epsilon^2 \nu^3) \nu^{2-3} \sim \epsilon^2 \nu^3$, which we have been consistently ignoring. We will see that even when we relax the assumption $\epsilon^2 \nu^3 \sim \mathcal{O}(1)$ and

![Figure 2](image_url)
probe smaller scales where \( \epsilon \nu^3 \ll 1 \), this correction can still be consistently ignored. Hence we can safely set \( \nu = \delta_c / \sigma_e \) in all of our expressions.

4 Consistency of the truncation

4.1 Comparative sizes of terms in the mass function

Now that all the derivative operators which we consider important have been accounted for, we can check whether our final result is consistently truncated, i.e. whether we have retained all terms at any given order in the expansion. Symbolically, our current result for the mass function can be written as

\[
 f \sim e^{-\frac{1}{2} \nu^2 (1 + \epsilon \nu + \nu^2 \nu^3 + O(\epsilon^2 \nu^3))} \left[ 1 + \epsilon \nu + \frac{\epsilon}{\nu} + \epsilon^2 \nu^2 + O(\epsilon \nu^{-3}, \epsilon^2, \epsilon^3 \nu^3) \right],
\]

with the understanding that coefficients are computed (but not displayed) for all terms except those indicated by the \( O() \) symbols. Also, \( \epsilon^2 \) refers to both \( \epsilon_1^2 \) and \( \epsilon_2^2 \).

Since the expansions involve two parameters, \( \nu \epsilon \) and \( \nu^{-2} \), they make sense only if we additionally prescribe a relation between these parameters. So far we assumed that \( \epsilon \) is fixed and \( \nu \) is such that \( \epsilon \nu^3 \approx 1 \), which was based on the observation that the term \( \epsilon \nu^3 \) naturally appears in the exponent and is not restricted in principle to small values. In Appendix B.1 we discuss this condition in more detail, and also analyse the consequences of relaxing this condition and probing smaller mass scales. We find that for observationally accessible mass scales larger than the scale where \( \epsilon \nu^3 \approx \nu^{-3} \), the single expression

\[
 f \sim e^{-\frac{1}{2} \nu^2 (1 + \epsilon \nu + \epsilon^2 \nu^2)} \left[ 1 + \epsilon \nu + \frac{\epsilon}{\nu} + O(\epsilon^3 \nu^5, \epsilon^2 \nu^2, \epsilon \nu^{-3}) \right],
\]

is parametrically consistent as it stands – the terms ignored are smaller than the smallest terms retained – and in fact it remains a very good approximation even when \( \epsilon \nu^3 \approx 1 \), since the only “inconsistent” term then is \( \epsilon \nu^{-1} \), whose effect reduces as \( \nu \) increases. On scales where \( \epsilon \nu^3 \approx \nu^{-4} \) and lower, the theoretical error becomes comparable to or larger than the quadratic term in the exponential. Plugging back all the coefficients, we have the following result for the mass function (excluding filter effects, see section 5),

\[
 f(\nu, t) = f_{\text{PS}}(\nu) \exp \left( \frac{1}{6} \epsilon_1 \nu^3 - \frac{1}{8} \left( \epsilon_2^2 - \frac{\nu}{3} \right) \nu^4 \right) \\
 \times \left\{ 1 - \frac{1}{4} \epsilon_1 \nu \left( 4 - c_1 \right) + \frac{1}{\nu^2} \left( c_1 - \frac{1}{4} c_2 - 2 \right) \right\} + O(\epsilon^3 \nu^5, \epsilon^2 \nu^2, \epsilon \nu^{-3}) \right\}.
\]

4.2 Comparing with previous work

In this subsection we compare our results with previous work on the non-Gaussian mass function. As mentioned in the introduction, this quantity has been computed by several authors in different ways [16, 17, 21]. If one considers the range of validity of the perturbative expansion, the strongest result so far has been due to MVJ [16], who explicitly retain the exponential dependence on \( \epsilon_1 \). Their expression for \( f \) can be written as

\[
 f_{\text{MVJ}} = f_{\text{PS}}(\nu) \frac{\epsilon_1 \nu^3 / 6}{(1 - \epsilon_1 \nu^3/3)^{1/2}} \left( 1 - \frac{1}{\nu} \epsilon_1 \nu \left( 1 - \frac{2}{3} \frac{d \ln \epsilon_1}{d \ln \nu} \right) \right).
\]

The major shortcoming of their result is that it is based on a Press-Schechter like prescription, and must therefore be normalized by an appropriate Gaussian mass function, typically taken to be the one due to

---

\[3\] The analysis presented by MVJ in fact allows one to retain terms like \( \sim \epsilon^2 \nu^4 \) in the exponential as well, and we have seen that when \( \epsilon \nu^3 \approx 1 \), these terms are as important as the polynomial \( \epsilon \nu \) term retained by MVJ. However, since the MVJ expression misses unequal time effects of order \( \sim \epsilon \nu \) anyway, it is reasonable to compare our results with the expression (35), which is also the one used by most other authors (see e.g. Refs. [39, 40]).
Sheth & Tormen [18]. Additionally, it always misses the contributions due to the unequal time correlators, which contribute to the terms $\sim \epsilon \nu, \epsilon \nu^{-1}$, etc. in Eqn. (34). When one considers formal correctness on the other hand, MR have presented a result based on explicit path integrals, which accounts for the unequal time contributions, and which also does not need any ad hoc normalizations (in this context see also Lam & Sheth [25]). Indeed, our calculations in section 3 were based on techniques discussed by MR in Refs. [19, 21]. As we discuss below however, the fact that MR do not explicitly retain the exponential dependence of $\epsilon_1 \nu^3$, means that their result is subject to significant constraints on the range of its validity. Their expression for $f$, ignoring filter effects, is

$$f_{MR} = f_{PS}(\nu) \left( 1 + \frac{1}{6} \epsilon_1 \nu^3 \left( 1 - \frac{3}{2\nu^2} (4 - c_1) - \frac{3}{2\nu^2} \left( c_1 - \frac{1}{4} c_2 - 2 \right) \right) \right).$$

(36)

This expression is precisely what one obtains by linearizing our expression (34) in $\epsilon_1$, which serves as a check on our calculations. LMSV [17] present a result based on an Edgeworth expansion of the type encountered when studying NG generated by nonlinear gravitational effects [41]. The result most often quoted in the literature is their expression linear in $\epsilon$ when studying NG generated by nonlinear gravitational effects [41]. The result most often quoted in the literature is their expression linear in $\epsilon$ when studying NG generated by nonlinear gravitational effects [41]. The result most often quoted in the literature is their expression linear in $\epsilon$ when studying NG generated by nonlinear gravitational effects [41]. The result most often quoted in the literature is their expression linear in $\epsilon$ when studying NG generated by nonlinear gravitational effects [41]. The result most often quoted in the literature is their expression linear in $\epsilon$ when studying NG generated by nonlinear gravitational effects [41]. The result most often quoted in the literature is their expression linear in $\epsilon$ when studying NG generated by nonlinear gravitational effects [41]. The result most often quoted in the literature is their expression linear in $\epsilon$ when studying NG generated by nonlinear gravitational effects [41]. The result most often quoted in the literature is their expression linear in $\epsilon$ when studying NG generated by nonlinear gravitational effects [41]. The result most often quoted in the literature is their expression linear in $\epsilon$ when studying NG generated by nonlinear gravitational effects [41].

$$f_{LMSV,lin} = f_{PS}(\nu) \left( 1 + \frac{1}{6} \epsilon_1 \nu^3 \left( 1 - \frac{1}{\nu^2} \left( 3 - 2 \frac{d \ln \epsilon_1}{d \ln t} \right) - \frac{2}{\nu^4} \frac{d \ln \epsilon_1}{d \ln t} \right) \right).$$

(37)

In Appendix B.3 of Ref. [17], LMSV also give an expression involving $\epsilon_1^2$ and $\epsilon_2$, which can be written as

$$f_{LMSV,quad} = f_{PS}(\nu) \left[ 1 + \frac{1}{6} \epsilon_1 \left( H_3(\nu) + \frac{2 d \ln \epsilon_1}{\nu d \ln t} H_2(\nu) \right) + \frac{1}{72} \epsilon_2 \left( H_6(\nu) + \frac{4 d \ln \epsilon_1}{\nu d \ln t} H_5(\nu) + \frac{1}{24} \epsilon_2 \left( H_4(\nu) + \frac{2 d \ln \epsilon_2}{\nu d \ln t} H_3(\nu) \right) \right) \right],$$

(38)

where the $H_n(\nu)$ are the Hermite polynomials of order $n$. This expression was used by LMSV only as a check on the validity of their linear expression. By comparing with our expression which is non-perturbative in $\epsilon_1 \nu^3$, we will see below that these quadratic terms in fact significantly improve LMSV’s prediction.

Sticking to the linearized results, we see that the expressions of both MR and LMSV have the symbolic form

$$f \sim e^{-\nu^2/2} \left[ 1 + \epsilon \nu^3 + \epsilon \nu + \frac{\epsilon}{\nu} + \ldots \right],$$

(39)

where the ellipsis denotes all terms of the type $\epsilon \nu^{-3}, \epsilon \nu^{-5}$, etc., as well as all terms containing $\epsilon^2$. As we have seen, deciding where to truncate the expression for $f$ is not trivial, and using our more detailed expression we can ask whether the expression (39) is consistent at all the relevant length scales. Immediately, we see that this expression cannot be correct once $\epsilon \nu^3$ becomes close to unity. In Appendix B.2 we discuss the situation for MR and LMSV at smaller mass scales.

