Excursion Sets and Non-Gaussian Void Statistics

Guido D'Amico[a,d,e], Marcello Musso[b], Jorge Noreña[c,d,e], Aseem Paranjape[b]

[a] *Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, New York, NY 10003, USA*

[b] *Abdus Salam International Centre for Theoretical Physics Strada Costiera 11, 34151, Trieste, Italy*

[c] *Institut de Ciències del Cosmos (ICC), Universitat de Barcelona (IEEC-UB), Martí Franquès 1, E08028 Barcelona, Spain*

[d] *SISSA, via Bonomea 265, 34136 Trieste, Italy*

[e] *INFN - Sezione di Trieste, 34151 Trieste, Italy*

Abstract

Primordial non-Gaussianity (NG) affects the large scale structure (LSS) of the universe by leaving an imprint on the distribution of matter at late times. Much attention has been focused on using the distribution of collapsed objects (i.e. dark matter halos and the galaxies and galaxy clusters that reside in them) to probe primordial NG. An equally interesting and complementary probe however is the abundance of extended *underdense* regions or *voids* in the LSS. The calculation of the abundance of voids using the excursion set formalism in the presence of primordial NG is subject to the same technical issues as the one for halos, which were discussed e.g. in Ref. [51]. However, unlike the excursion set problem for halos which involved random walks in the presence of one barrier $\delta_{\rm c}$, the void excursion set problem involves *two* barriers $\delta_{\rm v}$ and $\delta_{\rm c}$. This leads to a new complication introduced by what is called the "void-in-cloud" effect discussed in the literature, which is unique to the case of voids. We explore a path integral approach which allows us to carefully account for all these issues, leading to a rigorous derivation of the effects of primordial NG on void abundances. The void-in-cloud issue in particular makes the calculation conceptually rather different from the one for halos. However, we show that its final effect can be described by a simple yet accurate approximation. Our final void abundance function is valid on larger scales than the expressions of other authors, while being broadly in agreement with those expressions on smaller scales.

1 Introduction

A striking feature of the large scale structure (LSS) of the universe which has emerged from studies over the last few decades is the presence of a filamentary network or cosmic web in the matter distribution, with galaxies distributed along filaments which surround large apparently empty regions termed *voids* [1, 2, 3, 4, 5, 6]. A considerable amount of analytical and numerical effort has gone into understanding the nature of the cosmic web (see e.g. Refs. [7, 8, 9, 10], for a review see Ref. [11]). Apart from being intrinsically interesting, the LSS has the potential to be a powerful probe of cosmology since it is sensitive to both the expansion history as well as the initial conditions (in particular the physics of inflation). The advent of large galaxy surveys has realised this potential considerably over the last decade or so [12, 13, 14], with ongoing and upcoming surveys set to significantly increase the precision of the LSS as a cosmological tool (see e.g. Refs. [15, 16, 17]).

Our focus in this work is on probing the physics of the early universe. In the current understanding of structure formation, the present LSS has its seeds in the statistics of the tiny primordial curvature inhomogeneities generated in the very early universe, during the rapidly expanding inflationary phase. The simplest

E-mail: gda2@nyu.edu, musso@ictp.it, jorge.norena@icc.ub.edu, aparanja@ictp.it

models of inflation involving a single "slowly rolling" scalar field predict that the statistics of these primordial inhomogeneities are almost Gaussian [18, 19]. Constraining the amount of *non*-Gaussianity (NG) in the primordial distribution therefore provides a unique window into the physics of inflation. Until recently, the cosmic microwave background (CMB) radiation was considered the standard tool for constraining primordial NG, since inhomogeneities at the CMB epoch are small and the physics can be described by a perturbative treatment. In terms of the standard parametrization of the NG, the CMB constraints for the local model translate to $-10 < f_{NL}^{loc} < +74$ [20]. (For a brief introduction to primordial NG see Appendix A.) Over the last several years, however, the LSS has been shown to be an equally constraining and moreover complementary probe of primordial NG. For example, from Ref. [21] one finds $-29 < f_{NL}^{loc} < +69$, already comparable with the CMB constraints, with precisions of order $\Delta f_{NL}^{loc} \sim 10$ [22] and $\Delta f_{NL}^{loc} \sim 1$ [23] being claimed for future surveys. These constraints and forecasts rely on the statistics of massive collapsed dark matter halos and the galaxies and galaxy clusters that reside in them, chiefly using three tools – the scale dependent bias in the galaxy power spectrum [24, 25], the galaxy bispectrum [26] and the number density of collapsed objects (mass function) [27, 28]. For recent reviews, see Refs. [29, 30].

On the other hand, the voids found in the LSS are also potentially interesting probes of NG. While any given galaxy survey is expected to contain fewer voids than halos, the effect of NG on void abundances is opposite to that on halo abundances (see below). Whereas a positive $f_{\rm NL}$ will enhance the abundance of halos, it will reduce the abundance of voids and vice versa, making voids a complementary probe of NG. Although the literature contains several definitions of what a void exactly means (see e.g. Ref. [31] for a recent review and thorough comparison), the basic physical picture is that of a large expanding region which is underdense compared to the background. For a review of the structure and dynamics of cosmic voids, see Ref. [32]. From the point of view of using voids as statistical probes, detailed numerical studies [33, 34, 35] and analyses [36, 37] seem to indicate that the excursion set formalism [38, 39, 40], combined with simplified analytical models which track the underdensity of dark matter in a region, provides a good starting point. More precisely, the spherical ansatz [41, 42, 43, 36] which we discuss later in a bit more detail has been found to be a very useful tool when studying voids. (See also Refs. [44, 45] for a different approach and application.)

As one might expect from experience with halo abundances, analytical treatments of void abundances are subject to some caveats. Firstly, by choosing to describe voids using underdensities in the *dark matter* distribution, one is a step removed from actual observations which involve *galaxies*. As pointed out by Furlanetto & Piran [46], the visually striking empty patches of galaxy surveys are all *galaxy* voids, and it requires some nontrivial analysis to relate their distribution to that of the underlying dark matter. There is also some indication from N-body simulations (with Gaussian initial conditions) [34] that analyses of the distribution of voids based on the excursion set formalism, such as the one by Sheth & van de Weygaert [37], could fail to capture some non-universal features of the void abundance. Another concern is that the spherical assumption is of course an idealization – real structure is more complicated. Nevertheless, from the point of view of obtaining a better understanding of the underlying physics from an analytical perspective, an excursion set analysis based on the spherical ansatz remains the most robust starting point currently available. Once the calculation in this simplified setting is under control, one can explore improvements by relaxing the assumptions involved. We will adopt this point of view in this work, our goal being to obtain an expression for the abundance of voids in the presence of primordial NG. More precisely, we are after the differential comoving number density $dn_{\rm com}/dR_{\rm com}$ of voids of comoving size $R_{\rm com}$ which satisfies

$$R_{\rm com}\frac{dn_{\rm com}}{dR_{\rm com}} = \frac{1}{2}\,\frac{3}{4\pi R^3}\,f(S)\,\left|\frac{d\ln S}{d\ln R}\right|\,, \tag{1}$$

where R is the *Lagrangian* radius of the void related to the comoving radius by $R_{\rm com} = 1.7R$ (see below), $S \equiv \sigma^2(R)$ is the variance of the density contrast smoothed on scale R, and the multiplicity function $f(S)$ is what we will calculate analytically based on the excursion set formalism. Our results will be in a form that should be testable in N-body simulations which have complete control on the dark matter distribution. We will leave the second (and important) part of the exercise, namely that of relating our results to observationally relevant quantities, to future work.

We should mention that this approach has been adopted by other authors before us [47, 48] (see also Ref. [49]). The analysis of Kamionkowski *et al.* [47], while pioneering, was based on several simplifying

assumptions which we believe give an incomplete picture of the problem. In particular they used a Press-Schechter-like approach for the statistics which can be shown to miss certain scale dependent terms in the multiplicity f that arise from somewhat complex multi-scale correlations [50, 51]. Additionally their treatment of the non-Gaussianities involved an Edgeworth-like expansion (based on Ref. [28]) for the one point probability density of the smoothed density contrast, linearized in the non-Gaussianity amplitude. As pointed out in Ref. [51] for the case of dark matter halos, such an expansion tends to underestimate the effect of the NG on large scales, which is the regime we wish to probe using voids. And finally, Kamionkowski *et al.* also ignored a complication which arises when studying voids, which Sheth & van de Weygaert termed the "void-in-cloud" issue (which we discuss in detail below). Lam *et al.* [48] on the other hand used a more rigorous approach which accounted for multi-scale correlations and the void-in-cloud problem, but their treatment was also based on a linearized Edgeworth-like expansion and is hence subject to the same caveat mentioned above. Additionally their analysis made a technical assumption regarding the multi-scale correlations which is not strictly valid (see below). Our treatment below will be based on techniques introduced by Maggiore & Riotto [52, 53] and developed further by D'Amico *et al.* [51]. This will firstly allow us to derive a void multiplicity function which is valid on larger scales than the expressions of other authors. Secondly our approach will allow us to carefully account for the void-in-cloud issue (without making assumptions regarding the multi-scale correlations), and we will show that its final effects can actually be described using a simple yet accurate approximation.

The plan of this paper is as follows. In Section 2 we review the excursion set analysis of Sheth & van de Weygaert based on the spherical ansatz for Gaussian initial conditions, and describe the void-in-cloud issue mentioned above. To carefully account for this problem, we turn to a path integral description. We begin in Section 3 by rederiving the Gaussian result for the void multiplicity using path integrals, which allows us to introduce some formal machinery as well as discuss some of the subtleties of the calculation in a controlled setting. Following this, in Section 4 we generalize the calculation to the non-Gaussian case and derive our main result for the non-Gaussian void multiplicity, showing how to account for the void-in-cloud issue. Our final result is given in Eqn. (68) and is plotted in Figs. 2 and 3. We conclude in Section 5 with a brief discussion of the result and prospects for future work. Several technical details have been relegated to the Appendices.

2 Spherical void statistics: the SvdW result

The spherical ansatz for an expanding underdense region allows for a completely analytical treatment, exactly like what happens in the usual spherical *collapse* model [41]. Detailed numerical treatments [54, 33] also indicate that, although simplistic, this ansatz goes a long way in describing the evolution of individual underdense regions in realistic settings. For mathematical details of the model we refer the reader to Ref. [37], where Sheth and van de Weygaert (henceforth SvdW) also extensively discussed the physics of voids in the context of hierarchical structure formation. Since our primary concern in this paper is the statistical model built using the spherical ansatz, we will restrict ourselves to briefly describing some of the physical aspects relevant to our treatment.

An important feature of the spherical expansion scenario is the phenomenon of *shell crossing* [36, 37]. Physically shell crossing occurs because inner mass shells decelerate slower than the outer ones. With a sufficiently steep initial profile, this difference in decelerations is large enough that the inner shells catch up with the outer shells in a finite time, piling up mass in a very sharp ridge-like feature which then evolves self-similarly [43]. It can then be argued that the time of the *first* shell crossing is a sensible definition of the time of void formation. Such a definition brings us square into the realm of barriers and random walks, with the linearly extrapolated overdensity $\delta_{\rm c}$ from spherical collapse being replaced by a linearly extrapolated underdensity $(-\delta_{\rm v})$ corresponding to the time of first shell crossing. For an Einstein-de Sitter cosmology and assuming a tophat initial profile for the underdensity, this shell crossing threshold evaluates to $\delta_{\rm v} = 2.72$ [36]. Note that the excursion set ansatz relies on smoothing the (linearly extrapolated) initial conditions with a *Lagrangian* scale R associated with a conserved mass $M \propto R^3$. The observationally relevant scale however is the *comoving* scale $R_{\rm com}$ which corresponds to the physical size today of a sphere reaching first shell crossing. The spherical model predicts the relation [36]

$$R_{\rm com} = 1.7R\,. \tag{2}$$

It might seem that the statistical problem involving voids is now identical to the one with collapsed halos, and that a simple replacement $\delta_{\rm c} \to -\delta_{\rm v}$ in all halo statistics results would suffice to give the corresponding void statistics. The situation is not this simple however, as SvdW discuss extensively in Ref. [37]. Recall that the collapse problem had to deal with the subtlety of the "cloud-in-cloud" issue. This was the fact that, when counting the fraction of collapsed objects of mass M, one must only consider those trajectories (of the random walk of the smoothed density contrast) which *first* crossed the threshold $\delta_{\rm c}$ at a smoothing scale M, as the smoothing radius is decreased. Physically this corresponds to excluding overdense regions which are embedded in bigger overdense regions, since in such a case only the bigger region would survive as an independent collapsed object. Due to the nature of the problem, of course a similar requirement also holds in the case of voids, which SvdW call the "void-in-void" issue. This can be dealt with exactly as in the collapse case, and requires us to consider only first-crossing scales.

Additionally, SvdW also identified a second problem which is unique to the case of voids. Physically, this is the scenario where a region of size R_1 satisfies the threshold underdensity requirement and is a potential void candidate, but happens to be embedded in an *overdense* region of size $R_2 > R_1$ that satisfies the collapse criterion. This bigger region will then form a collapsed object, crushing the underdense void candidate out of existence. Explicit examples of such cases in N-body simulations were shown in Ref. [37]. The problem is therefore to exclude such situations from the statistics and was labeled the "void-in-cloud" issue by SvdW, who also showed how the problem could be tackled. The basic idea is that for the void formation problem there are now *two* barriers which are relevant – the negative void shell crossing threshold ($-\delta_{\rm v}$) which we will call the "void barrier", and also the positive halo formation threshold $\delta_{\rm c}$ which we will call the "halo barrier"[1]. Statistically, excluding the void-in-cloud cases amounts to counting only those trajectories in which the void barrier is first crossed *before* the halo barrier is ever crossed, as the smoothing radius is decreased from large values. SvdW showed that accounting for the void-in-cloud issue qualitatively changes the behaviour of the void multiplicity as compared to the halo multiplicity, by introducing a cutoff at small smoothing radii. This is again intuitively clear since small underdense regions would be more likely to find themselves embedded in larger scale overdensities, and should therefore not survive as voids. Our interest when extending the problem to the case of non-Gaussian initial conditions will primarily be in the *large* radius end, which is where non-Gaussianities are expected to play a significant role.

Before proceeding, a word on notation. In addition to the void barrier ($-\delta_{\rm v}$) and the halo barrier $\delta_{\rm c}$, we will frequently also require the *total* barrier height

$$\delta_{\rm T} \equiv \delta_{\rm c} + \delta_{\rm v} \,. \tag{3}$$

Additionally, all our later expressions will involve the "dimensionless" parameters $\delta/S^{1/2}$, where $S = S(R) = \langle\, \delta_R^2 \,\rangle$ is the variance of the density contrast smoothed on scale R. We will use the notation

$$\nu_{\rm c} \equiv \delta_{\rm c}/S^{1/2} \;\;;\;\; \nu_{\rm v} \equiv \delta_{\rm v}/S^{1/2} \;\;;\;\; \nu_{\rm T} \equiv \delta_{\rm T}/S^{1/2} \,. \tag{4}$$

In the following we will frequently refer to the variance S as "time", in keeping with standard terminology in excursion set theory. Statistically we are therefore interested in the probability distributions of first crossing (f.c.) times, or f.c. *rates* (conditional or otherwise), which will be denoted by $\mathcal{F}$ with appropriate subscripts. The multiplicity f which appears in the mass function (1) is related to $\mathcal{F}$ by $f = 2S\mathcal{F}$. SvdW showed that the conditional f.c. rate of the void barrier, accounting for the void-in-cloud issue and assuming Gaussian initial conditions, is given by

$$\mathcal{F}_{\rm SvdW}(S) = \sum_{j=1}^{\infty} \frac{j\pi}{\delta_{\rm T}^2} \sin\left(j\pi\delta_{\rm v}/\delta_{\rm T}\right) e^{-j^2\pi^2 S/2\delta_{\rm T}^2} \,, \tag{5}$$

whose Laplace transform is

$$\mathcal{L}_{\rm SvdW}(s) = \int_0^\infty dS e^{-sS} \mathcal{F}_{\rm SvdW}(S) = \frac{\sinh\left(\sqrt{2s}\delta_{\rm c}\right)}{\sinh\left(\sqrt{2s}\delta_{\rm T}\right)} \,. \tag{6}$$

[1]The actual value of $\delta_{\rm c}$ must be chosen carefully, since simply setting it equal to the usual spherical collapse value of 1.686 physically amounts to only excluding those underdense regions that would be completely crushed by collapsing overdensities. As SvdW discuss, this all-or-nothing approach fails to account for intermediate underdensities which are in the process of being crushed, thereby overestimating e.g. the typical void size in the Gaussian case (see also Ref. [46]). In this work we will ignore this complication and assume $\delta_{\rm c} = 1.686$.