### 5 Effects of the diffusing barrier and the filter

In Ref. [20], MR showed that the agreement between a Gaussian mass function calculated using the statistics of random walks, and mass functions observed in numerical simulations with Gaussian initial conditions, can be dramatically improved by allowing the barrier itself to perform a random walk. This approach is motivated by the fact that the ignorance of the details of the collapse introduces a scatter in the value of the collapse threshold for different virialized objects. The width of this scatter was found by Robertson et al. [42] to be a growing function of $\sigma(M)$, which is consistent with the physical expectation that deviations from spherical collapse become relevant at small scales. The barrier can thus be treated (at least on a first approximation) as a stochastic variable whose probability density function obeys a Fokker-Planck equation with a diffusion

---

4We have corrected a typo in MR’s result [21]: the object they define as $V_3$ should appear with an overall positive coefficient in their Eqns. (85), (87) and (92).

---
Although both approximations are reasonable on the scales of interest, non-Gaussian and scale dependent corrections range of length scales. In fact, MR show that a very good analytical approximation for the symmetric object obtained by perturbatively expanding in $\Delta_{jk}$ is 

$$\Delta_{jk} \simeq \kappa \min(t_j, t_k) \left(1 - \frac{\min(t_j, t_k)}{\max(t_j, t_k)}\right),$$

(40)

where in our case we find $\kappa(R) \simeq 0.464 + 0.002R$, with $R$ measured in $h^{-1}$Mpc. The mass function is then obtained by perturbatively expanding in $\Delta_{ij}$, with the leading effect being due to the integral

$$\int_{-\infty}^{\delta_c} d\delta_1 \ldots d\delta_n \frac{1}{2} \sum_{j,k=1}^n \Delta_{jk} \partial_j \partial_k W^{gm},$$

which on evaluation leads to

$$f_{\text{r, sharp}-x}(\nu, t) \sim \nu \left[(1 - \kappa)e^{-\nu^2/2} + \frac{\kappa}{2} \Gamma \left(0, \frac{\nu^2}{2}\right) + \mathcal{O}(\kappa^2)\right],$$

(41)

where the subscript stands for Gaussian noise with the top-hat filter in real space, and $\kappa$ introduces a weak explicit $t(= \sigma^2)$ dependence. In Ref. [21] MR proposed an extension of this result to the non-Gaussian case, by assuming that all the non-Gaussian terms that they computed with the sharp-$k$ filter, would simply get rescaled by the factor $(1 - \kappa)$ at the lowest order, but otherwise retain their coefficients. Symbolically, their result (Eqn. 88 of Ref. [21]) is

$$f_{\text{ng, sharp}-x}(\nu, t) \sim \nu \left[(1 - \kappa)e^{-\nu^2/2} (1 + \nu^3 + \nu + \nu^{-1}) + \frac{\kappa}{2} \Gamma \left(0, \frac{\nu^2}{2}\right)\right],$$

(42)

with the specific coefficients of the $\nu^3$, $\nu$ and $\nu^{-1}$ terms being identical to those in Eqn. (36). However, the coefficient of e.g. the $\kappa \nu$ term arises from the action of an operator $\sim \sum_{j,k} \Delta_{jk} \partial_j \partial_k$ combining with the first unequal time operator $\sim \varepsilon_1^{1/2} \sum_j (t - t_j) \partial_j \sum_k \partial_k \partial_i$, and there is no simple way of predicting its exact

---

5A potential issue in this argument lies in the assumption of a linear Langevin equation for the stochastic barrier $B$, resulting in a simple Fokker-Planck equation with a constant $D_B$ like the one in MR, while the distribution of $B$ was found to be approximately log-normal (and therefore non-Gaussian) in Ref. [42]. One can see that a Langevin equation of the type $\partial_t B = B\xi$ (which would produce a log-normal distribution) can be approximated by $\partial_t B = \langle B\rangle \xi$, whenever the fluctuations around $\langle B\rangle$ are small, and gives a constant diffusion coefficient as long as $\langle B\rangle$ is constant. Although both approximations are reasonable on the scales of interest, non-Gaussian and scale dependent corrections to the barrier diffusion should be studied, since in principle they could be of the same order as the other corrections retained here. This investigation is left for future work.
value beforehand. Since MR explicitly neglect such “mixed” terms, their formula is not strictly inconsistent, as long as one keeps in mind that the theoretical error in their expression is of the same order as the terms \( \sim \kappa \nu \) that they include. However, if one wants to consistently retain such terms, a detailed calculation is needed. Our calculations (not displayed) indicate that the coefficient of the \( \kappa \nu \) term depends on certain details of the continuum limit of the path integral near the barrier, which require a more careful study. We are currently investigating methods of computing these effects. At present however, we conclude that the mixed terms involving both filter effects and NG, must be truncated at order \( \sim \kappa \nu \).

Finally, the filter-corrected mass function is also subject to effects of barrier diffusion. Here we make the same assumptions as MR do in Ref. [20], namely that the barrier location satisfies a Langevin equation with white noise and diffusion constant \( D_B \), which can be accounted for by replacing \( \kappa \rightarrow \kappa = \kappa/(1 + D_B) = a \kappa \). However, it is difficult to theoretically predict the unequal time behaviour of the barrier correlations, and these simple assumptions must also be tested, perhaps by suitably comparing with the detailed results of Robertson et al. [42]. We leave this for future work. Our final expression for the mass function, corrected for effects of the diffusing barrier and the top-hat real space filter, is

\[
\begin{align*}
\hat{f}(\nu, t) &= f_{PS}(\nu) \left( 1 - \kappa + \mathcal{O}(\kappa^2) \right) \exp \left[ \frac{1}{6} \varepsilon_1 \nu^3 - \frac{1}{8} (\varepsilon_1^2 - \varepsilon_2^2) \nu^4 \right] \\
&\quad \times \left\{ 1 + \left( \frac{1 - 2d \ln \kappa/d \ln t}{1 - \kappa} \right)^2 \nu^2 \left( 1 - 2\nu^2 \right) - \frac{1}{4} \varepsilon_1 \nu \left( \frac{c_1}{1 - \kappa} + 4 - 2c_3 \right) \right. \\
&\quad \left. \quad \quad - \frac{1}{4} \varepsilon_1 \nu^{-1} \left( c_1 - \frac{1}{4} c_2 - 2 \right) \right. \\
&\quad \left. \quad \quad + \mathcal{O}(\kappa^2 \nu^{-2}, \kappa \nu, \kappa \nu^{-6}) + \mathcal{O}(\nu^2 \nu^2, \nu^3 \nu^5, \nu^3 \nu^5) \right\},
\end{align*}
\]

where we have chosen to account for the scale independent \( \mathcal{O}(\kappa^2) \) error arising from filter effects, as an overall normalization uncertainty, and have explicitly displayed the orders of the various terms we ignore. Here \( f_{PS}(\nu) \) is given by Eqn. (21) with \( \nu(M, z) \) defined in Eqn. (25).

To summarize, Eqn. (43) gives an analytical expression for the non-Gaussian mass function. This expression is based on approximations that are valid over a larger range of length scales than the ones presented by MR and LMSV, and incorporates effects which are ignored in the expression presented by MVJ and LMSV. Like all these other mass functions, it suffers from the errors introduced by filter effects. However, the largest of these can be accounted for as an overall normalization constant, which can be fixed using, for instance, results of a Gaussian simulation. Among the NG functions \( \varepsilon_1, \varepsilon_2 \) and the \( c_n \) which appear in the mass function, the most important ones over the mass range of interest are \( \varepsilon_1, \varepsilon_2 \) and \( c_1 \) which are nearly constant. In principle though, all these functions must be computed numerically for every mass scale of interest, and indeed all the plots in this paper use the results of such numerical calculations. However, since this is somewhat tedious to do in practice, in Table 1 we provide analytical approximations for \( \varepsilon_1, \varepsilon_2, c_1, c_2 \) and \( c_3 \), for the local and equilateral case as a function of \( \sigma^2 \). As mentioned earlier, all these quantities are independent of redshift, although they depend on the choice of cosmological parameters in a complicated way in general due to the presence of the transfer function in their definitions. However, the dependence on \( \sigma_8 \) is simple, and one can check that we have \( \varepsilon_1 \propto \sigma_8 \), \( \varepsilon_2 \propto \sigma_8^2 \) and that the \( c_n \) are independent of \( \sigma_8 \). Recall that the \( c_n \) are also independent of \( f_{NL} \) and \( g_{NL} \). For completeness, in Table 1 we also give approximations for the filter parameters \( \kappa / d \ln \kappa / d \ln t \) which appear in the mass function, as functions of \( \sigma^2 \).

### 6 Results and Discussion

In this section we conclude with our final results for the non-Gaussian halo mass function, comparing our approach with previous work. In principle, we should compare the full expressions for the mass functions of various authors with ours. However, recall that for MVJ and LMSV one has to multiply an analytically

---

\( ^{6} \) Notice that this issue is completely decoupled from the subtleties in truncation discussed in section 4 – this problem remains even at scales where MR’s expression is formally consistent.
predicted ratio $R_{\text{ng}} = f(\nu, M, f_{\text{NL}})/f(\nu, M, f_{\text{NL}} = 0)$ with a suitable Gaussian mass function based on fits to simulations. It is not clear how to compute theoretical error bars on the latter. On the other hand, the object $R_{\text{ng}}$ itself is an unambiguous theoretical prediction of every approach, that is MVJ, LMSV, MR and our work, and we can compute theoretical errors on it. In this work, we will restrict ourselves to comparing the different expressions for $R_{\text{ng}}$. In future work, we hope to compare both $R_{\text{ng}}$ and the full mass function with the results of $N$-body simulations.

In Fig. 3 and Fig. 4 we plot the ratio $R_{\text{ng}}$, respectively without and with the filter effects, at redshift $z = 1$. In this way we can explicitly disentangle the errors due to an approximate treatment of non-Gaussian effects, from those due to the filter effects. We compare our expression (43) with the expressions of MR (36), LMSV (37) and (38), and of MVJ (35). Notice that, when considering the filter effects, the Gaussian function that enters in the ratio $R_{\text{ng}}$ is defined to be the function with $f_{\text{NL}} = 0$ (i.e. without NG but with filter effects when present). We use the local model, setting $f_{\text{NL}} = 100$ and $g_{\text{NL}} = 0$, and use the reference $\Lambda$CDM cosmology described in section 2. We do not explicitly show the final results for the equilateral model, but they are qualitatively similar. As is commonly done in the literature, we modify the LMSV and MVJ curves by applying the Sheth et al. correction of $\delta_{\text{c},0} \to \sqrt{\alpha}\delta_{\text{c},0}$. An identical correction is already present in the expressions (43) and (36) due to the barrier diffusion. We set $\sqrt{\alpha} = 0.89$, which is the value inferred by MR in Ref. [20] using the simulations of Robertson et al. [42]. We wish to emphasize a feature that our calculation shares with that of MR, which is that the constant $a$ is the only parameter whose value depends on the output of $N$-body simulations. The rest of the calculation for the mass function is completely analytical and from first principles.

To make the comparison meaningful, we introduce theoretical error bars on the curves. These error bars have no intrinsic statistical meaning – they simply keep track of the absolute magnitude of the terms that are ignored in any given prescription for computing the mass function. As we have discussed at length in section 4, these theoretical errors are scale dependent. The estimated error magnitude for each point is the maximum among the terms ignored in the expression. More explicitly, the errors for the linearized LMSV expression (37) are estimated as the maximum of $(\epsilon \nu^3)^2$ which comes from the expansion of the exponential, $\epsilon \nu$ which is the order of the largest unequal time terms missing, and $\tilde{\kappa} \nu^{-2}$ which comes from the filter effects. The errors for the LMSV expression (38) are similarly estimated as the maximum of $(\epsilon \nu^3)^3$, $\epsilon \nu$ and $\tilde{\kappa} \nu^{-2}$. The largest error for the MVJ expression (35) is the maximum of $\epsilon \nu$ (unequal time terms) and $\tilde{\kappa} \nu^{-2}$ (filter effects). Finally, the error for the MR expression (36) is the maximum of $(\epsilon \nu^3)^2$ from the expansion of the exponential, $\epsilon \nu^{-3}$ from the largest unequal time terms ignored, and $\tilde{\kappa} \nu^{-2}$ and $\tilde{\kappa} \nu$ from the filter effects. We include the filter effects and the associated errors only in Fig. 4.
Figure 3: Theoretical comparison of the different mass functions at $z = 1$, without the filter effects, i.e. setting $\tilde{\kappa} = 0$. We plot the ratio $R_{ng}$ of the non-Gaussian and Gaussian mass functions, in the local model with $f_{\text{NL}} = 100$ and $g_{\text{NL}} = 0$. See main text for a discussion of the error bars. The arrow indicates the mass scale where $\varepsilon 1^{3/6} = 1$, i.e. where the expansions of LMSV (both linearized and quadratic) and MR break down.

From these figures, we can draw some interesting conclusions. First of all, we see that it is important to retain terms which are quadratic in the NG, either with a saddle point method like in MVJ and in our formula, or by expanding the exponential up to second order, like in LMSV. Actually, we argue that it is correct to keep the exponential, otherwise the expansion breaks down when $\varepsilon 1^{3}$ is of order unity. We notice in passing that the term proportional to $\varepsilon 2$ which comes from the trispectrum, partially cancels with the $\varepsilon 1^{3}$ term. Secondly, comparing our expression with MVJ’s, we can observe that keeping the unequal time terms allows us to sensibly reduce the theoretical errors due to the approximate treatment of NG. In fact, if these terms are missing, they provide the largest theoretical error on large scales. Instead, the largest theoretical error on small scales comes from the approximations involved in dealing with a real space top-hat filter, as is apparent from Fig. 4.

To conclude, in this work we have calculated the non-Gaussian halo mass function in the excursion set framework, improving over previous calculations. We started from a path integral formulation of the random walk of the smoothed density field, following Ref. [19]. This allows us to take into account effects due to multi-scale correlations of the smoothed density field (“unequal time” correlations), and due to the real space top-hat filter, which generates non-Markovianities in the random walk. We recognize two small parameters in which we perturb: $\epsilon$, defined below Eqn. (24), which measures the magnitude of the primordial NG; and $\nu^{-1} = \sigma R/\delta_c$, which is small on very large scales. In order to do a consistent expansion and to estimate the theoretical errors, one must study the (scale dependent) relation between these two parameters, which we have discussed in Sec. 4. We then used saddle point techniques which allowed us to non-perturbatively retain the dependence on $\epsilon \nu^3$, which naturally appears in the calculation and whose magnitude becomes of order unity at high masses and high redshift. Finally, we included effects due to the choice of filter function and due to deviations from spherical collapse, as explained in Sec. 5. Our final result is presented in Eqn. (43),
In Table 1 we provide analytical approximations for the various parameters that appear in this expression, which in general must be computed numerically. We also considered other expressions for the mass function found in the literature, which use different expansion methods but do not estimate the theoretical errors. We estimated the theoretical errors for each formula, and we show comparative plots in Fig. 3 and Fig. 4. In our work we have improved over the calculations of MVJ [16] and LMSV [17] (who ignore unequal time correlations) and of MR [21] (who do not retain the exponential dependence on $\nu^3$). We have also demonstrated that the (linearized) result of LMSV can be significantly improved by retaining the quadratic terms of their calculation which are usually ignored in the literature. We find that at large scales and high redshifts, the biggest theoretical errors are introduced by ignoring the exponential dependence on $\nu^3$, followed by the neglect of unequal time correlations. The errors on our expression (44) are therefore significantly smaller than those of the others. The strength of our approach lies in the combination of path integral methods as laid out by MR [21], and the saddle point approximation as used by MVJ [16].

Our work can be continued in several directions. First, a thorough calculation of the effects due to the choice of the filter should be performed, since they lead to significant uncertainties in our final expression. This would include a study of the details of the continuum limit of the path integral near the barrier, and also a study of the statistics of the barrier diffusion process in the presence of filter effects. Second, a comparison with $N$-body simulations should be performed, in order to quantitatively assess the possibility of constraining NG using our work. Third, it would be interesting to study how to account for the effects of ellipsoidal
collapse, in a framework such as the one employed in this paper. Finally, an application to the void statistics along the same lines should be feasible. The problem here is made more interesting by the presence of two barriers, as discussed by Sheth & van de Weygaert [43], and since voids probe larger length/mass scales than halos, they constitute a promising future probe of primordial NG [44].

Acknowledgements

It is a pleasure to thank Stefano Borgani, Paolo Creminelli, Francesco Pace, Emiliano Sefusatti, Ravi Sheth, Licia Verde and Filippo Vernizzi for useful discussions.

Appendix

A Mass function calculation

A.1 Equal time vs. unequal time terms

In this appendix we show how the exponentiated derivative operator in the path integral can be handled by separating the contributions of the equal time and unequal time correlations. While this calculation assumes the toy model introduced in Eqn. (28), it easily generalizes to the more realistic case as we discuss later.

Using the first few terms of the unequal time expansions, in our toy model one can write

$$
\sum_{j,k,l=1}^{n} \langle \hat{\delta}_j \hat{\delta}_l \hat{\delta}_k \rangle_c \partial_j \partial_k \partial_l = \varepsilon_1 t^{3/2} \left( \sum_{j,k,l=1}^{n} \partial_j \partial_k \partial_l - \frac{3}{2} \sum_{j=1}^{n} (1 - \frac{t_j}{t}) \partial_j \sum_{k,l=1}^{n} \partial_k \partial_l \right. \\
- \frac{3}{8} \sum_{j=1}^{n} (1 - \frac{t_j}{t})^2 \partial_j \sum_{k,l=1}^{n} \partial_k \partial_l + \frac{3}{4} \sum_{j,k=1}^{n} (1 - \frac{t_j}{t})(1 - \frac{t_k}{t}) \partial_j \partial_k \sum_{l=1}^{n} \partial_l + \ldots \right), \quad (A.1)
$$

$$
\sum_{j,k,l,m=1}^{n} \langle \hat{\delta}_j \hat{\delta}_k \hat{\delta}_l \hat{\delta}_m \rangle_c \partial_j \partial_k \partial_l \partial_m = \varepsilon_2 t^2 \left( \sum_{j,k,l,m=1}^{n} \partial_j \partial_k \partial_l \partial_m - 2 \sum_{j=1}^{n} (1 - \frac{t_j}{t}) \partial_j \sum_{k,l,m=1}^{n} \partial_k \partial_l \partial_m + \ldots \right). \quad (A.2)
$$

These derivative operators are exponentiated in the path integral, and act on $W^{sm}$. One simplification that occurs in our toy model, is that the path integral in Eqn. (22) becomes a function only of $\nu$ (although this is not obvious at this stage), and hence eventually only the $\nu \partial_\nu$ part of the overall derivative contributes. However, the structure of the exponentiated derivatives is still rather formidable. Moreover, the truncation of the series at this stage is based more on the intuition that higher order terms should somehow be smaller, rather than on a strict identification of the small parameters. In fact, we will see in detail in section 4 that the issue of truncation involves several subtleties.