SvdW derived this result based on probabilistic arguments and a clever use of Laplace transforms. This result is also known in the condensed matter literature on first crossing problems, and can be found derived e.g. in Ref. [55], based on a solution of the Fokker-Planck equation

$$\partial_S \Pi = \frac{1}{2}\partial_\delta^2 \Pi \,, \tag{7}$$

where $\Pi(\delta, S)$ is the probability density for a diffusing particle in the presence of absorbing barriers at $\delta = \delta_{\rm c}$ and $\delta = -\delta_{\rm v}$.

In the following we will be mainly interested in the large mass or small S limit of $\mathcal{F}_{\rm SvdW}$. As it stands, the expression (5) is not particularly useful in this limit, since for small S an increasing number of terms become important[2]. It is very useful therefore to recast this expression in a form that is better behaved as $S \to 0$. This is not hard to do, and we show in Appendix B.1 that the conditional f.c. rate $\mathcal{F}_{\rm SvdW}$ can be written as

$$\mathcal{F}_{\rm SvdW}(S) = \frac{1}{(2\pi)^{1/2} S} \sum_{j=-\infty}^{\infty} (\nu_{\rm v} - 2j\nu_{\rm T}) e^{-(\nu_{\rm v} - 2j\nu_{\rm T})^2/2} \,. \tag{8}$$

(One can check that the series above is also identical to the one in Eqn. 46 of Lam *et al.* [48], with our $j = 1$ term corresponding to their $n = 1$, our $j = -1$ to their $n = 2$, and so on.) As $S \to 0$, the $j = 0$ term rapidly becomes the most important and the resulting multiplicity reduces to

$$S \to 0 \quad : \quad f_{\rm SvdW}(\nu_{\rm v}, \nu_{\rm T}) = 2S\mathcal{F}_{\rm SvdW}(S) \to \sqrt{\frac{2}{\pi}} \nu_{\rm v} \; e^{-\nu_{\rm v}^2/2} = f_{\rm PS}(\nu_{\rm v}) \,, \tag{9}$$

which is just the 1-barrier Press-Schechter result for the void barrier. This is not surprising, since for small times it becomes increasingly unlikely for a trajectory to cross the halo barrier *and* return to cross the void barrier, reducing the result to a single barrier one. Figure 7 of Ref. [37] illustrates this effect, with the void-in-cloud effects becoming significant only at $\nu_{\rm v} \lesssim 1.5$ or so.

In our non-Gaussian extension we would like to address the question of whether this intuition continues to remain true, or whether the void-in-cloud issue is now relevant at very large radii. If we simply assume that the Gaussian reasoning still holds, then the large radius end should be describable as a single barrier problem. This was essentially the reasoning of Kamionkowski *et al.* [47], who applied the Press-Schechter approach to a single void barrier. If we follow this reasoning, then *at the least* the non-Gaussian void multiplicity $f_{\rm voids}$ should incorporate the single barrier effects discussed in Refs. [50, 51], and we should expect that $f_{\rm voids}$ is given by replacing $\delta_{\rm c} \to -\delta_{\rm v}$ in the expression derived by D'Amico *et al.* [51] (their Eqn. 68 with $\kappa = 0$ and $a = 1$ to remain within the sharp-k filter and fixed barrier approximations we are using here).

In this paper we will address this issue rigorously using path integral techniques. We will see that, while the void-in-cloud issue introduces some technical complications, the end result *is* in fact that at sufficiently large radii one can treat the problem using a single barrier. For smaller radii, we will see that the effects of the void-in-cloud issue are not negligible but can be described using a simple yet accurate approximation. Our results are broadly in agreement with those of Lam *et al.* [48], although our final expression for the void multiplicity is different from theirs (see below for a more detailed comparison with their work). We begin in the next section by introducing some path integral machinery.

3 Deriving the SvdW result from path integrals

Our goal in this section is to reproduce using path integrals the SvdW result which assumes Gaussian initial conditions. This will allow us to rigorously extend the result to the non-Gaussian case. Unfortunately the calculations we will end up doing are technically rather involved. It is instructive therefore to first go through an intuitive derivation of $\mathcal{F}_{\rm SvdW}$.

Consider the probability distribution $\Pi_{\rm allowed}(\delta, S)$ for the location of the diffusing particle, which solves the Fokker-Planck equation (7) in the presence of two absorbing barriers and is nonzero in the "allowed" region

[2]The expression (5) is more useful in the *large* time ($S \to \infty$) limit, which is typically encountered in condensed matter systems [55].

$-\delta_{\rm v} < \delta < \delta_{\rm c}$ between the barriers. One can interpret the rate $\mathcal{F}$ at time S as the amount of probability leaking across the chosen barrier per unit time. This is similar to constraining the probability to stay within the barriers until time S and then letting it evolve freely, as if the barriers were removed at time S. The rate at time S across let's say the void barrier can then be computed as a derivative of the probability that has leaked across this barrier at S, i.e.

$$\mathcal{F}_{\rm intuitive}(S) = \partial_S \int_{-\infty}^{-\delta_{\rm v}} d\delta\, \Pi(\delta, S)\,. \tag{10}$$

Here Π represents the *unconstrained* probability density, which also satisfies the same Fokker-Planck equation (7), with the "initial" condition that at time S we have $\Pi = \Pi_{\rm allowed}$. Using the Fokker-Planck equation one can simplify the expression above to find

$$\mathcal{F}_{\rm intuitive}(S) = \frac{1}{2}\partial_\delta \Pi_{\rm allowed}(\delta, S)\Big|_{\delta=-\delta_{\rm v}}\,. \tag{11}$$

The solution to the Fokker-Planck equation which vanishes at the two barriers $\delta = \delta_{\rm c}$ and $\delta = -\delta_{\rm v}$ and starts from a Dirac delta at the origin at the initial time is simply an infinite sum of Gaussians with shifted mean values,

$$\Pi_{\rm allowed}(\delta, S) = \sum_{j=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi S}} \left\{ \exp\left[-\frac{(\delta + 2j\delta_{\rm T})^2}{2S}\right] - \exp\left[-\frac{(\delta - 2j\delta_{\rm T} + 2\delta_{\rm v})^2}{2S}\right]\right\}. \tag{12}$$

This is easy to verify by inspection, since each Gaussian is separately a solution of the Fokker-Plank equation (7), and by construction the infinite sum satisfies the boundary conditions $\Pi(-\delta_{\rm v}, S) = \Pi(\delta_{\rm c}, S) = 0$ (as can be immediately verified for $(-\delta_{\rm v})$, while for $\delta_{\rm c}$ it is enough to shift j to $j+1$ in the second Gaussian). Straightforward algebra then shows that the conditional f.c. rate computed according to Eqn. (11) is precisely the series given in Eqn. (8),

$$\mathcal{F}_{\rm intuitive}(S) = \mathcal{F}_{\rm SvdW}(S)\,. \tag{13}$$

These ideas can be made more rigorous using the language of path integrals. The path integral approach to excursion sets for computing the halo multiplicity was developed by Maggiore & Riotto (MR) in a series of papers [52, 53, 50], and was improved upon by D'Amico *et al.* [51] for the non-Gaussian case. In the following we will mainly refer to the techniques developed in Ref. [52], restricting ourselves to the simplest case of a fixed barrier and a sharp-k filter (so that one is dealing with a Markovian stochastic process). The basic quantity one deals with is the probability distribution function $W(\delta_0; \{\delta_k\}_n; S)$ for a discrete random walk with n steps $\{\hat{\delta}_k\}|_{k=1}^n$, where $\hat{\delta}_k$ denotes the matter density contrast smoothed on a scale R_k corresponding to a variance S_k, with steps of equal spacing ΔS in the variance starting at S_0 with corresponding density contrast δ_0 (both of which we will assume to be zero) and with the last step denoted by $S_n \equiv S$. We have

$$W(\delta_0; \{\delta_k\}_n; S) \equiv \langle\, \delta_{\rm D}(\hat{\delta}_1 - \delta_1) \ldots \delta_{\rm D}(\hat{\delta}_n - \delta_n)\,\rangle\,, \tag{14}$$

which for the Gaussian, sharp-k filter case reduces to

$$W^{\rm gm} = \prod_{k=0}^{n-1} \Psi_{\Delta S}(\delta_{k+1} - \delta_k)\ ; \quad \Psi_{\Delta S}(x) = (2\pi\Delta S)^{-1/2} e^{-x^2/(2\Delta S)}\,, \tag{15}$$

with the superscript standing for "Gaussian, Markovian". Throughout our calculations we will deal with the path integral of this probability density over the "allowed" region – in the single barrier case that MR considered this would be the region $\delta_k < \delta_{\rm c}$, $1 \le k \le n-1$, while in our two barrier problem it will be the region $-\delta_{\rm v} < \delta_k < \delta_{\rm c}$. We will therefore be interested in objects of the type

$$\Pi_{\Delta S}(\delta_0, \delta_n; S) = \int_{\rm allowed} d\delta_1 \ldots d\delta_{n-1}\, W(\delta_0; \{\delta_k\}_n; S)\,. \tag{16}$$

This object is the probability density for the diffusing "particle" to remain inside its allowed region until time S. Our goal will be to calculate the rate at which this probability leaks out of the allowed region (i.e. the rate at which the "particles" escape) across one of the boundaries. Making these ideas rigorous requires us to introduce some technical aspects of the path integrals, and it will be easiest to do this in the more familiar single barrier case. Let us therefore begin with a brief recap of the (Gaussian) MR calculation for halos in the case of a fixed barrier and a sharp-k filter.

3.1 Recap of the single barrier problem

As mentioned above, it is useful to define the constrained probability density $\Pi^{\rm gm}_{1-{\rm bar},\Delta S}(\delta_0,\delta_n;S)$ for the density contrast at step n, given that at all previous steps the walk remained below the barrier, so that

$$\Pi^{\rm gm}_{1-{\rm bar},\Delta S}(\delta_0,\delta_n;S)=\int_{-\infty}^{\delta_{\rm c}} d\delta_1\ldots d\delta_{n-1}W^{\rm gm}(\delta_0;\{\delta_k\}_n;S)\,. \tag{17}$$

MR showed [52] that the continuum limit of this quantity recovers the result of Bond *et al.* [40],

$$\Pi^{\rm gm}_{1-bar,\Delta S\to 0}(\delta_0,\delta;S)=\Pi_{\rm Bond}(\delta_0,\delta;S)\equiv\frac{1}{\sqrt{2\pi S}}\left(e^{-(\delta-\delta_0)^2/2S}-e^{-(2\delta_{\rm c}-\delta-\delta_0)^2/2S}\right)\,, \tag{18}$$

being the solution of the Fokker-Planck equation (7) with initial condition $\Pi(\delta_0,\delta;0)=\delta_{\rm D}(\delta-\delta_0)$ and boundary conditions $\Pi(\delta_0,\delta_{\rm c};S)=0=\Pi(\delta_0,\delta\to-\infty;S)$. We will go through the derivation of a similar result for the two barrier case below.

We are looking for the distribution of the f.c. time $\hat{S}_{\rm c}$ at which the random walk first crosses the barrier $\delta_{\rm c}$. This can be constructed as follows. Start with the cumulative probability $P(\hat{S}_{\rm c}>S)$ that $\hat{S}_{\rm c}>S$, which is the same as the probability that the barrier has *not* been crossed until time S, i.e.

$$P(\hat{S}_{\rm c}>S)=\lim_{\Delta S\to 0}\int_{-\infty}^{\delta_{\rm c}} d\delta_n\Pi^{\rm gm}_{1-{\rm bar},\Delta S}(\delta_0,\delta_n;S)\,. \tag{19}$$

In the continuum limit this quantity is straightforward to compute using Eqn. (18). The f.c. rate is then just the negative derivative of $P(\hat{S}_{\rm c}>S)$ w.r.t S, leading to the multiplicity

$$f_{\rm PS}(\nu_{\rm c})=2S\mathcal{F}_{\rm PS}(S)=-2S\partial_S P(\hat{S}_{\rm c}>S)=\sqrt{\frac{2}{\pi}}\nu_{\rm c}\;e^{-\nu_{\rm c}^2/2}\,, \tag{20}$$

which is the celebrated Press-Schechter result (accounting for the infamous factor of 2), where $\nu_{\rm c}$ was defined in Eqn. (4).

On passing to the two barrier problem, in order to calculate the required constrained f.c. rate, the cumulative probability we need is $P(\hat{S}_{\rm v}>S,\hat{S}_{\rm c}>\hat{S}_{\rm v})$, which is the probability that the void barrier is first crossed *before* the halo barrier is ever crossed. In this object, the conditioning of the stochastic variable $\hat{S}_{\rm v}$ is thus on *another* stochastic variable $\hat{S}_{\rm c}$ which is the first crossing time for the halo barrier, and this complicates the issue. In the Gaussian case of course one can solve the problem the SvdW way, without resorting to path integrals. We are ultimately interested in the non-Gaussian case though, and we therefore explore a path integral derivation of the Gaussian SvdW result, which we can generalize later to the non-Gaussian case.

3.2 $\mathcal{F}_{\rm SvdW}$ from a path integral analysis

The basic trick we employ is to exploit the discretized nature of the path integral, and to pass to the continuum limit carefully. To keep the discussion as clear as possible, we relegate several technical derivations to the Appendices. As in the single barrier case, it is useful to construct the constrained probability density $\Pi^{\rm gm}_{\Delta S}(\delta_0,\delta_n;S_n)$ for the density contrast at time step n, given that the walk has not crossed *either* barrier at any previous time step,

$$\Pi^{\rm gm}_{\Delta S}(\delta_0,\delta_n;S_n)=\int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_1\ldots d\delta_{n-1}W^{\rm gm}(\delta_0;\{\delta_k\}_n;S_n)\,. \tag{21}$$

We will need to compute this quantity and its integrals under various limits and assumptions, which we will come to presently. To begin with, note that with a discretized variance parameter S one can explicitly construct a probability (rather than a density) for the constrained first crossing of the void barrier to occur at a *specific* time step n. Namely, the integral $\int_{-\infty}^{-\delta_{\rm v}} d\delta_n\Pi^{\rm gm}_{\Delta S}(\delta_0,\delta_n;S_n)$ is the probability that the walk did not cross either barrier for the first $n-1$ steps, and crossed the void barrier at step n. This is the same as the probability that the first crossing of the void barrier is at step n, and that the halo barrier has not yet been crossed. Using

this, the cumulative probability $P(\hat{S}_{\rm v} > S, \hat{S}_{\rm c} > \hat{S}_{\rm v})$ is obtained by summing over all possible choices of the final step n that have $S_n > S$, and then passing to the continuum limit,

$$P(\hat{S}_{\rm v} > S, \hat{S}_{\rm c} > \hat{S}_{\rm v}) = \sum_{S_n > S} \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \Pi^{\rm gm}_{\Delta S}(\delta_0, \delta_n; S_n) \to \int_S^\infty d\tilde{S} \lim_{\Delta S \to 0} \frac{1}{\Delta S} \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \Pi^{\rm gm}_{\Delta S}(\delta_0, \delta_n; \tilde{S}) \,. \tag{22}$$

The required f.c. rate is simply the negative time derivative of this object, which we can read off as

$$\mathcal{F}_{\rm WS}(S) = \lim_{\Delta S \to 0} \frac{1}{\Delta S} \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \Pi^{\rm gm}_{\Delta S}(\delta_0, \delta_n; S) \,, \tag{23}$$

which involves the integral of $\Pi^{\rm gm}_{\Delta S}$ on the "wrong side" of the void barrier, $\delta_n < -\delta_{\rm v}$ (hence the subscript WS). One can also see why this is the correct object to compute, since it corresponds to the amount of probability leaking out of the void barrier in a time interval ΔS, divided by ΔS. While the integral itself will vanish in the continuum limit, we will see that it does so like $\sim \Delta S$, leaving a finite limit in Eqn. (23) such that $\mathcal{F}_{\rm WS}(S) = \mathcal{F}_{\rm SvdW}(S)$.

The above arguments may seem like a convoluted way of arriving at the result. When we move to non-Gaussian initial conditions for the double barrier however, we are not left with much choice in the matter. Nevertheless, it would be reassuring to know that the "wrong side counting" described above actually works in some other situation which is under better control. The single barrier calculation provides us with such a check. In this case, we can apply exactly the same arguments as above to obtain the f.c. rate, *and* we have an alternative derivation of the rate due to MR. As a check therefore, we should find that these two methods lead to the same answer. In Appendix B.2, we show that this is indeed the case for *completely general* initial conditions (i.e. *without* assuming Gaussianity). In Appendix B.3 we also compare our approach with that of Lam *et al.* [48]. Their analysis effectively makes the assumption that the probability distribution $W(\delta_0; \{\delta_k\}_n; S)$ defined in Eqn. (14) is factorisable, which is certainly true in the Gaussian case for the sharp-k filter (see Eqn. (15)) but is not valid e.g. in the presence of non-Gaussianities. We show in Appendix B.3 that our expression for the rate (both the Gaussian one of Eqn. (23) as well as its non-Gaussian generalisation which we discuss in Section 4 below) is identical to what they aim to calculate in Eqn. 41 of Ref. [48], but *without* making this assumption of factorisability. With these reassurances, we proceed to the main two barrier calculation of this section. Since the calculation is rather technical in nature, the reader who is willing to take the result $\mathcal{F}_{\rm WS} = \mathcal{F}_{\rm SvdW}$ on faith may skip directly to the non-Gaussian generalization in Section 4.

To calculate the integral in Eqn. (23), we start by exploiting the factorisability of $W^{\rm gm}$ to write

$$\begin{aligned}
\Pi^{\rm gm}_{\Delta S}(\delta_0, \delta; S + \Delta S) &= \int_{\delta - \delta_{\rm c}}^{\delta + \delta_{\rm v}} dx \Psi_{\Delta S}(x) \Pi^{\rm gm}_{\Delta S}(\delta_0, \delta - x; S) \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \partial_\delta^n \Pi^{\rm gm}_{\Delta S}(\delta_0, \delta; S) \int_{\delta - \delta_{\rm c}}^{\delta + \delta_{\rm v}} dx \Psi_{\Delta S}(x) x^n \\
&\equiv \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \partial_\delta^n \Pi^{\rm gm}_{\Delta S}(\delta_0, \delta; S) \mathcal{I}_n^{\Delta S}(\delta) \,,
\end{aligned} \tag{24}$$

where the first equality follows from the definitions of $\Pi^{\rm gm}_{\Delta S}$, $W^{\rm gm}$ and some relabeling of dummy variables, the second follows from a Taylor expansion and exchanging the orders of integration and summation, and the last line defines the functions $\mathcal{I}_n^{\Delta S}(\delta)$, which reduce to

$$\mathcal{I}_n^{\Delta S}(\delta) = \frac{(2\Delta S)^{n/2}}{\pi^{1/2}} \int_{(\delta - \delta_{\rm c})/(2\Delta S)^{1/2}}^{(\delta + \delta_{\rm v})/(2\Delta S)^{1/2}} dy \, e^{-y^2} y^n \,. \tag{25}$$

Eqn. (24) is completely general, and holds for arbitrary δ. The first thing we can check is that this expression implies that in the continuum limit, $\Pi^{\rm gm}_{\Delta S=0}(\delta_0, \delta; S)$ vanishes on *both* barriers $\delta = -\delta_{\rm v}$ and $\delta = \delta_{\rm c}$,

$$\Pi^{\rm gm}_{\Delta S=0}(\delta_0, -\delta_{\rm v}; S) = 0 = \Pi^{\rm gm}_{\Delta S=0}(\delta_0, \delta_{\rm c}; S) \,. \tag{26}$$

This follows by Taylor expanding the l.h.s. of (24) for small ΔS and comparing the lowest order terms on both sides, using $\mathcal{I}_0^{\Delta S \to 0}(\delta_{\rm c}) = 1/2 = \mathcal{I}_0^{\Delta S \to 0}(-\delta_{\rm v})$. Here we set δ to $\delta_{\rm c}$ or $(-\delta_{\rm v})$ *before* taking the continuum limit.

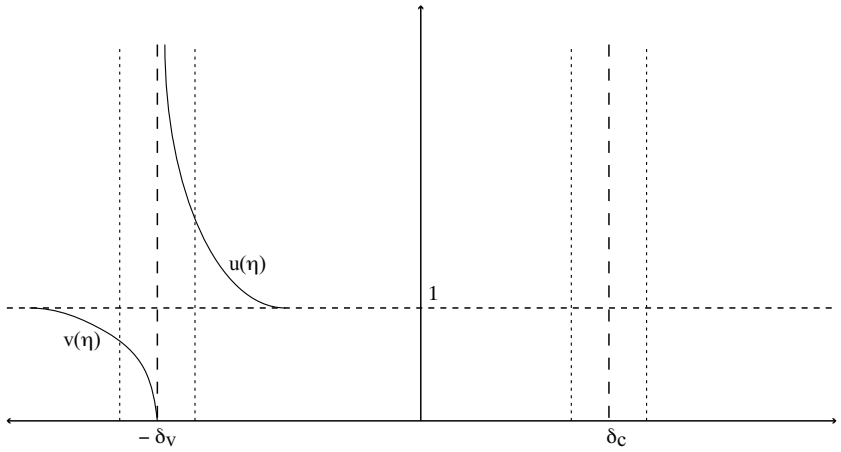


Figure 1: Schematic view of the two barriers (vertical dashed lines) with corresponding boundary layers (vertical dotted lines). The boundary layer functions $v(\eta)$ and $u(\eta)$ are also shown schematically.

This is a subtle point since these operations do not commute, which is clear from the structure of $\mathcal{I}_n^{\Delta S}(\delta)$. This brings us to the issue of "boundary layer" effects.

In evaluating the integral in Eqn. (23), we wish to take the continuum limit *after* integrating, in contrast to what we did above for $\mathcal{I}_0^{\Delta S}$. The integral therefore depends not only on the behaviour of $\Pi_{\Delta S}^{\rm gm}$ far from the void barrier, but also on its detailed behaviour in the "boundary layer" where $|\delta + \delta_{\rm v}| \sim \sqrt{\Delta S}$. In our wrong side counting approach, there are actually two boundary layers we must worry about. If we define the quantity η by

$$\eta \equiv \frac{\delta + \delta_{\rm v}}{\sqrt{2\Delta S}}\,, \tag{27}$$

then the two boundary layers correspond to regions where $|\eta| \sim \mathcal{O}(1)$, with the one discussed above corresponding to $\eta < 0$. We will also deal with the second layer at $\eta > 0$, which is in fact similar to the one discussed by MR in their single barrier calculation. In principle there are two more boundary layers, on either side of the halo barrier as well, but we will only need to deal with the ones near the void barrier. Fig. 1 illustrates the situation. We will find that it is in fact not possible in this approach to account for the boundary layer effects *ab initio*. However, these effects will reduce to a single number, which we can then fix by matching to the correct normalization for the f.c. rate. To account for boundary layer effects in the integral in Eqn. (23), we use the approach discussed by MR in Ref. [52]. This involves two steps: first, we calculate the leading behaviour of $\Pi_{\Delta S}^{\rm gm}(\delta_0, \delta; S)$ for small ΔS when we hold $\delta < -\delta_{\rm v}$ to be fixed – i.e. in the limit of large negative η. Denote this function by $C_{\Delta S}^{(<)}(\delta_0, \delta; S)$. Next, in order to calculate $\Pi_{\Delta S}^{\rm gm}$ for fixed but small ΔS and *arbitrary* $\delta \leq -\delta_{\rm v}$ (so that η is negative and arbitrary), we introduce a boundary layer function $v(\eta)$ which we choose to normalize so that $v(\eta \to -\infty) \to 1$, and then write

$$\Pi_{\Delta S}^{\rm gm}(\delta_0, \delta; S) = v(\eta) C_{\Delta S}^{(<)}(\delta_0, \delta; S) + \dots\,, \tag{28}$$

where the ellipsis indicates terms of higher order in ΔS than the leading order in $C_{\Delta S}^{(<)}$. The function $v(\eta)$ will in general behave nontrivially in the boundary layer where $|\eta| \lesssim \mathcal{O}(1)$, capturing effects missed by the function $C_{\Delta S}^{(<)}$ on its own.

Let us now calculate $C_{\Delta S}^{(<)}$ which can be done by starting with Eqn. (24). In Appendix C.1, we show that in this limit of large negative η the functions $\mathcal{I}_n^{\Delta S}(\delta)$ reduce to

$$\mathcal{I}_n^{\Delta S}(\delta < -\delta_{\rm v}) = \left(\frac{\Delta S}{2\pi}\right)^{1/2} \left[-e^{-(\delta+\delta_{\rm v})^2/2\Delta S}\,(\delta + \delta_{\rm v})^{n-1} \left(1 + \mathcal{O}\left(\frac{\Delta S}{(\delta+\delta_{\rm v})^2}\right)\right) + \dots\right]\,, \tag{29}$$

where the ellipsis indicates terms which are exponentially suppressed. Replacing this in Eqn. (24) we find that to leading order in ΔS, the summation over n can be carried out exactly using the identity

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \left(\delta + \delta_{\rm v}\right)^n \partial_\delta^n \Pi^{\rm gm}_{\Delta S}(\delta_0, \delta; S) = \Pi^{\rm gm}_{\Delta S}(\delta_0, -\delta_{\rm v}; S)\,, \tag{30}$$

which is not zero for finite ΔS. To leading order in ΔS, the l.h.s. of Eqn. (24) is simply the function $C^{(<)}_{\Delta S}(\delta_0, \delta; S)$ we are looking for, and we can set

$$C^{(<)}_{\Delta S}(\delta_0, \delta; S) \equiv \left(\frac{\Delta S}{2\pi}\right)^{1/2} \frac{e^{-(\delta+\delta_{\rm v})^2/2\Delta S}}{|\delta + \delta_{\rm v}|} \Pi^{\rm gm}_{\Delta S}(\delta_0, -\delta_{\rm v}; S)\,, \tag{31}$$

where it is understood that we are only interested in the leading order in ΔS. Since the object $\Pi^{\rm gm}_{\Delta S}(\delta_0, -\delta_{\rm v}; S)$ is independent of δ we will calculate it later. The function $\Pi^{\rm gm}_{\Delta S}(\delta_0, \delta; S)$ for arbitrary $\delta \leq -\delta_{\rm v}$ can be written as

$$\Pi^{\rm gm}_{\Delta S}(\delta_0, \delta \leq -\delta_{\rm v}; S) = \frac{1}{2\sqrt{\pi}} \frac{v(\eta)}{(-\eta)} e^{-\eta^2} \Pi^{\rm gm}_{\Delta S}(\delta_0, -\delta_{\rm v}; S)\,, \tag{32}$$

where we have expressed $C^{(<)}_{\Delta S}$ in terms of η. In the other limit of fixed ΔS but $\delta \to -\delta_{\rm v}$ from below, due to continuity of $\Pi^{\rm gm}_{\Delta S}(\delta_0, \delta; S)$ we have

$$v(\eta) \to 2\sqrt{\pi}(-\eta) \;\; \text{as } \eta \to 0^-\,. \tag{33}$$

Let us focus now on the integral in Eqn. (23). It is not hard to see that this can be re-expressed as an integral over η from $-\infty$ to 0, and reduces to

$$\int_{-\infty}^{-\delta_{\rm v}} d\delta\, \Pi^{\rm gm}_{\Delta S}(\delta_0, \delta; S) = \left(\frac{\Delta S}{2\pi}\right)^{1/2} \left[\int_{-\infty}^{0} \frac{d\eta}{(-\eta)} v(\eta) e^{-\eta^2}\right] \Pi^{\rm gm}_{\Delta S}(\delta_0, -\delta_{\rm v}; S)\,. \tag{34}$$

The integral in square brackets depends strongly on details of the function $v(\eta)$, and is in fact dominated by contributions from the boundary layer where $|\eta| \lesssim 1$ due to the exponential cutoff. We do not know of any way to calculate this object from first principles. However, this integral is a *finite constant* since the integrand is well behaved as $\eta \to 0^-$ (see Eqn. (33)) and is exponentially cutoff as $\eta \to -\infty$. We will therefore adopt a pragmatic point of view and allow this constant to be fixed later by a suitable normalization.

The calculation of $\mathcal{F}_{\rm WS}$ will be complete once we evaluate the object $\Pi^{\rm gm}_{\Delta S}(\delta_0, -\delta_{\rm v}; S)$ (which we need only to leading order in ΔS). We start by returning to Eqn. (24), where we are now interested in the limit $\delta \to -\delta_{\rm v}$ from *above* at fixed ΔS – i.e as $\eta \to 0^+$. As we shall see next, this will allow us to exploit the properties of the continuum probability density $\Pi^{\rm gm}_{\Delta S=0}$ which is non-zero in $-\delta_{\rm v} < \delta < \delta_{\rm c}$ and is moreover calculable. As one might expect, once again we need to account for a boundary layer, this time the one where η is *positive* and of order unity. This calculation closely mimics that of a similar object in the single barrier case [52]. We essentially repeat the previous analysis of $\Pi^{\rm gm}_{\Delta S}(\delta_0, \delta; S)$, now on the other side of the void barrier. Once again we define a function $C^{(>)}_{\Delta S}(\delta_0, \delta; S)$ which reproduces the leading behaviour in ΔS of $\Pi^{\rm gm}_{\Delta S}(\delta_0, \delta; S)$ when we hold $\delta \in (-\delta_{\rm v}, \delta_{\rm c})$ fixed and let $\Delta S \to 0$. This time we can easily see that $C^{(>)}_{\Delta S}(\delta_0, \delta; S)$ must simply be the continuum limit probability density

$$C^{(>)}_{\Delta S}(\delta_0, \delta; S) \equiv \Pi^{\rm gm}_{\Delta S=0}(\delta_0, \delta; S)\,. \tag{35}$$

Introducing a second boundary layer function $u(\eta)$, we can express $\Pi^{\rm gm}_{\Delta S}(\delta_0, \delta; S)$ for arbitrary $\delta \in [-\delta_{\rm v}, \delta_{\rm c})$ as

$$\Pi^{\rm gm}_{\Delta S}(\delta_0, \delta; S) = u(\eta) C^{(>)}_{\Delta S}(\delta_0, \delta; S)\,. \tag{36}$$

With $C^{(>)}_{\Delta S}(\delta_0, \delta; S)$ defined to be smooth in the boundary layer where $\eta = \mathcal{O}(1)$, we can obtain its leading behaviour in ΔS by a simple Taylor expansion, which leads to

$$\begin{aligned} C^{(>)}_{\Delta S}(\delta_0, \delta; S) &= \Pi^{\rm gm}_{\Delta S=0}(\delta_0, -\delta_{\rm v} + \eta\sqrt{2\Delta S}; S) \\ &= \eta\sqrt{2\Delta S}\, \partial_\delta \Pi^{\rm gm}_{\Delta S=0}(\delta_0, \delta = -\delta_{\rm v}; S) + \dots\,, \end{aligned} \tag{37}$$

where we used the fact that the continuum limit function $\Pi^{\rm gm}_{\Delta S=0}$ vanishes when evaluated at the barrier. Finally, we can write the leading order form of $\Pi^{\rm gm}_{\Delta S}(\delta_0, -\delta_{\rm v}; S)$ by taking the limit $\eta \to 0^+$ in Eqn. (36) to get

$$\Pi^{\rm gm}_{\Delta S}(\delta_0, -\delta_{\rm v}; S) = \gamma \sqrt{2\Delta S}\, \partial_\delta \Pi^{\rm gm}_{\Delta S=0}(\delta_0, \delta = -\delta_{\rm v}; S) \;\; ; \;\; \gamma \equiv \lim_{\eta\to 0^+} \eta\, u(\eta)\,. \tag{38}$$

Here γ is another constant which depends on details of the boundary layer, and is in fact very similar to the constant which MR compute in Ref. [52] (see their Eqn 78). We will however simply combine it with the other unknown constant which appears in Eqn. (34) and fix it by normalization. Plugging the above relation into Eqn. (34), we see that the integral in Eqn. (23) is proportional to ΔS at the leading order. Putting everything together, the two barrier conditional f.c. rate is then given by

$$\mathcal{F}_{\rm WS} = \mathcal{A}\, \frac{1}{2} \partial_\delta \Pi^{\rm gm}_{\Delta S=0}(\delta_0, \delta = -\delta_{\rm v}; S) \;\; ; \;\; \mathcal{A} \equiv \frac{2\gamma}{\sqrt{\pi}} \int_{-\infty}^{0} \frac{d\eta}{(-\eta)} v(\eta) e^{-\eta^2}\,. \tag{39}$$

We now need the derivative of the continuum limit solution between the barriers. If we hold $\delta \in (-\delta_{\rm v}, \delta_{\rm c})$ fixed and take the limit $\Delta S \to 0$ then it is easy to see, by expanding both sides of Eqn. (24) in powers of ΔS and equating the lowest order terms, that the continuum limit function satisfies the Fokker-Planck equation (7) with boundary conditions (26) and initial condition a Dirac delta centered at δ_0 (which we set to zero for convenience). In other words, our rigorous derivation has reproduced Eqn. (11), upto a constant factor. Since we saw earlier that the expression (11) is the same as $\mathcal{F}_{\rm SvdW}$ in Eqns. (8) and (5), we have

$$\mathcal{F}_{\rm WS}(S) = \mathcal{A}\, \mathcal{F}_{\rm SvdW}(S)\,, \tag{40}$$

and in order to match our solution with SvdW we must set $\mathcal{A} = 1$.