To make progress, it helps to analyze the effect on $W^{sm}$ of each of the terms in the above series, before exponentiation. The leading term in Eqn. (22) involves the multiple integral of $W^{sm}$, which is just the quantity $P_{\text{gauss}}$ encountered in Eqn. (20). The operator $\nu \partial_\nu$ acts on the error function to give the Gaussian rate of Eqn. (21). Next, notice that the action of the operator $\sum_{j=1}^{n} \partial_j$ on any function $g(\delta_1, \ldots, \delta_n)$ under the multiple integral, is simply

$$
\int_{-\infty}^{\delta_c} d\delta_1 \ldots d\delta_n \sum_{j=1}^{n} \partial_j g = \frac{\partial}{\partial \delta_c} \int_{-\infty}^{\delta_c} d\delta_1 \ldots d\delta_n g, \quad (A.3)
$$

Using this, and the fact that $t^{1/2}(\partial/\partial \delta_c)|_t = \partial_{\nu|t}$, we see that the leading term in Eqn. (A.1) (i.e. the term with no powers of $(1 - t_j/t)$), leads to a term involving

$$
\varepsilon_1 \nu \partial_\nu (\partial_{\nu})^3 \text{erf} \left( \nu / \sqrt{2} \right) \sim f_{PS} \varepsilon_1 \nu^3 (1 + O(\nu^{-2})),
$$

The problem with this term is that the quantity $\varepsilon_1 \nu^3$ can be of order unity, and hence cannot be treated perturbatively. To be consistent, we should keep all terms involving powers of $\varepsilon_1 \nu^3$. Luckily, this can be
done in a straightforward way due to the result in Eqn. (A.3). We see that the entire exponential operator \(\exp[-(\varepsilon_1 t^{3/2}/3!) \sum_{j,k,l=1}^n \partial_j \partial_k \partial_l]\) in Eqn. (22) can be pulled across the multiple integral and converted to \(\exp[-(\varepsilon_1 t^{3/2}/3!) \partial_0^3]\) acting on the remaining integral. Similarly, the operator \(\exp[(\varepsilon_2 t^{2}/4!) \sum_{j,k,l,m=1}^n \partial_j \partial_k \partial_l \partial_m]\) can be pulled out and converted to \(\exp[(\varepsilon_2 t^{2}/4!) \partial_0^n]\), and the same applies for all such equal time operators. We will see later that the action of these operators can be easily accounted for, using a saddle-point approximation. To summarize, the function \(f\) at this stage is given by

\[
f = \nu \epsilon^{-1/3} \partial_0^2 + (\varepsilon_2/4!) \partial_0^n + \ldots \partial_0 \int_{-\infty}^{\delta_c} d\delta_1 \ldots d\delta_n \exp \left[ \frac{1}{3!} \varepsilon_1 t^{3/2} \left( \frac{3}{2} \sum_{j=1}^n (1 - t_j/t) \partial_j \right) \right] \sum_{k,l=1}^n \partial_k \partial_l \right]
+ \frac{3}{8} \sum_{j=1}^n (1 - t_j/t)^2 \partial_j \right) \sum_{k,l=1}^n \partial_k \partial_l - \frac{3}{4} \sum_{j,k=1}^n (1 - t_j/t)(1 - t_k/t) \partial_j \partial_k \sum_{l=1}^n \partial_l + \ldots \right)
- \frac{1}{4!} \epsilon_2 t^2 \left( 2 \sum_{j=1}^n (1 - t_j/t) \partial_j \right) \sum_{k,l,m=1}^n \partial_k \partial_l \partial_m + \ldots \right] W^m. \tag{A.4}
\]

Now consider the action of the individual terms in the remaining exponential under the integrals, but without exponentiation. From MR [21], we have the following results\(^7\),

\[
\sum_{j=1}^n (1 - t_j/t) \sum_{k,l=1}^n \delta \sum_{l=1}^{\delta_c} d\delta_1 \ldots d\delta_n \partial_j \partial_k \partial_l W^m = \left( \frac{2}{\pi} \right)^{1/2} \frac{1}{t^{3/2}} e^{-\nu^2/2}, \tag{A.5a}
\]

\[
\sum_{j=1}^n (1 - t_j/t) \left[ \sum_{k,l=1}^n \delta \sum_{l=1}^{\delta_c} d\delta_1 \ldots d\delta_n \partial_j \partial_k \partial_l W^m = \left( \frac{2}{\pi} \right)^{1/2} \frac{3}{t^{3/2}} h(\nu), \tag{A.5b}
\]

\[
\sum_{j,k=1}^n (1 - t_j/t)(1 - t_k/t) \sum_{l=1}^{\delta_c} d\delta_1 \ldots d\delta_n \partial_j \partial_k \partial_l W^m = \left( \frac{2}{\pi} \right)^{1/2} \frac{2}{t^{3/2}} h(\nu), \tag{A.5c}
\]

where we have defined

\[
h(\nu) = e^{-\nu^2/2} - \left( \frac{\pi}{2} \right)^{1/2} \nu \text{erfc} \left( \frac{\nu}{\sqrt{2}} \right) = \frac{\nu}{2^{3/2}} \Gamma \left( -\frac{1}{2}, \frac{\nu^2}{2} \right), \tag{A.6}
\]

where \(\Gamma(-1/2, \nu^2/2)\) is an incomplete gamma function. Let us focus on the term in Eqn. (A.5a). If we linearize in \(\varepsilon_1\) in Eqn. (A.4), then this term appears with \(\varepsilon_1 t^{3/2} \partial_0\) acting on it, leading to \(\sim f_{PS} \varepsilon_1 \nu \ll f_{PS}\). This term can therefore be treated perturbatively. Similarly, one can check that the terms given by Eqns. (A.5b) and (A.5c) also lead to perturbatively small quantities, which are in fact further suppressed compared to \(\varepsilon_1 \nu\) by powers of \(\nu^2\). Specifically, one obtains terms involving \(\varepsilon_1 \text{erfc} \left( \nu/\sqrt{2} \right)\) which, for large \(\nu\), reduces to \(\sim f_{PS} \varepsilon_1 \nu \cdot \nu^{-2}(1 + O(\nu^{-2}))\).

A few comments are in order at this stage. First, this ordering in powers of \(\nu^{-2}\) is a generic feature of integrals involving an increasing number of powers of \((1 - t_j/t)\) being summed. This can be understood in a simple way from the asymptotic properties of the incomplete gamma function, as we show in Appendix C. We are therefore justified in truncating the Taylor expansion of the unequal time correlators, even though superficially (on dimensional grounds) each term in the series appears to be equally important. Secondly, we have not yet accounted for the effect of the exponential derivatives. In fact we will see in the next section that when \(\nu^2 \sim O(1)\), it is these terms that impose stricter conditions on the series truncations. For now, however, we have no guidance other than the fact that if we account for one term of order \(\sim \nu^3\), then we should account for all terms at this order. Given this, note that for \(\nu^3 \sim O(1)\) we have \(\nu^{-2} \sim \nu^3\), and hence the terms arising from Eqns. (A.5b) and (A.5c) are of order \(\sim \nu^3 \nu^2\). To consistently retain them, we must therefore also retain the term linear in \(\varepsilon_2\) and the one quadratic in \(\varepsilon_1\), when expanding the exponential. These

\(^7\)The terms in Eqns. (A.5a), (A.5b) and (A.5c) are, up to prefactors, the integrals of what MR denote as \(\Pi^{(3, NL)}, \Pi^{(3, NNLa)}\) and \(\Pi^{(3, NNLb)}\) respectively in Ref. [21].
involve the following quantities:

\[
\sum_{j=1}^{n} \left(1 - \frac{t_j}{t}\right) \sum_{k,l,m=1}^{n} \int_{-\infty}^{\delta_k} d\delta_1 \ldots d\delta_n \partial_j \partial_k \partial_l \partial_m W_{gm} = -\left(\frac{2}{\pi}\right)^{1/2} \frac{1}{t^2} e^{-\nu^2/2}, \tag{A.7a}
\]

\[
\sum_{j,k=1}^{n} \left(1 - \frac{t_j}{t} \right) \sum_{l,l_1,l_2,l_3}^{n} \int_{-\infty}^{\delta_c} d\delta_1 \ldots d\delta_n \partial_j \partial_k \partial_l \partial_{l_1} \partial_{l_2} \partial_{l_3} W_{gm} = -\left(\frac{2}{\pi}\right)^{1/2} \frac{4}{t^3} e^{-\nu^2/2}, \tag{A.7b}
\]

where we have used the result (A.3), and in Eqn. (A.7b) also the identity

\[
\partial^3 c h(\nu) = -\nu e^{-\nu^2/2}. \tag{A.8}
\]

We now see that the result of the path integral depends only on \( \nu \). Putting things together and computing the overall \( \nu \) derivative, we find the result in Eqn. (29).