4 Non-Gaussian voids

With the path integral machinery in place, it is formally staightforward (although still somewhat involved in practice) to extend the calculation to the case of non-Gaussian initial conditions. We simply replace $W^{\rm gm}$ and $\Pi^{\rm gm}_{\Delta S}$ from the Gaussian calculation with their appropriately generalized versions. Namely, we have [50]

$$\begin{aligned} W(\delta_0; \{\delta_k\}_n; S) &\equiv \langle\, \delta_{\rm D}(\hat{\delta}_1 - \delta_1) \ldots \delta_{\rm D}(\hat{\delta}_n - \delta_n)\, \rangle \\ &= \exp\Bigg[- \frac{1}{3!} \sum_{j,k,l=1}^{n} \langle\, \hat{\delta}_j \hat{\delta}_k \hat{\delta}_l\, \rangle_c \partial_j \partial_k \partial_l \\ &\qquad + \frac{1}{4!} \sum_{j,k,l,m=1}^{n} \langle\, \hat{\delta}_j \hat{\delta}_k \hat{\delta}_l \hat{\delta}_m\, \rangle_c \partial_j \partial_k \partial_l \partial_m + \ldots \Bigg] W^{\rm gm}(\delta_0; \{\delta_k\}_n; S)\,, \end{aligned} \tag{41}$$

where $\langle\, \hat{\delta}_j \hat{\delta}_k \hat{\delta}_l\, \rangle_c$, $\langle\, \hat{\delta}_j \hat{\delta}_k \hat{\delta}_l \hat{\delta}_m\, \rangle_c$, etc. are the connected moments correlating different length scales (which all vanish in the Gaussian case, giving back $W^{\rm gm}$). Using this, we also have

$$\Pi_{\Delta S}(\delta_0, \delta_n; S_n) = \int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_1 \ldots d\delta_{n-1} W(\delta_0; \{\delta_k\}_n; S_n)\,. \tag{42}$$

One can now apply exactly the same wrong side counting arguments as in the Gaussian case (see the application to the single barrier case in Appendix B.2), and find that the required non-Gaussian f.c. rate is given by

$$\mathcal{F}_{\rm NG}(S) = \lim_{\Delta S \to 0} \frac{1}{\Delta S} \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \Pi_{\Delta S}(\delta_0, \delta_n; S)\,. \tag{43}$$

We will focus on weak primordial non-Gaussianities (NG) which are expected to be generated in inflationary models. One expects that in this case a scale dependent perturbative treatment of the NG along the lines discussed by D'Amico *et al.* [51] for the halo multiplicity should work well for voids as well, with possible

complications due to the void-in-cloud issue. We proceed as in Ref. [51], beginning by defining the "equal time" correlation functions[3] (recall $S = \sigma_R^2$),

$$\varepsilon_{n-2} \equiv \frac{\langle\, \hat{\delta}_R^n \,\rangle_c}{\sigma_R^n} \quad ; \quad n \geq 3\,, \tag{44}$$

which are expected to follow the perturbative hierarchy $\varepsilon_n \sim \epsilon^n$ for $n \geq 1$, where $\epsilon \ll 1$, in generic inflationary models (see Appendix A for a brief introduction to primordial NG). Next we Taylor expand the "unequal time" correlators appearing in Eqn. (41) around the final time S [52] as e.g.

$$\langle\, \hat{\delta}_j \hat{\delta}_k \hat{\delta}_l \,\rangle_c = \sum_{p,q,r=0}^{\infty} \frac{(-1)^{p+q+r}}{p!q!r!} \mathcal{G}_3^{(p,q,r)}(S)(S-S_j)^p (S-S_k)^q (S-S_l)^r\,, \tag{45}$$

where we introduced the (scale dependent) coefficients

$$\mathcal{G}_3^{(p,q,r)}(S) \equiv \left[\frac{d^p}{dS_j^p} \frac{d^q}{dS_k^q} \frac{d^r}{dS_l^r} \langle\, \hat{\delta}(S_j) \hat{\delta}(S_k) \hat{\delta}(S_l) \,\rangle_c \right]_{S_j = S_k = S_l = S}\,, \tag{46}$$

and similarly for higher point correlations. Such an expansion works well when considering large scales for which S is small. In this work we will restrict ourselves to the effects of at most $\mathcal{G}_3^{(1,0,0)}$, ignoring the higher order correlations which will be suppressed by powers of S and can be tracked as part of the theoretical error we make in the calculation [51]. We also introduce a convenient parametrization of $\mathcal{G}_3^{(1,0,0)}$ as in Ref. [51] by defining the function $c_1(S)$ via

$$\mathcal{G}_3^{(1,0,0)}(S) = \frac{1}{2} \varepsilon_1(S) c_1(S) S^{1/2}\,, \tag{47}$$

which satisfies the identity

$$c_1(S) = 1 + \frac{2}{3} \frac{d \ln \varepsilon_1}{d \ln S}\,. \tag{48}$$

We will further follow the analysis of Ref. [51] and simplify the expression for the f.c. rate (43) by retaining equal time correlators in the exponential while linearizing the unequal time contributions, which is a strategy that works well at least for the single barrier problem. We then find, to the order we are interested in,

$$\mathcal{F}_{\rm NG}(S) = \lim_{\Delta S \to 0} \frac{1}{\Delta S} \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_{n-1} \ldots d\delta_1 \exp\left[-\frac{1}{3!} \langle\, \hat{\delta}_n^3 \,\rangle_c \sum_{j,k,l} \partial_j \partial_k \partial_l + \frac{1}{4!} \langle\, \hat{\delta}_n^4 \,\rangle_c \sum_{j,k,l,m} \partial_j \partial_k \partial_l \partial_m + \ldots \right]$$
$$\left(1 + \frac{1}{2} \mathcal{G}_3^{(1,0,0)}(S) \sum_j (S - S_j) \partial_j \sum_{k,l} \partial_k \partial_l + \ldots \right) W^{\rm gm}\,. \tag{49}$$

In the single barrier case, the analysis at this stage was simplified by the existence of a very useful identity which states that, for *any* function $g(\delta_1, \ldots, \delta_n)$,

$$\int_{-\infty}^{\delta_{\rm c}} d\delta_1 \ldots d\delta_n \sum_{j=1}^{n} \partial_j g = \frac{\partial}{\partial \delta_{\rm c}} \int_{-\infty}^{\delta_{\rm c}} d\delta_1 \ldots d\delta_n g\,. \tag{50}$$

This identity is quite easy to prove, in fact it takes only a little thought to see why it should be true. Remarkably, there is a similar identity which turns out to be very useful for the *two* barrier case also, which is not as obvious to write down as (50). As we show in Appendix D.1, for any function $g(\delta_1, \ldots, \delta_n)$ it is also true that

$$\int_{-\infty}^{-\delta_{\rm v}} d\delta_n \int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_{n-1} \ldots d\delta_1 \sum_{j=1}^{n} \partial_j g = -\left. \frac{\partial}{\partial \delta_{\rm v}} \right|_{\delta_{\rm T}} \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_{n-1} \ldots d\delta_1 g\,. \tag{51}$$

Notice that the final derivative holds fixed the sum $\delta_{\rm T} = \delta_{\rm v} + \delta_{\rm c}$ rather than $\delta_{\rm c}$ alone. For brevity, throughout the rest of the paper we will omit the explicit reference to this fact, and simply write $\partial_{\delta_{\rm v}}$ in place of $\partial / \partial \delta_{\rm v}|_{\delta_{\rm T}}$.

[3] In terms of the reduced cumulants $\mathcal{S}_3$, $\mathcal{S}_4$, etc. we have $\varepsilon_1 = \sigma \mathcal{S}_3$, $\varepsilon_2 = \sigma^2 \mathcal{S}_4$, and so on.

We can immediately see why this relation is useful. Consider the equal time exponentiated derivative operator in Eqn. (49). The identity (51) allows us to pull this entire operator outside the integral, exactly like in the single barrier case. The remaining terms involving the integrals and the continuum limit can now be treated individually. We can recognize the first of these as simply $\mathcal{F}_{\rm SvdW}$, and defining the second as

$$\mathcal{F}^{(3,\rm NL)} \equiv \lim_{\Delta S\to 0}\frac{1}{\Delta S}\int_{-\infty}^{-\delta_{\rm v}} d\delta_n \int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_{n-1}\ldots d\delta_1 \frac{1}{2}\mathcal{G}_3^{(1,0,0)}(S)\sum_j (S-S_j)\partial_j \sum_{k,l}\partial_k\partial_l\, W^{\rm gm} \tag{52}$$

(with the notation (3,NL) standing for 3 point, next to leading), the result at this order is

$$\mathcal{F}_{\rm NG} = e^{(1/3!)\varepsilon_1 S^{3/2}\partial_{\delta_{\rm v}}^3 + (1/4!)\varepsilon_2 S^2\partial_{\delta_{\rm v}}^4+\ldots}\left(\mathcal{F}_{\rm SvdW}+\mathcal{F}^{(3,\rm NL)}+\ldots\right). \tag{53}$$

The term $\mathcal{F}^{(3,\rm NL)}$ is tricky to evaluate, and in Appendix D.2 we show that it reduces to

$$\mathcal{F}^{(3,\rm NL)} = -\frac{1}{4\sqrt{2\pi}}\varepsilon_1 c_1 S^{1/2}\partial_{\delta_{\rm v}}^2\int_0^S \frac{d\tilde{S}}{\sqrt{S-\tilde{S}}}\mathcal{F}_{\rm SvdW}(\tilde{S})\,, \tag{54}$$

where we wrote $\mathcal{G}_3^{(1,0,0)}$ in terms of c_1. Note that $\mathcal{F}_{\rm SvdW}$ does depend on both $\delta_{\rm v}$ and $\delta_{\rm T}$, even though this is hidden by our compact notation. From Eqn. (8) we can write

$$\mathcal{F}_{\rm SvdW}(S) = \frac{1}{\sqrt{2\pi}\,S}\sum_{j=-\infty}^{\infty} B_j e^{-B_j^2/2}\quad;\quad B_j\equiv \nu_{\rm v}-2j\nu_{\rm T}\,, \tag{55}$$

and the integral over S in Eqn. (54) can be evaluated exactly, giving

$$\int_0^S \frac{d\tilde{S}}{\sqrt{S-\tilde{S}}}\mathcal{F}_{\rm SvdW}(\tilde{S}) = \sum_{j=-\infty}^{\infty}\frac{1}{S^{1/2}}e^{-B_j^2/2}\,, \tag{56}$$

using which we get an expression for $\mathcal{F}^{(3,\rm NL)}$,

$$\mathcal{F}^{(3,\rm NL)} = -\frac{1}{\sqrt{2\pi}\,S}\frac{1}{4}\varepsilon_1 c_1\sum_{j=-\infty}^{\infty}\left(B_j^2-1\right)e^{-B_j^2/2}\,. \tag{57}$$

Further, noting that in Eqn. (53) the combination $S^{1/2}\partial_{\delta_{\rm v}}$ can be written as $\partial/\partial\nu_{\rm v}|_{\nu_{\rm T},S}\equiv\partial_{\nu_{\rm v}}$, we can write $\mathcal{F}_{\rm NG}$ as

$$\mathcal{F}_{\rm NG} = \frac{1}{\sqrt{2\pi}\,S}e^{(1/3!)\varepsilon_1\partial_{\nu_{\rm v}}^3+(1/4!)\varepsilon_2\partial_{\nu_{\rm v}}^4+\ldots}\sum_{j=-\infty}^{\infty}e^{-B_j^2/2}\left(B_j-\frac{1}{4}\varepsilon_1 c_1\left(B_j^2-1\right)+\ldots\right). \tag{58}$$

Consider the $j=0$ term, for which $B_0=\nu_{\rm v}$. The exponential derivative can be computed using the saddle point approximation as discussed in Ref. [51], and gives precisely the single barrier f.c. rate computed there for the fixed barrier excluding filter effects (see their Eqn. 60 with $\nu\to-\nu_{\rm v}$),

$$\begin{aligned}\sqrt{\frac{2}{\pi}}\,e^{(1/3!)\varepsilon_1\partial_{\nu_{\rm v}}^3+(1/4!)\varepsilon_2\partial_{\nu_{\rm v}}^4+\ldots}&\left[e^{-\nu_{\rm v}^2/2}\left(\nu_{\rm v}-\frac{1}{4}\varepsilon_1 c_1\left(\nu_{\rm v}^2-1\right)+\ldots\right)\right]\\ &= \sqrt{\frac{2}{\pi}}\,\nu_{\rm v}e^{-\frac{1}{2}\nu_{\rm v}^2\left(1+\varepsilon_1\nu_{\rm v}/3+\left(\varepsilon_1^2-\varepsilon_2/3\right)\nu_{\rm v}^2/4\right)}\left(1-\frac{1}{4}\varepsilon_1\nu_{\rm v}(c_1-4)+\ldots\right)\\ &\equiv f_{\rm NG,1-bar}(\nu_{\rm v})\,.\end{aligned} \tag{59}$$

As discussed in detail in Ref. [51], the ellipsis in the second line denotes all terms that are parametrically smaller than the ones written down, which correspond to terms of order $\mathcal{O}(\epsilon\nu_{\rm v}^{-1},\epsilon^2\nu_{\rm v}^2,\epsilon^3\nu_{\rm v}^5)$, where $\epsilon\ll 1$ is the parameter controlling the NG via $\varepsilon_n\sim\epsilon^n$. We will assume that the largest scales we access satisfy $\epsilon\nu_{\rm v}^3\sim\mathcal{O}(1)$. One can check that for $f_{\rm NL}\sim 100$ this corresponds to Lagrangian scales $R\sim 25h^{-1}$Mpc, or very large voids (see Eqn. (2)). In this case one finds that the ignored terms listed above are all of the same order of magnitude.

At all smaller scales, the term of the form $\epsilon\nu_{\rm v}^{-1}$ gives the largest theoretical error. We will use this in our arguments below.

For $j \neq 0$, notice first that the B_j are all linear in $\nu_{\rm v}$, so that the effect of a derivative $\partial_{\nu_{\rm v}}|_{\nu_{\rm T}}$ is identical to that of a derivative w.r.t B_j. We therefore have

$$\mathcal{F}_{\rm NG} = \frac{1}{\sqrt{2\pi}\,S}\sum_{j=-\infty}^{\infty} e^{(1/3!)\varepsilon_1\partial_{B_j}^3+(1/4!)\varepsilon_2\partial_{B_j}^4+\ldots}\left[e^{-B_j^2/2}\left(B_j - \frac{1}{4}\varepsilon_1 c_1\left(B_j^2-1\right)+\ldots\right)\right]. \tag{60}$$

At least formally, one might say that the solution is therefore just a series of single barrier results, leading to a multiplicity $f_{\rm NG} = 2S\mathcal{F}_{\rm NG}$ given by

$$\begin{aligned} f_{\rm NG}(\nu_{\rm v},\nu_{\rm T}) &= \sqrt{\frac{2}{\pi}}\sum_{j=-\infty}^{\infty} B_j e^{-\frac{1}{2}B_j^2\left(1+\varepsilon_1 B_j/3+\left(\varepsilon_1^2-\varepsilon_2/3\right)B_j^2/4+\ldots\right)}\left(1-\frac{1}{4}\varepsilon_1 B_j(c_1-4)+\ldots\right) \\ &= \sum_{j=-\infty}^{\infty} f_{\rm NG,1-bar}(B_j)\,, \end{aligned} \tag{61}$$

with $B_j = \nu_{\rm v} - 2j\nu_{\rm T}$ and $f_{\rm NG,1-bar}$ defined in Eqn. (59). The problem with this expression is that for *any* fixed $\nu_{\rm v}$, $\nu_{\rm T}$ and NG parameters ε_1, ε_2, etc., for large enough j the terms being ignored in the ellipsis will become comparable to the ones being retained. This is not such an important issue for the polynomial NG terms, which on their own would always be suppressed by the Gaussian factor $e^{-\frac{1}{2}B_j^2}$. This is also the reason why analyses such as those of Lam *et al.* [48] and Kamionkowski *et al.* [47], which are based on the Edgeworth expansion and therefore have multiplicities of the form $e^{-\nu^2/2}$ multiplying a polynomial in ν, are not susceptible to the void-in-cloud issue for large enough voids. As D'Amico *et al.* [51] argued however, the Edgeworth series results break down when the combination $\epsilon\nu^3$ becomes of order unity. The D'Amico *et al.* analysis, which we have used here, instead effectively resums potentially troublesome terms and leads to the non-trivial series in the exponential in $f_{\rm NG,1-bar}(\nu)$, which can give significantly different results for the halo multiplicity than the Edgeworth-like analysis (see e.g. Fig. 5 of Ref. [51]).