### A.2 Saddle point calculation

To compute the action of the exponentiated derivative operators, we start by writing the expression in square brackets in Eqn. (29) in terms of its Fourier transform, using the relations

\[
e^{-\nu^2/2} = \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{i\lambda \nu} e^{-\lambda^2/2},
\]

\[
-\nu e^{-\nu^2/2} = \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} (i\lambda) e^{i\lambda \nu} e^{-\lambda^2/2},
\]

\[
\nu^2 e^{-\nu^2/2} = -\int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} (\lambda^2 - 1) e^{i\lambda \nu} e^{-\lambda^2/2},
\]

\[
\left(\frac{\pi}{2}\right)^{1/2} \text{erfc} \left(\frac{\nu}{\sqrt{2}}\right) = \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} \frac{i}{\lambda} e^{i\lambda \nu} e^{-\lambda^2/2}. \tag{A.9}
\]

Together with the identity \( e^{A(-d/d\nu)^\alpha} e^{i\lambda \nu} = e^{A(-i\lambda)^\alpha} e^{i\lambda \nu} \), for constant \( A \) and \( B \), this gives

\[
f(\nu) = \left(\frac{2}{\pi}\right)^{1/2} \nu \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{i\lambda \nu} e^{-\lambda^2/2 + (-i\lambda)^3 \varepsilon_1 / 6 + (-i\lambda)^4 \varepsilon_2 / 24 + \ldots} \mathcal{P}(\lambda) \tag{A.10}
\]

where \( \mathcal{P}(\lambda) \) is the truncated series given by

\[
\mathcal{P}(\lambda) = 1 + \frac{1}{4} i \varepsilon_1 \lambda + \frac{5}{16} \frac{i \varepsilon_1}{\lambda} - \frac{1}{4} \lambda^2 \left(\frac{\varepsilon_1^2}{2} - \frac{\varepsilon_2}{3}\right) + \ldots \tag{A.11}
\]

The integral in eq. (A.10) can be performed using the saddle point approximation. We write it as

\[
f(\nu) = \left(\frac{2}{\pi}\right)^{1/2} \nu \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{\phi(\lambda)}, \tag{A.12}
\]

where

\[
\phi(\lambda) \equiv i \lambda \nu - \frac{1}{2} \lambda^2 + \frac{i \varepsilon_1}{6} \lambda^3 + \frac{\varepsilon_2}{24} \lambda^4 + \ln \mathcal{P}(\lambda) + \ldots \tag{A.13}
\]

The location of the saddle point, \( \lambda = \lambda_* \), is the solution of \( \phi'(\lambda_*) = 0 \), and the saddle point approximation then tells us that

\[
\int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{\phi(\lambda)} = e^{\phi(\lambda_*) (|\phi''(\lambda_*)|)^{-1/2}}, \tag{A.14}
\]

\footnote{We are using a regulator which shifts the pole at \( \lambda = 0 \) in the last expression in Eqn. (A.9), to \( \lambda = -i\alpha \) where \( \alpha \) is real, positive and small.}
(see Appendix D for a discussion of the errors introduced by this approximation). It turns out that in order to obtain \( f(\nu) \) correctly up to order \( \sim \epsilon^2 \nu^2 \), we only need \( \lambda_* \) correct up to order \( \sim \epsilon \nu \). The expression for \( \phi' \) at the relevant order is,
\[
\phi'(\lambda) = i \nu - \lambda + \frac{\epsilon \nu}{2} \lambda^2 + \ldots,
\]
and solving for \( \lambda_* \) perturbatively up to order \( \epsilon \nu \), we find
\[
\lambda_* = i \nu \left[ 1 - \frac{1}{2} \epsilon \nu + \mathcal{O}(\epsilon^2 \nu^2) \right].
\]
This leads to the expression for \( f(\nu) \) in Eqn. (30).

### A.3 Result with full unequal time terms

In the more realistic case of slowly-varying \( \epsilon_n \), we choose to parametrize the coefficients \( G_3 \) and \( G_4 \) (see Eqn. (26)) in a convenient way as follows:
\[
G_3^{(1,0.0)} = \frac{1}{2} \epsilon_1(t) c_1(t) t^{1/2} ; \quad G_3^{(2,0.0)} = -\frac{1}{4} \epsilon_1(t) c_2(t) t^{-1/2},
\]
\[
G_3^{(1,1.0)} = \frac{1}{4} \epsilon_1(t) c_3(t) t^{-1/2} ; \quad G_4^{(1,0.0)} = \frac{1}{2} \epsilon_2(t) c_4(t),
\]
where the coefficients \( c_n(t) \) are smoothly varying functions and depend on the NG model. They are defined in such a way that they all reduce to unity in the toy model defined by Eqn. (28). Fig. 5 shows the behaviour of \( c_1, c_2 \) and \( c_3 \) with \( \sigma^2 \), for the local and equilateral models. The \( \epsilon_n \) and \( c_n \) are independent of redshift by construction, since the linear growth rate \( D(z) \) always drops out in their definitions. Further, the \( c_n \) do not depend on the values of \( f_{NL} \) and \( g_{NL} \). One can then use the definitions of \( \epsilon_1 \) and the \( c_n \) to prove the following useful relations
\[
\frac{d \ln \epsilon_1}{d \ln t} = \frac{3}{2} (c_1 - 1) ; \quad \frac{d \ln c_1}{d \ln t} = 1 - \frac{3}{2} c_1 + \frac{1}{c_1} \left( c_3 - \frac{1}{2} c_2 \right).
\]
The calculation of the mass function for this general case proceeds completely analogously to that for the toy model, apart from a few subtleties which we will discuss later. In this case Eqn. (A.4) is replaced with
\[
f = (\nu \partial_\nu - 2t \partial_t) e^{- (\epsilon_1(t)/3!) \partial_\nu^2 + (\epsilon_2(t)/4!) \partial_T^4 + \ldots} \cdot g(\nu, t)
\]
\[
= \left[ \nu + \frac{1}{3} \frac{d \ln \epsilon_1}{d \ln t} \epsilon_1 \partial_\nu^2 - \frac{1}{12} \frac{d \ln \epsilon_2}{d \ln t} \epsilon_2 \partial_\nu^4 \right] e^{- (\epsilon_1(t)/3!) \partial_\nu^2 + (\epsilon_2(t)/4!) \partial_T^4 + \ldots} \partial_\nu g(\nu, t)
\]
\[
- 2te^{- (\epsilon_1(t)/3!) \partial_\nu^2 + (\epsilon_2(t)/4!) \partial_T^4 + \ldots} \partial_t g(\nu, t),
\]
where the function \( g(\nu, t) \) can be shown to be
\[
g(\nu, t) = \left( \frac{2}{\pi} \right)^{1/2} \left[ \left( \frac{\pi}{2} \right)^{1/2} \text{erf} \left( \frac{\nu}{\sqrt{2}} \right) + \frac{1}{4} \epsilon_1 c_1 e^{-\nu^2/2} + \frac{1}{4} c_3 \left( \frac{1}{4} \epsilon_2 - 2c_2 \right) h(\nu) \right.
\]
\[
- \frac{1}{8} \epsilon_1 \epsilon_2 \nu e^{-\nu^2/2} + \frac{1}{12} \epsilon_2 c_4 \nu e^{-\nu^2/2} + \ldots \right],
\]
The expression in Eqn. (A.19) can be evaluated analogously to Eqn. (29), since the additional derivatives pose no conceptual difficulty. The result of the saddle point calculation, correct up to quadratic order assuming \( \epsilon \nu^3 \sim \mathcal{O}(1) \), and after using the relations (A.18), is given in Eqn. (31).
Figure 5: The derivative coefficients $c_1$ (panel (a)), $c_2$ (panel (b)) and $c_3$ (panel (c)), as a function of $\sigma^2$, for local and equilateral NG models. These quantities are independent of redshift and the NG amplitudes $f_{\text{NL}}$ and $g_{\text{NL}}$. The axes are logscale.

B Truncation of the perturbative series

B.1 Analysis of “transition points”

In this appendix we analyse the consistency of our series truncation at various mass scales. When $\epsilon \nu^3 \simeq 1$, in the polynomial in (32) we retain the terms $\epsilon \nu \simeq \nu^{-2}$, $(\epsilon \nu^{-1}, \epsilon^2 \nu^2) \simeq \nu^{-4}$, and we discard $(\epsilon \nu^{-3}, \epsilon^2, \epsilon^3 \nu^3) \simeq \nu^{-6}$. It would seem that our expression is then correct up to order $\sim \nu^{-4}$. However, the terms discarded in the exponential have the form $\exp(\mathcal{O}(\epsilon^3 \nu^5)) \sim \exp(\mathcal{O}(\nu^{-4})) \sim 1 + \mathcal{O}(\nu^{-4})$. The error we are making is thus of the same order as the smallest terms we are retaining, and it therefore makes sense to also ignore all the terms of order $\sim \nu^{-4}$ which we computed in the polynomial. The consistent expression when $\epsilon \nu^3 \simeq 1$ is then given by

$$f \sim e^{-\frac{1}{2} \epsilon^2 (1 + \epsilon \nu + \epsilon^2 \nu^2)} \left[1 + \epsilon \nu + \mathcal{O}(\nu^{-4})\right].$$  \hfill (B.1)

Clearly, similar arguments can be applied at smaller scales where, e.g. one might have $\epsilon \nu^3 \simeq \nu^{-1}, \nu^{-2}$, etc. It is then important to ask which mass scales correspond to these “transition points”. In Fig. 6 we plot $\nu(M, z)$ given by Eqn. (25) in an observationally interesting mass range, for three different redshifts. The horizontal lines mark the transition points where $\epsilon \nu^3$ becomes equal to (from top to bottom) $1, \nu^{-1}, \nu^{-2}, \nu^{-3}$, $\nu^{-4}$ and $\nu^{-5}$. We fix $\epsilon = 1/300$ which follows from the fact that in the local model with $f_{\text{NL}} = 100$ we have $\epsilon_1 \simeq 0.02$ (see Fig. 1), and the expression for $f(\nu, M)$ contains the quantity $\epsilon_1/6$ in the exponential. From the intersections of the horizontal lines with the curves, we see that different transition points are relevant.
Figure 6: $\nu \equiv \delta_c(z)/\sigma(M)$ in the range $5 \cdot 10^{13} < (M/h^{-1} M_{\odot}) < 5 \cdot 10^{15}$ for three different redshifts, with $\epsilon = 1/300$. The solid, long dashed and short dashed curves correspond to redshifts $z = 1, 0.5$ and 0 respectively. The horizontal lines mark the transition points where $\epsilon \nu^3$ becomes equal to (from top to bottom) $1, \nu^{-1}, \nu^{-2}$, $\nu^{-3}$, $\nu^{-4}$ and $\nu^{-5}$.

at different redshifts, and their locations also obviously depend on the value of $\epsilon$. For example, we find that the transition point where $\epsilon \nu^3 \simeq \nu^{-2}$, remains accessible even when $\epsilon$ is an order of magnitude smaller (with $\epsilon \simeq 1/3000$, this transition occurs at $\nu \simeq 4.96$). The transitions at $\epsilon \nu^3 \simeq 1, \nu^{-1}$ on the other hand, are not accessible for this level of NG. The transition at $\epsilon \nu^3 \simeq \nu^{-2}$ is therefore observationally very interesting.

We will now discuss in some detail the truncation of our expression for $f$, at various transition points. The goal is to try and settle on a single expression which is valid over a wide range of scales (i.e. across several transition points). This can then be applied without worrying about truncation inconsistencies. Of course, the order of the discarded terms will then depend on the particular transition point being considered, leading to a scale dependent theoretical error.

**B.1.1 $\epsilon \nu^3 \simeq \nu^{-1}$**

At this transition point, the terms we retain in the exponential are

$$\epsilon \nu^3 \simeq \nu^{-1}, \quad \epsilon^2 \nu^4 \simeq \nu^{-4},$$

while discarding $O(\epsilon^3 \nu^5) = O(\nu^{-7})$. In the polynomial meanwhile, we retain

$$\epsilon \nu \simeq \nu^{-3}, \quad \epsilon \nu^{-1} \simeq \nu^{-5}, \quad \epsilon^2 \nu^2 \simeq \nu^{-6},$$

while discarding

$$O(\epsilon \nu^{-3}) = O(\nu^{-7}) \quad ; \quad O(\epsilon^2) = O(\nu^{-8}) \quad ; \quad O(\epsilon^3 \nu^3) = O(\nu^{-9}).$$

Our expression (32) therefore retains all terms correctly up to order $\sim \nu^{-6}$, and is consistent. With some foresight however, it turns out to be more convenient to degrade this expression somewhat by also discarding the polynomial quadratic term $\epsilon^2 \nu^2 \simeq \nu^{-6}$. The remaining expression,

$$f \sim e^{-\frac{\epsilon}{\nu^3} \nu^2 (1+\epsilon \nu + \epsilon^2 \nu^2) \left[1 + \epsilon \nu + \frac{\epsilon}{\nu} + O\left(\nu^{-6}\right)\right]}, \quad \text{(B.2)}$$

is also consistent at this transition point, and has a form which is identical to the ones we will see next.

**B.1.2 $\epsilon \nu^3 \simeq \nu^{-2}$**

As we mentioned earlier, this transition point is observationally quite interesting. The terms we retain in the exponential are

$$\epsilon \nu^3 \simeq \nu^{-2}, \quad \epsilon^2 \nu^4 \simeq \nu^{-6},$$
while discarding $\mathcal{O}(\epsilon^3 \nu^5) = \mathcal{O}(\nu^{-10})$, and in the polynomial we retain

$$
\epsilon \nu \simeq \nu^{-4}; \quad \epsilon \nu^{-1} \simeq \nu^{-6}; \quad \epsilon^2 \nu^2 \simeq \nu^{-8},
$$

while discarding

$$
\mathcal{O}(\epsilon \nu^{-3}) = \mathcal{O}(\nu^{-8}); \quad \mathcal{O}(\epsilon^3) = \mathcal{O}(\nu^{-10}); \quad \mathcal{O}(\epsilon^3 \nu^3) = \mathcal{O}(\nu^{-12}).
$$

This time we see that the term $\epsilon \nu^{-3}$ has become as important as the quadratic term $\epsilon^2 \nu^2$ in the polynomial, and to be consistent we should discard the quadratic term. The expansion should read

$$
f \sim e^{-\frac{1}{2} \nu^2 (1+\epsilon + \epsilon^2 \nu^2)} \left[ 1 + \epsilon \nu + \frac{\epsilon}{\nu} + \mathcal{O}(\nu^{-8}) \right].
$$

(B.3)

### B.1.3 $\epsilon \nu^3 \simeq \nu^{-3}$

A similar analysis as above shows that at this stage $\epsilon \nu^{-3} \simeq \nu^{-9} > \epsilon^2 \nu^2$, and a consistent expression again requires dropping the quadratic term in the polynomial, leaving

$$
f \sim e^{-\frac{1}{2} \nu^2 (1+\epsilon + \epsilon^2 \nu^2)} \left[ 1 + \epsilon \nu + \frac{\epsilon}{\nu} + \mathcal{O}(\nu^{-9}) \right].
$$

(B.4)

### B.1.4 $\epsilon \nu^3 \simeq \nu^{-4}$ and smaller

Beyond this point, the term $\epsilon \nu^{-3}$ which we discard in the polynomial, becomes comparable or larger than the quadratic term of the exponential as well, and a consistent expression becomes

$$
f \sim e^{-\frac{1}{2} \nu^2 (1+\epsilon \nu^2)} \left[ 1 + \epsilon \nu + \frac{\epsilon}{\nu} + \ldots \right]
$$

(B.5)

The parametric order of the terms now discarded, depends on the exact relation between $\epsilon \nu^3$ and $\nu^{-1}$.

Finally, note that the error introduced by setting $\nu \to \nu_0$ where $\nu_0$ is defined using the variance of a Gaussian field, was estimated in section 3 as $\mathcal{O}(\epsilon^2)$. When $\epsilon \nu^3 \simeq 1$, this error is of order $\mathcal{O}(\epsilon^3 \nu^3)$ and can therefore be consistently ignored. It is not hard to see that at all lower transition points, this error continues to be comparable to or smaller than the largest terms being discarded, and can hence be consistently ignored. This finally leads to the conclusion stated in the main text.

### B.2 Truncation in MR and LMSV results

Since the mass scale where $\epsilon \nu^3 \simeq 1$ is on the border of the observed mass window (for galaxy cluster observations), even at high redshifts, let us therefore directly look at the case $\epsilon \nu^3 \simeq \nu^{-2}$ which, as we saw, is accessible over a wide range of redshifts for $\epsilon \sim 10^{-2}$, and at high redshifts also for $\epsilon \sim 10^{-3}$. In this case the terms MR and LMSV retain have magnitudes

$$
\epsilon \nu^3 \simeq \nu^{-2}; \quad \epsilon \nu \simeq \nu^{-4}; \quad \epsilon \nu^{-1} \simeq \nu^{-6},
$$

and terms like $\epsilon \nu^{-3} \simeq \nu^{-8}$ are discarded. We know from our expression however, that $\epsilon \nu^3$ appears in the exponential, and therefore leads to terms like $(\epsilon \nu^3)^2 \simeq \nu^{-4}$ and $(\epsilon \nu^3)^3 \simeq \nu^{-6}$ when the exponential is expanded, which are of the same order as the terms retained in (39). The exponential also contributes a term $\epsilon^2 \nu^4 \simeq \nu^{-6}$, which in fact involves the trispectrum of NG, again at the order retained by MR and LMSV. The error in the expression (39) when $\epsilon \nu^3 \simeq \nu^{-2}$, is therefore $\mathcal{O}(\epsilon \nu)$ (A similar analysis shows that the error at transition point where $\epsilon \nu^3 \simeq \nu^{-1}$, is $\mathcal{O}(\nu^{-2}) > \mathcal{O}(\epsilon \nu)$.

From a purely parametric point of view, the situation for MR and LMSV improves as $\nu$ is decreased further, and the expression (39) as it stands, becomes exactly consistent (in the sense discussed in the previous subsection, see below Eqn. (33)) when $\epsilon \nu^3 \simeq \nu^{-5}$, because at this stage $\epsilon \nu^{-1} \simeq \nu^{-9}$ while $(\epsilon \nu^3)^2 \simeq \nu^{-10}$ and $\epsilon^2 \nu^4 \simeq \nu^{-12}$, and hence the exponential only contributes a single linear term $\epsilon \nu^3$. More importantly, LMSV’s expression also has errors due to the absence of the unequal time terms discussed earlier, which are of order
~ εν and can be dominant over the others. For the intermediate transitions, the analysis shows that when \(\epsilon \nu^3 \simeq \nu^{-3}\), the error in (39) is \(\mathcal{O}(\nu^{-6}) \gg \mathcal{O}(\epsilon \nu^{-1})\), and when \(\epsilon \nu^3 \simeq \nu^{-4}\), the error is \(\mathcal{O}(\epsilon \nu^{-1})\). This should be compared with our result (34), in which the error (at least on large scales) is always parametrically smaller than the smallest terms we retain.

C Hierarchy of terms in Eqn. (A.4)

Here we argue why the hierarchy of terms ordered by powers of \(\nu^{-2}\) emerges on expanding the exponentiated derivative operators in Eqn. (A.4). Focusing on terms involving the 3-point correlator, one sees that a generic term in the expansion contains some powers of \((\epsilon_1 t^{3/2})\), multiplying an \(n\)-dimensional integral containing some summations \(\sim \sum_{j_1,j_2,\ldots=1}^{n} (1-t_{j_1}/t)^{p_1} (1-t_{j_2}/t)^{p_2} \ldots \partial_{j_1} \partial_{j_2} \ldots\), and also some summations over “free” derivatives \(\sim \sum_{n,k_2=1}^{n} \partial_{k_1} \partial_{k_2} \ldots\), all of this acting on \(W^{\text{rgm}}\). More precisely, the structure of the terms is

\[
\sim (\epsilon_1 t^{3/2})^m \sum_{j_1,\ldots,j_{3m}} \int_{-\infty}^{\delta} d\delta_1 \ldots d\delta_n [(1-t_{j_1}/t) \ldots (1-t_{j_m}/t)]^p \left[(1-t_{j_{m+1}/t}) \ldots (1-t_{j_{3m}/t})\right]^q \times [(1-t_{j_{m+1}/t}) \ldots (1-t_{j_{3m}/t})]^r \partial_{j_1} \ldots \partial_{j_{3m}} W^{\text{rgm}},
\]

for \(m \geq 1\) and non-negative \(p, q, r\) such that not all three are zero. The terms we have considered in the text are \((m, p, q, r) = (1, 1, 0, 0), (1, 1, 1, 0), (1, 2, 0, 0)\) and \((2, 1, 0, 0)\). We have already discussed how the “free” derivatives can be pulled out of the integral and converted to \(\partial\). For the “non-free” derivatives, we see that what is important is the total number of \((1-t_j/t)\) factors accompanying these derivatives. For example, the \((1,1,1,0)\) term in Eqn. (A.5c) has the same structure as the \((1,2,0,0)\) term in Eqn. (A.5b) – the effect of \(\sum_{j,k} (1-t_j/t)(1-t_k/t)\partial_j \partial_k\), up to numerical factors, is identical to that of \(\sum_{j,k} (1-t_j/t)^2 \partial_j \partial_k\). This is expected to be true also with higher numbers of non-free derivatives.