In the two barrier case as well, these exponentiated terms are expected to be important for the $j = 0$ piece at large enough $\nu_{\rm v}$. However for $j \neq 0$ they are problematic, being a series in the supposedly small parameter ϵB_j. Clearly $|\epsilon B_j|$ becomes larger than unity for large enough j (positive or negative) at *any* fixed $\nu_{\rm v}$, and the series expansions for all j values beyond this point break down. This is not surprising, considering that this expression depends on the saddle point approximation, which for each j is only valid provided $|\epsilon B_j| < 1$ [51]. A more intuitive way of understanding this breakdown is to note that the result for the f.c. rate is effectively a series of *single barrier* f.c. rates with successively larger barrier heights. Since each single barrier f.c. rate involves the behaviour of a non-Gaussian conditional p.d.f. ($\Pi_{1-{\rm bar},\Delta{\rm S}=0}$) at the barrier, for successively larger j we are effectively sampling further and further extremes of the non-Gaussian tails of the distributions, which eventually can no longer be described perturbatively. In practice of course, we don't expect that the resummation of this series would dominate the Gaussian suppression to the extent of giving order unity features in the single barrier f.c. rate.

In fact one can make a stronger statement based on the limit in which we artificially send $\delta_{\rm c} \to \infty$, which *must* recover the single barrier result for the void barrier. In the expression (61), sending $\delta_{\rm c} \to \infty$ for fixed $\delta_{\rm v}$ and S is the same as sending $\nu_{\rm T} \to \infty$ at fixed $\nu_{\rm v}$. In this limit, if the single barrier result is to be recovered, then *each* term with $j \neq 0$ must individually vanish. For a given j, as $\nu_{\rm T} \to \infty$ we have $B_j \to \pm\infty$ with the sign depending on the sign of j. A given term with $j \neq 0$ will then vanish only if the series $1+\varepsilon_1 B_j/3+\left(\varepsilon_1^2-\varepsilon_2/3\right)B_j^2/4+\ldots$ which appears in the exponential, resums as $|B_j| \to \infty$ into a form which is bounded both above and below by strictly positive numbers. In other words, the exponential suppression of the single barrier multiplicity must qualitatively remain intact. We will make the mild assumption that the lower bound on the resummed series is not arbitrarily close to zero but is closer to unity, which excludes pathological features such as e.g. sharp repeating spikes in the single barrier multiplicity with a slowly decreasing maximum height.

With this we can extend the argument to finite $\nu_{\rm T}$, by noting that in $f_{\rm NG,1-bar}(B_j)$ sending $\nu_{\rm T} \to \infty$ for fixed j is the same as sending $|j| \to \infty$ for fixed $\nu_{\rm T}$. The previous paragraph immediately implies that

the contribution of terms with increasing $|j|$ is progressively suppressed. Consider first a term in which the series expansion has broken down, so that $|\epsilon B_j| = \alpha_0 > 1$ (while $\epsilon\nu_{\rm v}$ is still significantly less than unity). The arguments above suggest that the quantity $f_{\rm NG,1-bar}(B_j)$ resums to the form

$$f_{\rm NG,1-bar}(B_j) \sim \alpha_1 B_j e^{-\frac{1}{2}\alpha_2 B_j^2} = \frac{1}{\epsilon}\alpha_1\alpha_0 e^{-\frac{1}{2\epsilon^2}\alpha_2\alpha_0^2}\,, \tag{62}$$

where α_1 and α_2 are positive numbers with a possible mild dependence on $|B_j|$. Let us compare this term with the biggest term err_0 we do not calculate in $f_{\rm NG,1-bar}(B_0)$, which is

$$\mathrm{err}_0 = \nu_{\rm v} e^{-\frac{1}{2}\nu_{\rm v}^2\left(1+\varepsilon_1\nu_{\rm v}/3+\left(\varepsilon_1^2-\varepsilon_2/3\right)\nu_{\rm v}^2/4\right)} \times \mathcal{O}\left(\epsilon\nu_{\rm v}^{-1}\right)\,. \tag{63}$$

The ratio of these terms is

$$r \equiv \frac{f_{\rm NG,1-bar}(B_j)}{\mathrm{err}_0} \sim \frac{\alpha_1\alpha_0}{\epsilon^2} e^{-\frac{1}{2\epsilon^2}\left(\alpha_2\alpha_0^2-(\epsilon\nu_{\rm v})^2(1+\mathcal{O}(\epsilon\nu_{\rm v}))\right)}\,, \tag{64}$$

and as long as this ratio is less than unity the error we make by ignoring $f_{\rm NG,1-bar}(B_j)$ is smaller than the one made by truncating the series in $f_{\rm NG,1-bar}(B_0)$. We argued above that the quantity α_2 is strictly positive. Moreover we expect α_1 to be of the same order as α_2, since both are proxies for resummed series which are structurally similar. Since we cannot compute these objects, let us separately analyse the situations where α_2 and α_1 are of order unity or are much smaller.

- If α_2, α_1 are of order unity, then since $\alpha_0 > 1$ the first term in the exponential is much larger than the term containing $(\epsilon\nu_{\rm v})^2$, and clearly we have $r \ll 1$.
- Even if α_2, α_1 are not close to unity, in practice r remains small provided only that α_2 is not very close to zero, this assumption following from our comment above on the lower bound on the exponentiated series. For example if we assume $\alpha_2, |\alpha_1| > (\epsilon\nu_{\rm v})^2$ then we find

$$r < \alpha_0\nu_{\rm v}^2 e^{-\frac{1}{2}\nu_{\rm v}^2(\alpha_0^2-1)}\,. \tag{65}$$

  Now even if α_0 is not much larger than unity the ratio r will remain small for all interesting values of $\nu_{\rm v}$. Say $\alpha_0 \simeq 1.5$, then r is already less than unity at $\nu_{\rm v} = 1$, which is the lower limit for our formalism in any case.

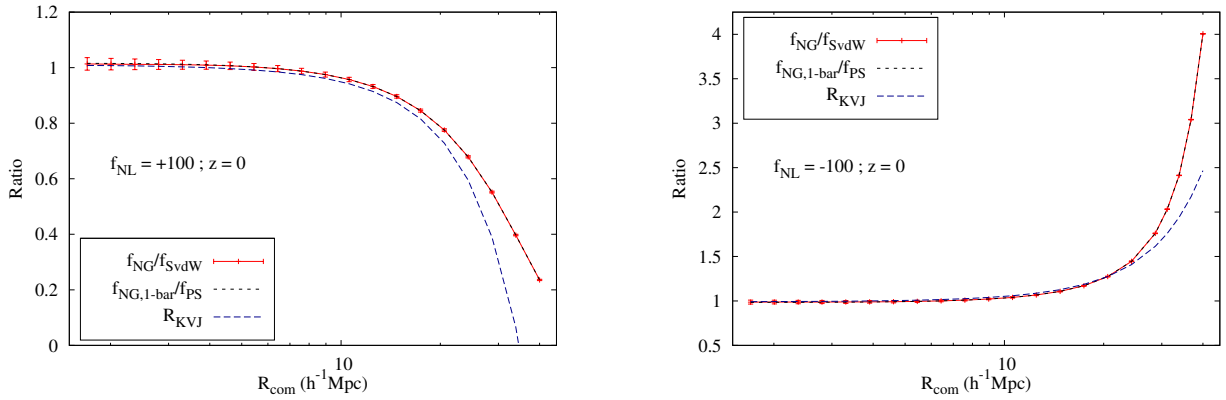
To summarize, j-values for which the series expansion in $|\epsilon B_j|$ has formally broken down (even mildly) are not expected to give contributions larger than the terms which are already being ignored in the $j = 0$ term.

Finally, for terms in Eqn. (61) with non-zero j where the series expansion has *not* broken down, one might still expect that the contribution of these terms is always smaller than err_0. We will now see that this is not quite true, although it is possible to ignore the $j \neq 0$ terms if we only probe the largest voids. We start by comparing the leading order piece of $f_{\rm NG,1-bar}(B_j)$ which is given by $B_j e^{-\frac{1}{2}B_j^2(1+\mathcal{O}(\epsilon B_j))}$, with err_0. The ratio of these terms is

$$\left|\frac{B_j e^{-\frac{1}{2}B_j^2(1+\mathcal{O}(\epsilon B_j))}}{\mathrm{err}_0}\right| = \frac{B_j}{\epsilon} e^{-\frac{1}{2}\left(B_j^2-\nu_{\rm v}^2+\ldots\right)}\,, \tag{66}$$

with the ellipsis denoting parametrically smaller quantities. In order to be able to ignore this term, we need to place an upper bound on the above ratio and show that this bound is less than unity. Since this entire discussion is valid only if $f_{\rm NL} \neq 0$ (else we simply use $f_{\rm SvdW}$), suppose now that we have a lower bound for $|f_{\rm NL}|$, say $|f_{\rm NL}| > 1$ in the local model, which translates to $\epsilon > 2\cdot 10^{-4}$ or $\epsilon^{-1} < 5000$. Writing $B_j = \nu_{\rm v}\left(1 - 2j\delta_{\rm T}/\delta_{\rm v}\right)$ with $\delta_{\rm T}/\delta_{\rm v} \approx 5/3$ for $\delta_{\rm c} \simeq 1.7$ and $\delta_{\rm v} \simeq 2.7$, we can see that $|B_j| > 2\nu_{\rm v}$ for all $j \neq 0$. Together with the bound on ϵ, this allows us to calculate a minimum value $\nu_{\rm v\,min}$ such that the ratio (66) is always less than unity. We do this by setting $|B_j| = 2\nu_{\rm v}$ since all larger values will give smaller ratios, and find that the ratio (66) is guaranteed to be less than unity if $\nu_{\rm v} > \nu_{\rm v\,min} \simeq 2.5$ corresponding to voids with Lagrangian radius larger than $\sim 6.5h^{-1}$Mpc (or comoving radius larger than $\sim 11h^{-1}$Mpc) in a *WMAP*-compatible ΛCDM cosmology [20]. If we assume a smaller lower bound on $|f_{\rm NL}|$, the value of $\nu_{\rm v\,min}$ will increase slowly.

We do not need to be this conservative however. We can also consider values $\nu_{\rm v} < \nu_{\rm v\,min}$ by retaining an appropriate number of terms with $j \neq 0$, depending on the chosen value of $f_{\rm NL}$ or ϵ. In fact as we will



(a) (b)

Figure 2: The ratio of non-Gaussian to Gaussian multiplicity for $f_{\rm NL} = +100$ (left panel) and $f_{\rm NL} = -100$ (right panel) in the local model, at redshift $z = 0$. The two barrier and single barrier ratios are indistinguishable, as discussed in the text. There is a clear departure at large radii from the ratio $R_{\rm KVJ}$ proposed by Kamionkowski *et al.* [47], which is the one calculated by LoVerde *et al.* [28]. The error bars represent theoretical errors of order $\mathcal{O}(\epsilon\nu_{\rm v}^{-1})$ as discussed in the text.

see presently, in practice there is an even simpler way of accounting for the void-in-cloud effect. Firstly, our arguments above indicate that at a given $\nu_{\rm v} > 1$ we must keep the leading order behaviour (schematically $\sim B_j e^{-\frac{1}{2}B_j^2(1+\epsilon B_j+\ldots)}$) of those $j \neq 0$ terms for which $|\epsilon B_j| < 1$ and the ratio in Eqn. (66) is larger than unity. (An analysis similar to the one above shows that the subleading terms can always be ignored in this case, so that the dominant error is still given by err_0.) This is straightforward to implement numerically, and in Fig. 2 we show the ratio $f_{\rm NG}(\nu_{\rm v},\nu_{\rm T})/f_{\rm SvdW}(\nu_{\rm v},\nu_{\rm T})$ (solid red) as a function of $R_{\rm com}$ for $f_{\rm NL} = \pm 100$ at $z = 0$ with $\delta_{\rm v} = 2.72$ and $\delta_{\rm c} = 1.686$. We also plot the ratio of the *single barrier* functions $f_{\rm NG,1-bar}(\nu_{\rm v})/f_{\rm PS}(\nu_{\rm v})$ (short-dashed black), see Eqn. (59). As we see, these curves are indistinguishable, and one can also check that even at the smallest radii we consider, the difference between the curves is less than 0.5%. In fact this result is also easy to see analytically, by recognizing that the $j \neq 0$ terms give increasingly smaller contributions to the sum in $f_{\rm NG}(\nu_{\rm v},\nu_{\rm T})$. One then has

$$\frac{f_{\rm NG}(\nu_{\rm v},\nu_{\rm T})}{f_{\rm SvdW}(\nu_{\rm v},\nu_{\rm T})} = \frac{f_{\rm NG,1-bar}(\nu_{\rm v}) + \sum_{j\neq 0} f_{\rm NG,1-bar}(B_j)}{f_{\rm PS}(\nu_{\rm v}) + \sum_{j\neq 0} f_{\rm PS}(B_j)}$$
$$\approx \frac{f_{\rm NG,1-bar}(\nu_{\rm v})}{f_{\rm PS}(\nu_{\rm v})}\left[1 + \sum_{j\neq 0}\left(\frac{f_{\rm NG,1-bar}(B_j)}{f_{\rm NG,1-bar}(\nu_{\rm v})} - \frac{f_{\rm PS}(B_j)}{f_{\rm PS}(\nu_{\rm v})}\right)\right], \qquad (67)$$

where we linearized in the $j \neq 0$ terms. Now, for large $\nu_{\rm v}$ the summation in the second line of (67) will be suppressed simply because of the Gaussian factor $e^{-\frac{1}{2}(B_j^2-\nu_{\rm v}^2)}$ in each term. For $\nu_{\rm v} \to 1$ this supression will not be very strong at least for small values of $|j| \neq 0$. However, for such terms the *single barrier* ratio $f_{\rm NG,1-bar}/f_{\rm PS}$ approaches a constant, leading to a cancellation of the terms in the summation above. In practice therefore, to extremely good accuracy the two barrier non-Gaussian multiplicity can be simply written as the product of the single barrier non-Gaussian ratio with the two barrier *Gaussian* multiplicity $f_{\rm SvdW}$,

$$f_{\rm NG}(\nu_{\rm v},\nu_{\rm T}) = \sum_{j=-\infty}^{\infty} f_{\rm NG,1-bar}(B_j) \approx \left(\frac{f_{\rm NG,1-bar}(\nu_{\rm v})}{f_{\rm PS}(\nu_{\rm v})}\right) \times f_{\rm SvdW}(\nu_{\rm v},\nu_{\rm T})\,, \qquad (68)$$

where $f_{\rm NG,1-bar}(\nu)$ was defined in Eqn. (59), $f_{\rm PS}(\nu)$ in Eqn. (20) and $B_j = \nu_{\rm v} - 2j\nu_{\rm T}$. Of course this discussion is subject to the caveat that there is always a theoretical error at least due to the terms we do not compute in the single barrier multiplicity. For comparison, in Fig. 2 we also show (dashed blue) the ratio proposed by Kamionkowski *et al.* [47], which (as expected) deviates from our prediction at large radii. The same will be