It is then possible to understand the hierarchy of terms by only considering terms containing \(\sum_j (1-t_j/t)^p \partial_j\), and no other non-free derivatives. The basic object to study now becomes

\[
\sum_{j} (1-t_j/t) \int d\delta_1 \ldots d\delta_n \partial_j W^{\text{rgm}},
\]

which in the continuum limit can be shown to reduce to the integral

\[
g(0) \left(\frac{\nu^2}{2}\right) = \int_0^1 \frac{dy}{y^{3/2}} (1-y)^{1/2} e^{-\nu^2/2y} = \frac{\sqrt{\pi}}{2} \Gamma\left(-\frac{1}{2}, \frac{\nu^2}{2}\right). \tag{C.2}
\]

Notice the similarity with the function \(h(\nu)\) in Eqn. (A.8), which of course is not accidental given the definitions of these objects. It is now easy to check that increasing the powers of \((1-t_j/t)\) in the summation amounts to increasing the powers of \((1-y)\) in \(g(0)\). We are then comparing (with \(A = \nu^2/2\)) \(g(0)(A)\) with \(g(p)(A)\) where

\[
g(p)(A) \equiv \int_0^1 \frac{dy}{y^{3/2}} (1-y)^{1/2+p} e^{-A/y}. \tag{C.3}
\]

Starting with \(p = 1\) and manipulating the integrals, it is straightforward to establish the recurrence

\[
g(p+1)(A) = g(p)(A) - \int_A^{\infty} d\tilde{A} g(p)(\tilde{A}). \tag{C.4}
\]

The argument is now almost complete. We know that for large \(A = \nu^2/2\), we have \(\Gamma(n, A) = e^{-A} A^{n-1}(1 + \mathcal{O}(A^{-1}))\). Hence \(g(0)(A) = (\sqrt{\pi}/2) A^{-3/2} e^{-A} (1 + \mathcal{O}(A^{-1}))\), and its integral from \(A\) to \(\infty\) gives a leading term proportional to \(\Gamma(-3/2, A) = e^{-A} A^{-5/2} (1 + \mathcal{O}(A^{-1}))\). The pattern is now clear: \(g(p)(A) \sim A^{-3/2+p} e^{-A} (1 + \mathcal{O}(A^{-1}))\), and since \(A = \nu^2/2\), this explains the hierarchy of terms in powers of \(\nu^{-2}\), in Eqn. (A.4).
D The saddle point approximation

In this appendix we discuss the saddle point approximation of the integrals of the type appearing in section 3.1, and estimate the error it induces. We will argue that the errors introduced by the saddle point approximation are much smaller than those due to truncating the perturbative series in the small parameters $\epsilon$ and $\nu^{-1}$. For an introduction to the saddle point approximation see Ref. [45]. Since we only wish to discuss the saddle point method in this appendix, we will ignore here the complications introduced by the unequal time correlators, i.e. in Eqn. (A.10) we set $\mathcal{P}(\lambda) = 1$. We will also work here to first order in $\epsilon \nu$. The extension to a more general case is straightforward and the result is given by (31) as described in section 3.1. We begin with expression (A.10):

$$f(\nu) = \left(\frac{2}{\pi}\right)^{1/2} \nu \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{g(\lambda)},$$  \hspace{1cm} (D.1)

where $g(\lambda) \equiv i \nu \lambda - \lambda^2/2 + (-i\lambda)^3\varepsilon_1/6 + \mathcal{O}(\epsilon^2 \lambda^4)$.

We first find the location of a saddle point $\lambda_*$ of the function $g(\lambda)$, by perturbatively solving $g'(\lambda_*) = 0$ using $\epsilon \nu$ as the small parameter and demanding $g''(\lambda_*) < 0$. The first-order solution is

$$\lambda_* = i \nu (1 - \varepsilon_1 \nu/2 + \mathcal{O}(\epsilon^2 \nu^2))^2, \hspace{1cm} (D.2)$$

$$g(\lambda_*) = -\frac{\nu^2}{2} (1 - \frac{1}{3} \varepsilon_1 \nu + \mathcal{O}(\epsilon^2 \nu^2))^2, \hspace{1cm} (D.3)$$

$$g''(\lambda_*) = -1 - \varepsilon_1 \nu + \mathcal{O}(\epsilon^2 \nu^2). \hspace{1cm} (D.4)$$

The saddle point approximation consists roughly of performing a Taylor expansion of $g(\lambda)$ to second order around $\lambda_*$ in the integrand of (D.1) and performing the resulting Gaussian integral. We will carry this out explicitly below. The saddle point prescription will give a good approximation to the integral as long as $g(\lambda)$ attains a global maximum at $\lambda_*$ (along the contour of integration); this is indeed our case since the integrand in Eqn. (D.1) will be nearly a Gaussian centered at $\lambda_*$ in the complex plane.

Notice that $\text{Im} \lambda_* \neq 0$, requiring a deformation of the contour of integration such that it passes through $\lambda_*$. The deformation of the path of integration can be performed by taking a closed contour formed by four pieces: The real axis $C_1$, the line $\text{Im} \lambda = \text{Im} \lambda_*$ which we call here $-C_2$, and the closures of this contour at positive and negative infinity. The integral in this closed contour must be zero, and since the integral on the closures of the contour at infinity can be assumed to vanish, we have $\int_{C_1} = \int_{C_2}$. Therefore $C_2$ is the desired deformation of the contour which passes through $\lambda_*$. We can then make a series of approximations in the integral (D.1), which we discuss below.

$$\int_{C_1} \frac{d\lambda}{\sqrt{2\pi}} e^{g(\lambda)} \approx \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{g(\lambda_*) + g''(\lambda_*)(\lambda - \lambda_*)^2/2}$$

$$= e^{g(\lambda_*)} \left(-g''(\lambda_*)\right)^{-1/2}$$

$$= e^{-\frac{\nu^2}{2} (1 - \varepsilon_1 \nu/3 + \mathcal{O}(\epsilon^2 \nu^2)) \left(1 + \varepsilon_1 \nu + \mathcal{O}(\epsilon^2 \nu^2)\right)^{-1/2}}. \hspace{1cm} (D.5)$$

Here the integrations are performed along the deformed contour.

In order to estimate the errors induced by the approximation done in equation (D.5), one can keep higher

\footnote{Technically, one should also require that $\text{Im} g(\lambda)$ be nearly constant along the deformed contour for the saddle point approximation to work. In our case one can show that $\text{Im} g$ will be suppressed by $\epsilon$. This and all errors induced by the saddle point are accounted for in equation (D.6).}
orders in the Taylor expansion of the function in the exponential:

\[
\int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{g(\lambda)} \approx \int_{-\infty}^{\infty} \frac{d\lambda}{\sqrt{2\pi}} e^{g(\lambda_*)+g''(\lambda_*)(\lambda-\lambda_*)^2/2+g^{(3)}(\lambda_*)(\lambda-\lambda_*)^3/6+g^{(4)}(\lambda_*)(\lambda-\lambda_*)^4/24+...}
\]

\[
\approx e^{g(\lambda_*)} ( -g''(\lambda_*) )^{-1/2}
\]

\[
+ \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \left( \frac{1}{6} g^{(3)}(\lambda_*) z^3 + \frac{1}{12} [g^{(3)}(\lambda_*)]^2 z^6 + \frac{1}{24} g^{(4)}(\lambda_*) z^4 + \ldots \right) e^{g(\lambda_*)+g''(\lambda_*) z^2/2}
\]

\[
= e^{g(\lambda_*)} ( -g''(\lambda_*) )^{-1/2} ( 1 + \mathcal{O}(\epsilon^2) ) . \tag{D.6}
\]

Here we used the fact that \( g^{(3)}(\lambda_*) = \mathcal{O}(\epsilon) \) and \( g^{(4)}(\lambda_*) = \mathcal{O}(\epsilon^2) \). The integrals in the second equality of this derivation can be computed analytically, which allows one to go to arbitrary accuracy with the saddle point technique. Notice that the results of these integrations are of higher order than the terms we retain. In the main text, where the integral contains also a polynomial \( \mathcal{P}(\lambda) \) one can again compute the errors via similar Taylor expansions. These errors can be shown to be of order \( \mathcal{O}(\epsilon^2) \), comparable to other terms which we ignore.