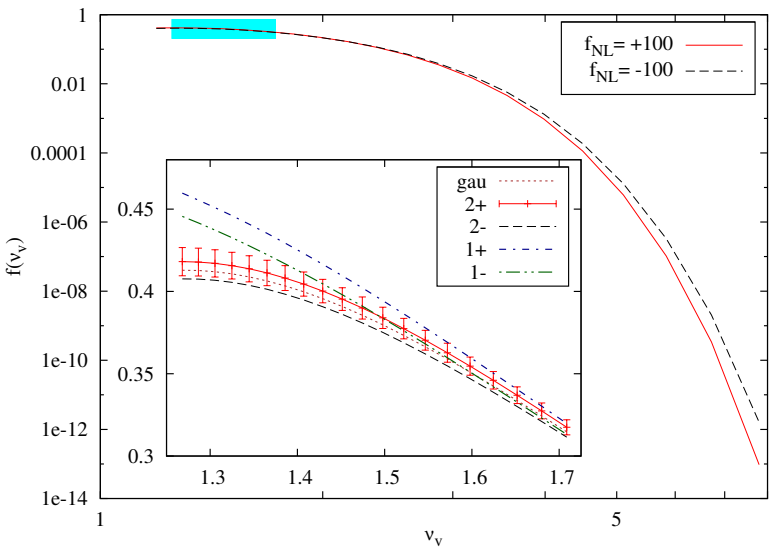


Figure 3: The two barrier non-Gaussian multiplicity $f_{\rm NG}$ as a function of $\nu_{\rm v}$, for $f_{\rm NL} = \pm 100$. The inset shows a zoomed in view of the range highlighted by the cyan (shaded) area, corresponding to the range $R_{\rm com}$ = 2-4h^{-1}Mpc. The curves in the inset correspond to $f_{\rm SvdW}$ (gau), $f_{\rm NG}^{(f_{\rm NL}=\pm 100)}$ (2$\pm$) and $f_{\rm NG,1-bar}^{(f_{\rm NL}=\pm 100)}$ (1$\pm$). The single barrier result overestimates the void abundance by order $\sim 10\%$ for *both* signs of $f_{\rm NL}$. The correct, two barrier result on the other hand displays a behaviour opposite to that at large $R_{\rm com}$: positive $f_{\rm NL}$ slightly enhances the abundance at small $R_{\rm com}$. Also shown are the theoretical errors on the two barrier result as discussed in the text. For clarity we only show these for $f_{\rm NL} = +100$.

true of the ratio calculated by Lam *et al.* [48] at large radii. Fig. 2 also illustrates the complementary nature of the void abundances as a probe of NG, since the abundance at large radii is reduced compared to the Gaussian case for positive values of $f_{\rm NL}$ and vice versa for negative $f_{\rm NL}$, which is the opposite of what happens for halo abundances.

In Fig. 3 we show the non-Gaussian multiplicity as a function of $\nu_{\rm v}$, together with a zoomed in view at small comoving radii ($\nu_{\rm v} \gtrsim 1$). We see that ignoring the $j \neq 0$ void-in-cloud terms entirely will overpredict the void abundance at small radii for either sign of $f_{\rm NL}$. On the other hand, accounting for void-in-cloud effects gives a result which is approximately the same as the *Gaussian* one, with a slight enhancement for positive $f_{\rm NL}$ and a slight reduction for negative $f_{\rm NL}$. Notice that this is the opposite of what happens at *large* radii. Unfortunately the magnitude of the reversed effect at small radii appears to be too small to be observationally relevant. Finally, in Fig. 4 we show the differential comoving number density $dn_{\rm com}/d\log R_{\rm com}$ defined in Eqn. (1) as a function of comoving radius $R_{\rm com}$.

5 Discussion

Primordial non-Gaussianity (NG) can be probed by the imprints it leaves on the late time large scale structure of the universe by modifying the distribution of matter, which is manifested for example in the abundance of collapsed objects. In this paper we have explored a second manifestation of this effect, which is the abundance of *voids* or underdense regions. While such calculations have been performed earlier in the literature [47, 48], they have been subject at least to the caveat that their treatment of the NG was based on a linearization in $f_{\rm NL}$, which is known to potentially misestimate the abundance of very massive objects [51]. Our calculation was based on path integral techniques introduced by Maggiore & Riotto [52, 50], and importantly also on the improvements to these techniques developed in Ref. [51]. The latter allow us to access a larger range of length scales than treatments based on linearizing in $f_{\rm NL}$, while the use of path integrals also allows us to carefully account for the "void-in-cloud" issue first pointed out by Sheth & van de Weygaert [37], which is unique to the case of voids. We showed that in the final analysis, the complication introduced by the void-in-cloud issue

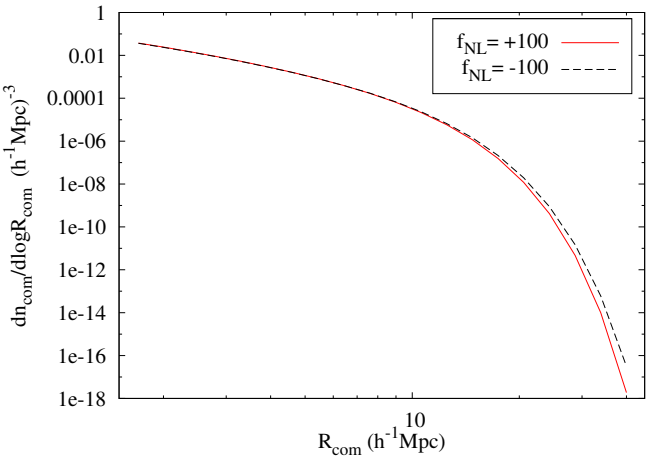


Figure 4: Comoving number density defined in Eqn. (1), as a function of comoving radius $R_{\rm com}$ at redshift $z = 0$ for $f_{\rm NL} = \pm 100$.

can actually be accounted for in a fairly simple manner, with the void multipicity $f_{\rm NG}(\nu_{\rm v}, \nu_{\rm T})$ being given by the approximation in Eqn. (68). Our treatment of the multi-scale correlations arising from primordial NG is more rigorous than that of Lam *et al.* [48] (although these effects are small), and additionally our explicit expression for the multiplicity is different from theirs due to the presence of terms involving the NG functions ε_1, ε_2, etc. in an exponential (see e.g. Eqn. (59)). Our prediction for the void multiplicity (or number density) is also simple to implement numerically, especially when combined with the fits for the NG functions given in Ref. [51].

It is worth spending a moment to compare the relative merits of using voids as opposed to halos as a probe of primordial NG. In terms of absolute numbers, a given survey (or simulation box) will always contain fewer voids of some Lagrangian scale R than halos of the corresponding mass scale $M \propto R^3$. This is simply because the void barrier height $\delta_{\rm v}$ is larger than the halo barrier height $\delta_{\rm c}$, leading to a stronger cutoff $\sim e^{-\frac{1}{2}\nu_{\rm v}^2}$ for the void abundance. Furthermore on comparing the typical masses of the largest clusters observed ($M \sim 10^{15} h^{-1} M_{\rm sol}$) with the typical comoving sizes of the largest voids ($R_{\rm com} \sim 50 h^{-1}$Mpc), one finds that the largest *Lagrangian* length scales probed by both halos and voids are roughly the same ($R \sim 25 h^{-1}$Mpc). In this sense voids would be a poorer statistical probe of primordial NG than halos. The strength of voids however comes from the complementarity which is demonstrated in Fig. 2, and was also highlighted in Refs. [47, 48]. In contrast to the halo abundance which is e.g. reduced compared to the Gaussian for negative $f_{\rm NL}$, the void abundance is enhanced at large radii, and vice versa for positive $f_{\rm NL}$. It will be interesting to see how these characteristics ultimately play out in determining the constraining power of voids.

Our work can be extended in more than one direction. Firstly, our arguments regarding the terms with large $|j|$ in Eqn. (68) were somewhat qualitative, although we expect them to be robust. In a future work, we will test these arguments by numerically generating the appropriate random walks and explicitly determining the multiplicity using the resulting distribution of conditional first crossing times [56]. Of course it will also be interesting to compare our predictions (and those of others) with full-fledged N-body simulations. On the analytical side, it will be interesting to study the regime $\nu_{\rm v} \gtrsim 1$ in more detail, since apart from void-in-cloud effects it is also likely that triaxial effects would become important here [57, 58]. It would be interesting to try and account for such effects (even approximately) within our framework, perhaps along the lines explored in Ref. [59] (see also Refs. [60, 48]). Finally, an equally interesting avenue would be to convert our predictions which hold for the dark matter distribution, to predictions for voids in the *galaxy* distribution, perhaps by generalizing the treatment of Ref. [46] to the non-Gaussian case.

Acknowledgements

It is a pleasure to thank Paolo Creminelli, Michele Maggiore, Jorge Moreno and Licia Verde for useful discussions, and especially Ravi Sheth for detailed discussions and comments on an earlier draft.

Appendix

A Primordial non-Gaussianity

The physics of inflation governs the statistics of the initial seeds of inhomogeneities which grow into the large scale structure we see today. These initial conditions can be characterised by the primordial comoving curvature perturbation $\mathcal{R}(\vec{x})$ (with Fourier transform $\mathcal{R}(\mathbf{k})$) which remains constant on superhorizon scales. The function $\langle\, \mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\mathcal{R}(\mathbf{k}_3)\,\rangle_c$ (with the subscript denoting the connected part) is then an example of a function which probes the physics of the inflationary epoch. By translational invariance, it is proportional to a momentum-conserving delta function:

$$\langle\, \mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\mathcal{R}(\mathbf{k}_3)\,\rangle_c = (2\pi)^3\delta_D(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)B_{\mathcal{R}}(k_1,k_2,k_3)\,, \tag{A.1}$$

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 $k_1 \ll k_2 \simeq k_3$, and the equilateral NG, peaked on equilateral triangles $k_1 \simeq k_2 \simeq k_3$. 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 [61]. 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 [62, 63] or in the inhomogeneous reheating scenario [64]. In these models, the curvature perturbation can be written in the following form,

$$\mathcal{R}(\mathbf{x}) = \mathcal{R}_g(\mathbf{x}) + \frac{3}{5}f_{\rm NL}^{\rm loc}\left(\mathcal{R}_g^2(\mathbf{x}) - \langle\,\mathcal{R}_g^2\,\rangle\right) + \frac{9}{25}g_{\rm NL}\mathcal{R}_g^3(\mathbf{x})\,, \tag{A.2}$$

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

$$B_{\mathcal{R}}(k_1,k_2,k_3) = \frac{6}{5}f_{\rm NL}^{\rm loc}\left[P_{\mathcal{R}}(k_1)P_{\mathcal{R}}(k_2) + \text{cycl.}\right]\,, \tag{A.3}$$

where "cycl." denotes the 2 cyclic permutations of the wavenumbers, and $P_{\mathcal{R}}(k)$ is the power spectrum given by $P_{\mathcal{R}}(k) = Ak^{n_s-4}$. The trispectrum is given by

$$\begin{aligned}\langle\, \mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\mathcal{R}(\mathbf{k}_3)\mathcal{R}(\mathbf{k}_4)\,\rangle_c &= (2\pi)^3\delta_D(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3+\mathbf{k}_4)\\ &\times\left[\frac{36}{25}f_{\rm NL}^2\sum_{\substack{b<c\\ a\neq b,c}}P_{\mathcal{R}}(|\mathbf{k}_a+\mathbf{k}_b|)P_{\mathcal{R}}(k_b)P_{\mathcal{R}}(k_c) + \frac{54}{25}g_{\rm NL}\sum_{a<b<c}P_{\mathcal{R}}(k_a)P_{\mathcal{R}}(k_b)P_{\mathcal{R}}(k_c)\right]\,.\end{aligned} \tag{A.4}$$

The equilateral model:

Models with derivative interactions of the inflaton field [65, 66, 67] 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. [68],

$$B_{\mathcal{R}}(k_1,k_2,k_3)=\frac{18}{5}f_{\rm NL}^{\rm equil}A^2\Big[\frac{1}{2k_1^{4-n_s}k_2^{4-n_s}}+\frac{1}{3(k_1k_2k_3)^{2(4-n_s)/3}}-\frac{1}{(k_1k_2^2k_3^3)^{(4-n_s)/3}}+5\text{ perms.}\Big]\,. \tag{A.5}$$

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.

To connect the statistics of $\mathcal{R}$ with large scale structure, we use the fact that the excursion set ansatz only requires us to know the linearly extrapolated present day behaviour of the density contrast δ_R smoothed on various length scales. We can relate this quantity to the initial conditions $\mathcal{R}$ via the following relations. We start from the Bardeen potential Φ on subhorizon scales, given by

$$\Phi(\mathbf{k},z)=-\frac{3}{5}T(k)\frac{D(z)}{a}\mathcal{R}(k)\,, \tag{A.6}$$

where $T(k)$ 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 (see e.g. Ref. [69]); 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 cosmological Poisson equation, which in Fourier space reads

$$\begin{aligned}\delta(\mathbf{k},z)&=-\frac{2ak^2}{3\Omega_m H_0^2}\Phi(\mathbf{k},z)=\frac{2k^2}{5\Omega_m H_0^2}T(k)D(z)\mathcal{R}(k)\\&\equiv\mathcal{M}(k,z)\mathcal{R}(k)\,,\end{aligned} \tag{A.7}$$

with Ω_m the present time fractional density of matter (cold dark matter and baryons), and $H_0=100h\,\mathrm{km\,s^{-1}Mpc^{-1}}$ the present time Hubble constant. The redshift dependence is trivially accounted for by the linear growth factor $D(z)$. Introducing a filter function $W_R(|\mathbf{x}|)$, the smoothed density field (around one point, which we take as the origin) is given by

$$\delta_R(z)=\int\frac{d^3k}{(2\pi)^3}\widetilde{W}(kR)\delta(\mathbf{k},z)\,, \tag{A.8}$$

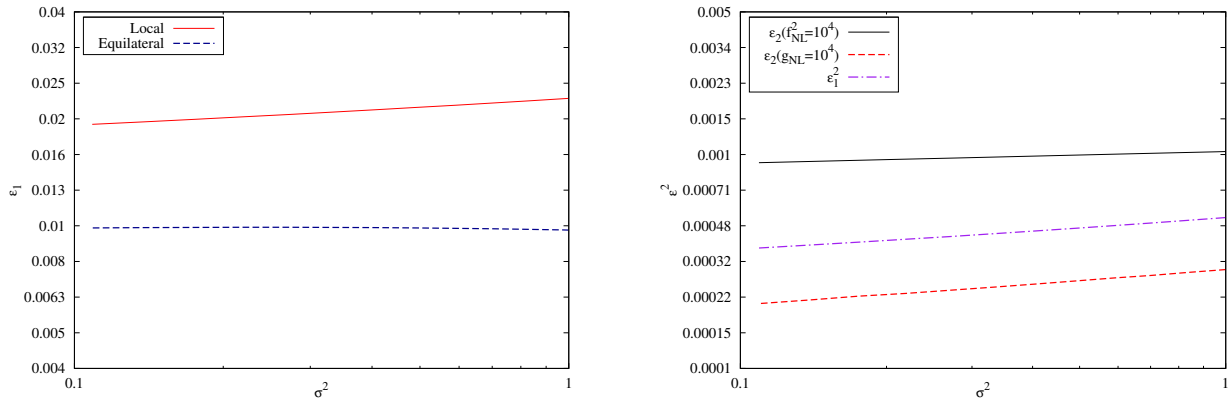
where $\widetilde{W}(kR)$ is the Fourier transform of the filter function. The results of this paper are strictly valid only for a sharp filter in k-space, although a physically more relevant filter would be the spherical top-hat filter in real space, whose Fourier transform $\widetilde{W}(kR)$ is given by

$$\widetilde{W}(y)=\frac{3}{y^3}\left(\sin y-y\cos y\right)\,. \tag{A.9}$$

While this choice allows us to have a well-defined relation between length scales and masses, namely $M=(4\pi/3)\Omega_m\rho_cR^3$ with $\rho_c=3H_0^2/(8\pi G)=2.77\cdot10^{11}h^{-1}M_{\rm sol}(h^{-1}\mathrm{Mpc})^{-3}$, it spoils the Markovianity of the random walk of δ_R (see e.g. Ref. [52]). We will therefore continue to present results for the sharp-k filter. By using Eqns. (A.8) and (A.7) we have, for the 3-point function,

$$\langle\,\delta_{R_1}\delta_{R_2}\delta_{R_3}\,\rangle_c=\int\frac{d^3k_1}{(2\pi)^3}\frac{d^3k_2}{(2\pi)^3}\frac{d^3k_3}{(2\pi)^3}\widetilde{W}(k_1R_1)\widetilde{W}(k_2R_2)\widetilde{W}(k_3R_3)\mathcal{M}(k_1)\mathcal{M}(k_2)\mathcal{M}(k_3)\langle\,\mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\mathcal{R}(\mathbf{k}_3)\,\rangle_c\,, \tag{A.10}$$

where we suppressed the redshift dependence, and analogous formulae are valid for the higher order correlations. Fig. 5 shows the behaviour of the NG functions ε_1 and ε_2 defined in the text (Eqn. (44)) as a function of $S=\sigma_R^2$ in the local and equilateral cases. (ε_2 is shown only for the local case.) We see that these functions remain approximately constant over a range of S values which corresponds to roughly the range 2-25h^{-1}Mpc of Lagrangian smoothing length scales.