One can also estimate the errors introduced by our approximations by using the following toy model in which everything is computable: Take the 3-point cumulant \( \varepsilon_1 \) to be different from zero and all higher order cumulants \( \varepsilon_n \) for \( n \geq 2 \) to be zero\(^{10}\). For such a model the integral is

\[
\int_{-\infty}^{\infty} d\lambda e^{i\omega\lambda-\lambda^2/2+(-i\lambda)^3\varepsilon_1/6} \approx \left( \frac{2\pi}{\sqrt{1+2\varepsilon_1\nu}} \right)^{1/2} \exp \left( \frac{1 - \sqrt{1 + 2\varepsilon_1\nu} + \varepsilon_1\nu(3 - 2\sqrt{1 + 2\varepsilon_1\nu})}{3\varepsilon_1^2} \right) . \tag{D.7}
\]

In the r.h.s of this equation we have used the saddle point approximation but have made no expansion in \( \epsilon \). By comparing the numerical integration of the l.h.s. with the expression on the r.h.s. (panel (a) of Fig. 7), one can see that the errors introduced by the saddle point approximation are indeed of order \( \epsilon^2 \) as indicated by (D.6). On the other hand, one can use the numerical integration of the left hand side of this equation and

\(^{10}\)This toy model is inconsistent because if the third cumulant is different from zero, then all higher cumulants must also be different from zero. We use it here only to estimate how good the saddle point prescription is in approximating an integral, and compare it with errors induced by a perturbative expansion in \( \epsilon \).
compare it with the approximation (D.5) (panel (b) of Fig. 7), to see that the biggest error is of order $\epsilon^2 \nu^2$ induced by the fact that we perform a perturbative expansion in $\epsilon \nu$. Notice that here we considered only the leading order in $\nu$ and ignored unequal time correlators, while in the main text we present a result which is more precise (to next to leading order in $\epsilon \nu$) and complete (using the excursion set formalism rigorously).

References

[1] 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. D 77 (2008) 123514 [arXiv:0710.4560 [astro-ph]].

[2] S. Matarrese and L. Verde, “The effect of primordial non-Gaussianity on halo bias,” Astrophys. J. 677 (2008) L77 [arXiv:0801.4826 [astro-ph]].

[3] A. Slosar, C. Hirata, U. Seljak, S. Ho and N. Padmanabhan, “Constraints on local primordial non-Gaussianity from large scale structure,” JCAP 0808 (2008) 031 [arXiv:0805.3580 [astro-ph]].

[4] E. Komatsu et al., “Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation,” arXiv:1001.4538 [astro-ph.CO].

[5] B. Sartoris, S. Borgani, C. Fedeli, S. Matarrese, L. Moscardini, P. Rosati and J. Weller, “The potential of X-ray cluster surveys to constrain primordial non-Gaussianity,” arXiv:1003.0841 [astro-ph.CO].

[6] C. Carbone, L. Verde and S. Matarrese, “Non-Gaussian halo bias and future galaxy surveys,” Astrophys. J. 684, L1 (2008) [arXiv:0806.1950 [astro-ph]].

[7] C. Carbone, O. Mena and L. Verde, “Cosmological Parameters Degeneracies and Non-Gaussian Halo Bias,” JCAP 1007, 020 (2010) [arXiv:1003.0456 [astro-ph.CO]].

[8] C. Cunha, D. Huterer and O. Dore, “Primordial non-Gaussianity from the covariance of galaxy cluster counts,” arXiv:1003.2416 [astro-ph.CO].

[9] E. Sefusatti, “1-loop Perturbative Corrections to the Matter and Galaxy Bispectrum with non-Gaussian Initial Conditions,” Phys. Rev. D 80 (2009) 123002 [arXiv:0905.0717 [astro-ph.CO]].

[10] R. Jimenez and L. Verde, “Implications for Primordial Non-Gaussianity ($f_{NL}$) from weak lensing masses of high-z galaxy clusters,” Phys. Rev. D 80, 127302 (2009) [arXiv:0909.0403 [astro-ph.CO]].

[11] L. Verde, “Non-Gaussianity from Large-Scale Structure Surveys,” arXiv:1001.5217 [astro-ph.CO].

[12] V. Desjacques and U. Seljak, “Primordial non-Gaussianity from the large scale structure,” arXiv:1003.5020 [astro-ph.CO].

[13] J. E. Gunn and J. R. I. Gott, “On the infall of matter into cluster of galaxies and some effects on their evolution,” Astrophys. J. 176, 1 (1972).

[14] W. H. Press and P. Schechter, “Formation of galaxies and clusters of galaxies by selfsimilar gravitational condensation,” Astrophys. J. 187, 425 (1974).

[15] J. R. Bond, S. Cole, G. Efstathiou and N. Kaiser, “Excursion set mass functions for hierarchical Gaussian fluctuations,” Astrophys. J. 379, 440 (1991).

[16] 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].

[17] M. LoVerde, A. Miller, S. Shandera and L. Verde, “Effects of Scale-Dependent Non-Gaussianity on Cosmological Structures,” JCAP 0804 (2008) 014 [arXiv:0711.4126 [astro-ph]].

[18] R. K. Sheth and G. Tormen, “An Excursion Set Model Of Hierarchical Clustering : Ellipsoidal Collapse And The Moving Barrier,” Mon. Not. Roy. Astron. Soc. 329, 61 (2002) [arXiv:astro-ph/0105113].
[19] M. Maggiore and A. Riotto, “The Halo Mass Function from the Excursion Set Method. I. First principle derivation for the non-Markovian case of Gaussian fluctuations and generic filter,” arXiv:0903.1249 [astro-ph.CO].
[20] M. Maggiore and A. Riotto, “The halo mass function from the excursion set method. II. The diffusing barrier,” arXiv:0903.1250 [astro-ph.CO].
[21] M. Maggiore and A. Riotto, “The halo mass function from the excursion set method. III. First principle derivation for non-Gaussian theories,” arXiv:0903.1251 [astro-ph.CO].
[22] P. Valageas, “Mass function and bias of dark matter halos for non-Gaussian initial conditions,” arXiv:0906.1042 [astro-ph.CO].
[23] J. M. Bardeen, J. R. Bond, N. Kaiser and A. S. Szalay, “The Statistics Of Peaks Of Gaussian Random Fields,” Astrophys. J. 304, 15 (1986).
[24] R. K. Sheth, H. J. Mo and G. Tormen, “Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes,” Mon. Not. Roy. Astron. Soc. 323, 1 (2001) [arXiv:astro-ph/9912025].
[25] T. Y. Lam and R. K. Sheth, “Halo abundances in the $f_{\text{NL}}$model,” arXiv:0905.1702 [astro-ph.CO].
[26] N. Sugiyama, “Cosmic background anistropies in CDM cosmology,” Astrophys. J. Suppl. 100, 281 (1995) [arXiv:astro-ph/9412025].
[27] U. Seljak and M. Zaldarriaga, “A Line of Sight Approach to Cosmic Microwave Background Anisotropies,” Astrophys. J. 469 (1996) 437 [arXiv:astro-ph/9603033].
[28] A. Lewis, A. Challinor and A. Lasenby, “Efficient Computation of CMB anisotropies in closed FRW models,” Astrophys. J. 538 (2000) 473 [arXiv:astro-ph/9911177].
[29] D. Babich, P. Creminelli and M. Zaldarriaga, “The shape of non-Gaussianities,” JCAP 0408 (2004) 009 [arXiv:astro-ph/0405356].
[30] H. Lyth, C. Ungarelli and D. Wands, “The primordial density perturbation in the curvaton scenario,” Phys. Rev. D 67 (2003) 023503 [arXiv:astro-ph/0208055].
[31] N. Bartolo, S. Matarrese and A. Riotto, “On non-Gaussianity in the curvaton scenario,” Phys. Rev. D 69 (2004) 043503 [arXiv:hep-ph/0309033].
[32] G. Dvali, A. Gruzinov and M. Zaldarriaga, “A new mechanism for generating density perturbations from inflation,” Phys. Rev. D 69 (2004) 023505 [arXiv:astro-ph/0303591].
[33] M. Alishahiha, E. Silverstein and D. Tong, “DBI in the sky,” Phys. Rev. D 70, 123505 (2004) [arXiv:hep-th/0404084].
[34] N. Arkani-Hamed, P. Creminelli, S. Mukohyama and M. Zaldarriaga, “Ghost Inflation,” JCAP 0404, 001 (2004) [arXiv:hep-th/0402100].
[35] P. Creminelli, “On non-Gaussianities in single-field inflation,” JCAP 0310, 003 (2003) [arXiv:astro-ph/0306122].
[36] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark and M. Zaldarriaga, “Limits on non-Gaussianities from WMAP data,” JCAP 0605 (2006) 004 [arXiv:astro-ph/0509029].
[37] J. Zinn-Justin, “Quantum field theory and critical phenomena,” Int. Ser. Monogr. Phys. 113 (2002) 1.
[38] S. Chandrasekhar, “Stochastic problems in physics and astronomy,” Rev. Mod. Phys. 15, 1 (1943).
[39] T. Giannantonio and C. Porciani, “Structure formation from non-Gaussian initial conditions: multivariate biasing, statistics, and comparison with N-body simulations,” arXiv:0911.0017 [astro-ph.CO].
[40] C. Carbone et al., “The properties of the dark matter halo distribution in non-Gaussian scenarios,” Nucl. Phys. Proc. Suppl. 194, 22 (2009).
[41] F. Bernardeau, S. Colombi, E. Gaztanaga and R. Scoccimarro, “Large-scale structure of the universe and cosmological perturbation theory,” Phys. Rept. 367, 1 (2002) [arXiv:astro-ph/0112551].

[42] B. Robertson, A. Kravtsov, J. Tinker and A. Zentner, “Collapse Barriers and Halo Abundance: Testing the Excursion Set Ansatz,” Astrophys. J. 696 (2009) 636 [arXiv:0812.3148 [astro-ph]].

[43] R. K. Sheth and R. van de Weygaert, “A hierarchy of voids: Much ado about nothing,” Mon. Not. Roy. Astron. Soc. 350, 517 (2004) [arXiv:astro-ph/0311260].

[44] M. Kamionkowski, L. Verde and R. Jimenez, “The Void Abundance with Non-Gaussian Primordial Perturbations,” JCAP 0901, 010 (2009) [arXiv:0809.0506 [astro-ph]].

[45] A. Erdélyi, “Asymptotic Expansions,” New York, NY: Dover (1956) 108pp.