(a) (b)

Figure 5: Scale dependence of the ε_n. Panel (a) : Behaviour of ε_1 vs. σ^2 in the local and equilateral models, for $f_{\rm NL} = 100$ in each case. Panel (b) : Behaviour of ε_2 for the local model with $f_{\rm NL} = 100$ and $g_{\rm NL} = 10^4$. The terms proportional to $f_{\rm NL}^2$ and $g_{\rm NL}$ are shown separately. Also shown is ε_1^2 for the same model. The axes are logscale.

B Some results concerning statistics of random walks

B.1 Expressing $\mathcal{F}_{\rm SvdW}$ as a sum of Gaussians

Let $\mathcal{F}$ denote the r.h.s of Eqn. (8). To prove the equivalence of the two expressions for $\mathcal{F}_{\rm SvdW}$ in Eqns. (5) and (8), we will show that the Laplace transform $\mathcal{L}(\omega)$ of $\mathcal{F}$ is the SvdW result (6). We have

$$\begin{aligned}
\mathcal{L}(\omega) &= \int_0^{+\infty} {\rm d}S\, e^{-\omega S} \mathcal{F}(S) \\
&= \sum_{j=-\infty}^{+\infty} \frac{\delta_{\rm v} - 2j\delta_{\rm T}}{(2\pi)^{1/2}} \int_0^{+\infty} {\rm d}S\, \frac{1}{S^{3/2}} e^{-\omega S - (\delta_{\rm v} - 2j\delta_{\rm T})^2/2S} \\
&= -\sum_{j=-\infty}^{+\infty} {\rm sgn}(2j\delta_{\rm T} - \delta_{\rm v})\, e^{-|2j\delta_{\rm T} - \delta_{\rm v}|\sqrt{2\omega}} \\
&= e^{-\delta_{\rm v}\sqrt{2\omega}} \sum_{j\le 0} e^{2j\delta_{\rm T}\sqrt{2\omega}} - e^{\delta_{\rm v}\sqrt{2\omega}} \sum_{j>0} e^{-2j\delta_{\rm T}\sqrt{2\omega}}\,,
\end{aligned} \tag{B.1}$$

where the integral in the second line can be calculated using Eqn. 3.472(5) of Ref. [70], and we used the fact that $2j\delta_{\rm T} - \delta_{\rm v}$ is negative for $j \le 0$. By changing $j \to -j$ in the first summation in the last line one gets

$$\begin{aligned}
\mathcal{L}(\omega) &= e^{-\delta_{\rm v}\sqrt{2\omega}} \sum_{j=0}^{+\infty} \left(e^{-2\delta_{\rm T}\sqrt{2\omega}}\right)^j - e^{\delta_{\rm v}\sqrt{2\omega}} \sum_{j=1}^{+\infty} \left(e^{-2\delta_{\rm T}\sqrt{2\omega}}\right)^j \\
&= \frac{e^{-\delta_{\rm v}\sqrt{2\omega}} - e^{\delta_{\rm v}\sqrt{2\omega}} e^{-2\delta_{\rm T}\sqrt{2\omega}}}{1 - e^{-2\delta_{\rm T}\sqrt{2\omega}}} = \frac{e^{\delta_{\rm c}\sqrt{2\omega}} - e^{-\delta_{\rm c}\sqrt{2\omega}}}{e^{\delta_{\rm T}\sqrt{2\omega}} - e^{-\delta_{\rm T}\sqrt{2\omega}}}\,,
\end{aligned} \tag{B.2}$$

which reproduces exactly the result of Eqn. (6).

B.2 Wrong side counting for the single barrier

Here we show that the wrong side (WS) counting argument for the single barrier, is equivalent to MR's calculation of the f.c. rate for generic random walks. For a single barrier problem with discrete time steps, the probability for the walk to remain below the barrier $\delta_{\rm c}$ for the first $n-1$ steps, and to cross the barrier at the $n^{\rm th}$ step is $\int_{\delta_{\rm c}}^{\infty} d\delta_n \int_{-\infty}^{\delta_{\rm c}} d\delta_{n-1} \ldots d\delta_1 W(\delta_0; \{\delta_k\}_n; S_n)$. Here $W(\delta_0; \{\delta_k\}_n; S_n)$ defined in Eqn. (14) is the p.d.f.

of a completely generic random walk – we are *not* assuming that the walk is Markovian or even Gaussian. Using an argument very similar to the one used in writing Eqn. (23), the WS f.c. rate is then given by

$$\mathcal{F}_{\mathrm{WS,1-bar}}(S) = \lim_{\Delta S\to 0}\frac{1}{\Delta S}\int_{\delta_c}^{\infty} d\delta_n \int_{-\infty}^{\delta_c} d\delta_{n-1}\ldots d\delta_1 W(\delta_0;\{\delta_k\}_n;S_n = S)\,. \tag{B.3}$$

In general we also have the identity

$$\int_{-\infty}^{\infty} d\delta_n \int_{-\infty}^{\delta_c} d\delta_{n-1}\ldots d\delta_1 W(\delta_0;\{\delta_k\}_n;S_n) = \int_{-\infty}^{\delta_c} d\delta_{n-1}\ldots d\delta_1 W(\delta_0;\{\delta_k\}_{n-1};S_{n-1})\,, \tag{B.4}$$

which is a simple marginalisation over the location of the walk at the final time step, and follows from the definition of $W(\delta_0;\{\delta_k\}_n;S_n)$. Let $P_{\Delta S}(>S_n)$ denote the probability that the f.c. time is larger than S_n. Since this is the same as the probability that the walk has not crossed the barrier in the first n steps, we have

$$P_{\Delta S}(>S_n) = \int_{-\infty}^{\delta_c} d\delta_n\ldots d\delta_1 W(\delta_0;\{\delta_k\}_n;S_n)\,, \tag{B.5}$$

and the r.h.s. of Eqn. (B.4) equals $P_{\Delta S}(>S_{n-1})$. Splitting the integral over δ_n on the l.h.s. of Eqn. (B.4) as $\int_{-\infty}^{\infty}\to\int_{-\infty}^{\delta_c}+\int_{\delta_c}^{\infty}$ and using Eqn. (B.5) we then get

$$\int_{\delta_c}^{\infty} d\delta_n \int_{-\infty}^{\delta_c} d\delta_{n-1}\ldots d\delta_1 W(\delta_0;\{\delta_k\}_n;S_n) = P_{\Delta S}(>S_{n-1}) - P_{\Delta S}(>S_n)\,. \tag{B.6}$$

From Eqn. (B.3) it then follows that

$$\mathcal{F}_{\mathrm{WS,1-bar}}(S) = \lim_{\Delta S\to 0}\frac{1}{\Delta S}\left[P_{\Delta S}(>S_{n-1}) - P_{\Delta S}(>S_n)\right] = -\partial_S P_{\Delta S=0}(>S) = \mathcal{F}_{\mathrm{MR,1-bar}}(S)\,, \tag{B.7}$$

where $P_{\Delta S=0}(>S)$ is the continuum limit cumulative probability for the f.c. time, which is what MR use to compute the f.c. rate. This proves our result.

B.3 Comparison with Lam *et al.* [48]

To compare with the analysis of Lam *et al.* [48] it is useful to express the two barrier conditional f.c. rate in a slightly different form. First, by manipulating the integrals in its definition, the conditional probability density $\Pi_{\Delta S}(\delta_0,\delta_n;S)$ can be rewritten as

$$\begin{aligned}\Pi_{\Delta S}(\delta_0,\delta_n;S_n) &\equiv \int_{-\delta_v}^{\delta_c} d\delta_{n-1}\ldots d\delta_1 W(\delta_0;\{\delta_k\}_n;S_n)\\ &= \int_{-\delta_v}^{\infty} d\delta_{n-1}\ldots d\delta_1 W(\delta_0;\{\delta_k\}_n;S_n)\\ &\qquad - \sum_{k=1}^{n-1}\int_{-\delta_v}^{\infty} d\delta_{n-1}\ldots d\delta_{k+1}\int_{\delta_c}^{\infty} d\delta_k\int_{-\delta_v}^{\delta_c} d\delta_{k-1}\ldots d\delta_1 W(\delta_0;\{\delta_k\}_n;S_n)\,,\end{aligned} \tag{B.8}$$

where it is understood that for $k=1$ the integral in the second term begins with $\int_{\delta_c}^{\infty} d\delta_1$ on the far right, and for $k=n-1$ it ends with $\int_{\delta_c}^{\infty} d\delta_{n-1}$ on the far left. This form of the expression is useful because it helps us to intuitively understand the calculation of the conditional f.c. rate, which becomes

$$\begin{aligned}\mathcal{F}_{\mathrm{NG}}(S) &\equiv \lim_{\Delta S\to 0}\frac{1}{\Delta S}\int_{-\infty}^{-\delta_v} d\delta_n \Pi_{\Delta S}(\delta_0,\delta_n;S)\\ &= \lim_{\Delta S\to 0}\frac{1}{\Delta S}\int_{-\infty}^{-\delta_v} d\delta_n\int_{-\delta_v}^{\infty} d\delta_{n-1}\ldots d\delta_1 W(\delta_0;\{\delta_k\}_n;S)\\ &\qquad - \lim_{\Delta S\to 0}\frac{1}{\Delta S}\sum_{k=1}^{n-1}\int_{-\infty}^{-\delta_v} d\delta_n\int_{-\delta_v}^{\infty} d\delta_{n-1}\ldots d\delta_{k+1}\int_{\delta_c}^{\infty} d\delta_k\int_{-\delta_v}^{\delta_c} d\delta_{k-1}\ldots d\delta_1 W(\delta_0;\{\delta_k\}_n;S)\,.\end{aligned} \tag{B.9}$$

The first term here is simply the single barrier rate across the void barrier (see Appendix B.2). The second term subtracts from this rate the fraction of trajectories which crossed the halo barrier at times before S, without having crossed the void barrier before. This is exactly what Lam *et al.* write in Eqn. 41 of Ref. [48] (in the continuum limit the summation over k above would become an integral over time). Their expression however assumes a factorisation of the second term into a two barrier rate convolved with a conditional probability. This is certainly true for the Gaussian case with the sharp-k filter, as one can immediately see by replacing W above with $W^{\rm gm}$ which factorises, giving the two barrier probability across the *halo* barrier, convolved with the usual Gaussian conditional probability. In the non-Gaussian case however, $W(\delta_0;\{\delta_k\}_n;S)$ does not factorise due to the presence of multi-scale correlations which make the process non-Markovian. In practice though these effects are small, and hence factorisability should be a reasonable assumption. Notice however that our approach does not require this assumption, and we also do not need to linearize the effects of non-Gaussianities as in Ref. [48].

C Details of the calculation of $\mathcal{F}_{\rm SvdW}$

C.1 $\mathcal{I}_n^{\Delta S}(\delta)$ for fixed $\delta < -\delta_{\rm v}$ and small ΔS

With $\delta < -\delta_{\rm v}$ held fixed, a few manipulations allow us to write $\mathcal{I}_n^{\Delta S}(\delta)$ as

$$\begin{aligned}\mathcal{I}_n^{\Delta S}(\delta<-\delta_{\rm v})) &= \frac{(2\Delta S)^{n/2}}{\pi^{1/2}}\left[\int_{-\infty}^{(\delta+\delta_{\rm v})/\sqrt{2\Delta S}} dy\, y^n e^{-y^2} - \int_{-\infty}^{(\delta-\delta_{\rm c})/\sqrt{2\Delta S}} dy\, y^n e^{-y^2}\right] \\ &= \frac{(2\Delta S)^{n/2}}{\pi^{1/2}}(-1)^n\left[\int_{|\delta+\delta_{\rm v}|/\sqrt{2\Delta S}}^{\infty} dq\, q^n e^{-q^2} - \dots\right], \end{aligned} \tag{C.1}$$

where we ignore the second term since one can explicitly check that it leads to an exponentially suppressed contribution which eventually vanishes in the appropriate continuum limit. (This essentially follows from the fact that $\delta_{\rm T} = \delta_{\rm c} + \delta_{\rm v}$ is finite and hence $\delta_{\rm T}/\sqrt{2\Delta S} \to +\infty$ as $\Delta S \to 0$.) Now, since the variable q in the last integral is strictly positive under our assumptions, one can check that the transformation to $x = q^2$ will reduce the integral above to an incomplete Gamma function,

$$\mathcal{I}_n^{\Delta S}(\delta<-\delta_{\rm v}) = \frac{(2\Delta S)^{n/2}}{\pi^{1/2}}(-1)^n\left[\frac{1}{2}\Gamma\left(\frac{n+1}{2},\frac{(\delta+\delta_{\rm v})^2}{2\Delta S}\right)+\dots\right], \tag{C.2}$$

and in the limit of small ΔS, the asymptotic expansion of the incomplete Gamma gives us Eqn. (29).

D Details of the calculation of $\mathcal{F}_{\rm NG}$

D.1 The two barrier derivative exchange property

To prove Eqn. (51), for any function $g(\delta_1,\dots,\delta_n)$ define the function $h(\delta_n;\delta_{\rm c},\delta_{\rm v})$ of three variables as the multiple integral

$$h(\delta_n;\delta_{\rm c},\delta_{\rm v}) = \int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_{n-1}\dots d\delta_1\, g(\delta_1,\dots,\delta_n)\,. \tag{D.1}$$

The l.h.s. of Eqn. (51) can then be reduced as

$$\begin{aligned}\text{L.h.s.} &= \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_{n-1}\dots d\delta_1 \sum_{j=1}^{n}\partial_j g \\ &= h(-\delta_{\rm v};\delta_{\rm c},\delta_{\rm v}) + \sum_{j=1}^{n-1}\int_{-\infty}^{-\delta_{\rm v}} d\delta_n \int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_{n-1}\dots d\delta_1 \partial_j g \\ &= h(-\delta_{\rm v};\delta_{\rm c},\delta_{\rm v}) + \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \left(\partial_{\delta_{\rm c}}|_{\delta_{\rm v}} - \partial_{\delta_{\rm v}}|_{\delta_{\rm c}}\right) h(\delta_n;\delta_{\rm c},\delta_{\rm v})\,, \end{aligned} \tag{D.2}$$

where in the second line we integrated the term involving ∂_{δ_n} and the third line follows since for each $1 \leq j \leq n-1$, the integral becomes

$$\int_{-\infty}^{-\delta_{\rm v}} d\delta_n \int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_{n-1} \dots d\delta_{j+1} d\delta_{j-1} \dots d\delta_1 \Big[g(\delta_1, \dots, \delta_j = \delta_{\rm c}, \dots, \delta_n) - g(\delta_1, \dots, \delta_j = -\delta_{\rm v}, \dots, \delta_n) \Big] , \quad \text{(D.3)}$$

the summation of which is the same as the integral in the third line. The derivative $\partial_{\delta_{\rm c}}|_{\delta_{\rm v}}$ now simply comes across the integral over δ_n since the boundary is independent of $\delta_{\rm c}$. The term involving $\partial_{\delta_{\rm v}}|_{\delta_{\rm c}}$ can be handled by noting that when this derivative acts only on the boundary of the integral $\int_{-\infty}^{-\delta_{\rm v}} d\delta_n \, h(\delta_n; \delta_{\rm c}, \delta_{\rm v})$, we simply get $-h(-\delta_{\rm v}; \delta_{\rm c}, \delta_{\rm v})$. We can then write

$$\partial_{\delta_{\rm v}}|_{\delta_{\rm c}} \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \, h(\delta_n; \delta_{\rm c}, \delta_{\rm v}) = -h(-\delta_{\rm v}; \delta_{\rm c}, \delta_{\rm v}) + \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \, \partial_{\delta_{\rm v}}|_{\delta_{\rm c}} h(\delta_n; \delta_{\rm c}, \delta_{\rm v}) \, . \quad \text{(D.4)}$$

This simplifies the expression (D.2) to

$$\text{L.h.s.} = (\partial_{\delta_{\rm c}}|_{\delta_{\rm v}} - \partial_{\delta_{\rm v}}|_{\delta_{\rm c}}) \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \, h(\delta_n; \delta_{\rm c}, \delta_{\rm v}) \, . \quad \text{(D.5)}$$

It is also straightforward to show that under the change of variables $(\delta_{\rm c}, \delta_{\rm v}) \to (\delta_{\rm T}, \delta_{\rm v})$ where $\delta_{\rm T} = \delta_{\rm c} + \delta_{\rm v}$, we have $\partial_{\delta_{\rm c}}|_{\delta_{\rm v}} - \partial_{\delta_{\rm v}}|_{\delta_{\rm c}} = -\partial_{\delta_{\rm v}}|_{\delta_{\rm T}}$, which gives us

$$\text{L.h.s.} = -\left. \frac{\partial}{\partial \delta_{\rm v}} \right|_{\delta_{\rm T}} \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_{n-1} \dots d\delta_1 \, g(\delta_1, \dots, \delta_n) = \text{R.h.s.} \, , \quad \text{(D.6)}$$

which proves the result.

D.2 Leading unequal time contribution

Here we sketch a proof of Eqn. (54). Defining the quantity M_j as

$$M_j \equiv \sum_{k,l=1}^{n} \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_{n-1} \dots d\delta_1 \partial_k \partial_l \left(\partial_j W^{\rm gm} \right) \, , \quad \text{(D.7)}$$

we can write

$$\mathcal{F}^{(3,\rm NL)} = \lim_{\Delta S \to 0} \frac{1}{2\Delta S} \mathcal{G}_3^{(1,0,0)}(S) \sum_j (S - S_j) M_j \, . \quad \text{(D.8)}$$

Using the derivative exchange property (51) we can write M_j as

$$\begin{aligned} M_j &= \partial_{\delta_{\rm v}}^2 \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \int_{-\delta_{\rm v}}^{\delta_{\rm c}} d\delta_{n-1} \dots d\delta_1 \, \partial_j W^{\rm gm} \\ &= -\partial_{\delta_{\rm v}}^2 \int_{-\infty}^{-\delta_{\rm v}} d\delta_n \left[\Pi^{\rm gm}_{\Delta S}(\delta_0, -\delta_{\rm v}; S_j) \Pi^{\rm gm}_{\Delta S}(-\delta_{\rm v}, \delta_n; S - S_j) - \Pi^{\rm gm}_{\Delta S}(\delta_0, \delta_{\rm c}; S_j) \Pi^{\rm gm}_{\Delta S}(\delta_{\rm c}, \delta_n; S - S_j) \right] . \end{aligned} \quad \text{(D.9)}$$

Recall that $\partial_{\delta_{\rm v}} = \partial / \partial \delta_{\rm v}|_{\delta_{\rm T}}$. In the second line, we already have an expression for $\Pi^{\rm gm}_{\Delta S}(\delta_0, -\delta_{\rm v}; S_j)$ from Eqn. (38). One can also check that in the continuum limit, the second term involving $\delta_{\rm c}$ will be exponentially suppressed. The only non-trivial quantity we need then is the integral $\int_{-\infty}^{-\delta_{\rm v}} d\delta_n \Pi^{\rm gm}_{\Delta S}(-\delta_{\rm v}, \delta_n; S - S_j)$. Following techniques similar to those discussed in Section 3.2 leading up to Eqn. (34), we can write this integral by simply replacing δ_0 in (34) with $(-\delta_{\rm v})$ to get

$$\int_{-\infty}^{-\delta_{\rm v}} d\delta_n \Pi^{\rm gm}_{\Delta S}(-\delta_{\rm v}, \delta_n; S - S_j) = \left(\frac{\Delta S}{2\pi} \right)^{1/2} \left[\int_{-\infty}^{0} \frac{d\eta}{(-\eta)} v(\eta) e^{-\eta^2} \right] \Pi^{\rm gm}_{\Delta S}(-\delta_{\rm v}, -\delta_{\rm v}; S - S_j) \, , \quad \text{(D.10)}$$

with the same boundary layer function $v(\eta)$ appearing, since this is independent of the value of δ_0. Notice that the integral involving $v(\eta)$ is identical to that appearing in the expression (39) for $\mathcal{F}_{\rm SvdW}$. In fact the

other constant γ in that expression also appears in M_j through the object $\Pi^{\rm gm}_{\Delta S}(\delta_0, -\delta_{\rm v}; S_j)$, and will lead us to identify a factor of $\mathcal{F}_{\rm SvdW}$ in the final expression, see below. The object $\Pi^{\rm gm}_{\Delta S}(-\delta_{\rm v}, -\delta_{\rm v}; S - S_j)$ is similar to one discussed by MR in Ref. [52], where they showed that $\Pi^{\rm gm}_{1-bar,\Delta S}(\delta_{\rm c}, \delta_{\rm c}; \tilde{S}) = \Delta S/(2\pi\tilde{S}^3)^{1/2}$, independent of $\delta_{\rm c}$. In fact, one can show that $\Pi^{\rm gm}_{\Delta S}(-\delta_{\rm v}, -\delta_{\rm v}; \tilde{S})$ is also independent of *both* $\delta_{\rm c}$ and $\delta_{\rm v}$, and has the same value

$$\Pi^{\rm gm}_{\Delta S}(-\delta_{\rm v}, -\delta_{\rm v}; \tilde{S}) = \frac{\Delta S}{\sqrt{2\pi}} \frac{1}{\tilde{S}^{3/2}} \,. \tag{D.11}$$

This can be checked in two steps. Firstly for arbitrary n one can show that the path integral with n steps involved in $\Pi^{\rm gm}_{\Delta S}(-\delta_{\rm v}, -\delta_{\rm v}; S_n)$ becomes independent of $\delta_{\rm c}$ and $\delta_{\rm v}$ in the continuum limit (since $\delta_{\rm c}$ and $\delta_{\rm v}$ only appear in the limits of integration, and a suitable change of variables sends these to infinity as $\Delta S \to 0$). The remaining integral is therefore only a function of n. Following MR, a dimensional argument coupled with a straightforward calculation for $n = 2$ then fixes the time dependence completely, leading to Eqn. (D.11). Putting things together,

$$\begin{aligned} M_j &= -\partial^2_{\delta_{\rm v}} \left[\Pi^{\rm gm}_{\Delta S}(\delta_0, -\delta_{\rm v}; S_j) \left(\left(\frac{\Delta S}{2\pi} \right)^{1/2} \left[\int_{-\infty}^{0} \frac{d\eta}{(-\eta)} v(\eta) e^{-\eta^2} \right] \right) \left(\frac{\Delta S}{\sqrt{2\pi}} \frac{1}{(S - S_j)^{3/2}} \right) \right] \\ &= -\frac{(\Delta S)^2}{\sqrt{2\pi}} \partial^2_{\delta_{\rm v}} \frac{\mathcal{F}_{\rm SvdW}(S_j)}{(S - S_j)^{3/2}} \,, \end{aligned} \tag{D.12}$$

where we used Eqns. (38) and Eqn. (39) in writing the last line. Using this result for M_j in Eqn. (D.8) with the continuum limit $\sum_{j=1}^{n} h(S_j) \to \int_0^S d\tilde{S} h(\tilde{S}) (\Delta S)^{-1}$, we see that the factors of ΔS cancel, and further expressing $\mathcal{G}_3^{(1,0,0)}(S)$ in terms of $c_1(S)$ finally leads to the result of Eqn. (54).

References

[1] R. P. Kirshner, A. . J. Oemler, P. L. Schechter and S. A. Shectman, "A million cubic megaparsec void in Bootes," Astrophys. J. **248**, L57 (1981).

[2] S. G. Patiri, J. Betancort-Rijo, F. Prada, A. Klypin and S. Gottlober, "Statistics of Voids in the 2dF Galaxy Redshift Survey," Mon. Not. Roy. Astron. Soc. **369**, 335 (2006) [arXiv:astro-ph/0506668].

[3] D. M. Goldberg, T. D. Jones, F. Hoyle, R. R. Rojas, M. S. Vogeley and M. R. Blanton, "The Mass Function of Void Galaxies in the SDSS Data Release 2," Astrophys. J. **621**, 643 (2005) [arXiv:astro-ph/0406527].

[4] D. J. Croton *et al.* [The 2dFGRS Team Collaboration], "The 2dF Galaxy Redshift Survey: Voids and hierarchical scaling models," Mon. Not. Roy. Astron. Soc. **352**, 828 (2004) [arXiv:astro-ph/0401406].

[5] F. Hoyle and M. S. Vogeley, "Voids in the PSCz Survey and the Updated Zwicky Catalog," Astrophys. J. **566**, 641 (2002) [arXiv:astro-ph/0109357].

[6] F. Hoyle, M. S. Vogeley and R. R. Rojas, "Void Galaxies in the Sloan Digital SKy Survey," Bulletin of the AAS **37**, 443 (2005)

[7] J. E. Forero-Romero, Y. Hoffman, S. Gottloeber, A. Klypin and G. Yepes, "A Dynamical Classification of the Cosmic Web," Mon. Not. Roy. Astron. Soc. **396**, 1815 (2009) [arXiv:0809.4135 [astro-ph]].

[8] J. Lee and B. Lee, "The Variation of Galaxy Morphological Type with the Shear of Environment," Astrophys. J. **688**, 78 (2008) [arXiv:0801.1558 [astro-ph]].

[9] M. A. Aragon-Calvo, B. J. T. Jones, R. van de Weygaert and M. J. van der Hulst, "The Multiscale Morphology Filter: Identifying and Extracting Spatial Patterns in the Galaxy Distribution," Astron. Astrophys. **474**, 315 (2004) [arXiv:0705.2072 [astro-ph]].

[10] S. Shandarin, S. Habib and K. Heitmann, "Origin of the Cosmic Network: Nature vs Nurture," Phys. Rev. D **81**, 103006 (2010) [arXiv:0912.4471 [astro-ph.CO]].

[11] R. van de Weygaert, M. A. Aragon-Calvo, B. J. T. Jones and E. Platen, "Geometry and Morphology of the Cosmic Web: Analyzing Spatial Patterns in the Universe," arXiv:0912.3448 [astro-ph.IM].

[12] J. A. Peacock *et al.*, "A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey," Nature **410**, 169 (2001) [arXiv:astro-ph/0103143].

[13] D. J. Eisenstein *et al.* [SDSS Collaboration], "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies," Astrophys. J. **633**, 560 (2005) [arXiv:astro-ph/0501171].

[14] K. N. Abazajian *et al.* [SDSS Collaboration], "The Seventh Data Release of the Sloan Digital Sky Survey," Astrophys. J. Suppl. **182**, 543 (2009) [arXiv:0812.0649 [astro-ph]].

[15] A. Cimatti *et al.*, "Euclid Assessment Study Report for the ESA Cosmic Visions," arXiv:0912.0914 [astro-ph.CO].

[16] N. Cappelluti *et al.*, "eROSITA on SRG: a X-ray all-sky survey mission," arXiv:1004.5219 [astro-ph.IM].

[17] D. Schlegel, M. White and D. Eisenstein [with input from the SDSS-III collaboration and with input from the SDSS-III], "The Baryon Oscillation Spectroscopic Survey: Precision measurements of the absolute cosmic distance scale," arXiv:0902.4680 [astro-ph.CO].

[18] J. M. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," JHEP **0305**, 013 (2003) [arXiv:astro-ph/0210603].

[19] V. Acquaviva, N. Bartolo, S. Matarrese and A. Riotto, "Second-order cosmological perturbations from inflation," Nucl. Phys. B **667**, 119 (2003) [arXiv:astro-ph/0209156].

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

[21] 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]].

[22] 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].

[23] C. Cunha, D. Huterer and O. Dore, "Primordial non-Gaussianity from the covariance of galaxy cluster counts," Phys. Rev. D **82**, 023004 (2010) [arXiv:1003.2416 [astro-ph.CO]].

[24] 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]].

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

[26] 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]].

[27] 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].

[28] 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]].

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

[30] V. Desjacques and U. Seljak, "Primordial non-Gaussianity from the large scale structure," Class. Quant. Grav. **27**, 124011 (2010) [arXiv:1003.5020 [astro-ph.CO]].

[31] J. M. Colberg *et al.*, "The Aspen–Amsterdam Void Finder Comparison Project," Mon. Not. Roy. Astron. Soc. **387**, 933 (2008) [arXiv:0803.0918 [astro-ph]].

[32] R. van de Weygaert and E. Platen, "Cosmic Voids: structure, dynamics and galaxies," arXiv:0912.2997 [astro-ph.CO].

[33] J. Dubinski, L. Nicolaci da Costa, D. S. Goldwirth, M. Lecar and T. Piran, "Void evolution and the large scale structure," Astrophys. J. **410**, 458 (1993).

[34] J. M. Colberg, R. K. Sheth, A. Diaferio, L. Gao and N. Yoshida, Mon. Not. Roy. Astron. Soc. **360**, 216 (2005) [arXiv:astro-ph/0409162].

[35] R. van de Weygaert and E. van Kampen, "Voids in Gravitational Instability Scenarios - Part One - Global Density and Velocity Fields in an Einstein - De-Sitter Universe," Mon. Not. Roy. Astron. Soc. **263**, 481 (1993).

[36] G. R. Blumenthal, L. Nicolaci da Costa, D. S. Goldwirth, M. Lecar and T. Piran, "The largest possible voids," Astrophys. J. **388**, 234 (1992).

[37] 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].

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

[39] R. I. Epstein, "Proto-galactic perturbations," Mon. Not. Roy. Astron. Soc. **205**, 207 (1983).

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

[41] 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).

[42] J. A. Fillmore and P. Goldreich, "Self-similar spherical voids in an expanding universe," Astrophys. J. **281**, 9 (1984).

[43] E. Bertschinger, "The self-similar evolution of holes in an Einstein-de Sitter universe," Astrophys. J. Suppl. **58**, 1 (1985).

[44] G. Lavaux and B. D. Wandelt, "Precision cosmology with voids: definition, methods, dynamics," Mon. Not. Roy. Astron. Soc. **403**, 1392 (2010) [arXiv:0906.4101 [astro-ph.CO]].

[45] R. 1. Biswas, E. Alizadeh and B. D. Wandelt, "Voids as a Precision Probe of Dark Energy," Phys. Rev. D **82**, 023002 (2010) [arXiv:1002.0014 [astro-ph.CO]].

[46] S. Furlanetto and T. Piran, "The Evidence of Absence: Galaxy Voids in the Excursion Set Formalism," Mon. Not. Roy. Astron. Soc. **366**, 467 (2006) [arXiv:astro-ph/0509148].

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

[48] T. Y. Lam, R. K. Sheth and V. Desjacques, "The initial shear field in models with primordial local non-Gaussianity and implications for halo and void abundances," arXiv:0905.1706 [astro-ph.CO].

[49] S. Chongchitnan and J. Silk, "A Study of High-Order Non-Gaussianity with Applications to Massive Clusters and Large Voids," Astrophys. J. **724**, 285 (2010) [arXiv:1007.1230 [astro-ph.CO]].

[50] M. Maggiore and A. Riotto, "The halo mass function from the excursion set method. III. First principle derivation for non-Gaussian theories," Astrophys. J. **717**, 526 (2010) [arXiv:0903.1251 [astro-ph.CO]].

[51] G. D'Amico, M. Musso, J. Noreña and A. Paranjape, "An Improved Calculation of the Non-Gaussian Halo Mass Function," arXiv:1005.1203 [astro-ph.CO].

[52] 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," Astrophys. J. **711**, 907 (2010) [arXiv:0903.1249 [astro-ph.CO]].

[53] M. Maggiore and A. Riotto, "The halo mass function from the excursion set method. II. The diffusing barrier," Astrophys. J. **717**, 515 (2010) [arXiv:0903.1250 [astro-ph.CO]].

[54] G. L. Hoffman, E. E. Salpeter and I. Wasserman, "Spherical simulations of holes and honeycombs in Friedmann universes," Astrophys. J. **268**, 527 (1983).

[55] S. Redner, "A guide to first-passage processes," Cambridge, UK : Cambridge Univ. Press (2001) 312pp.

[56] G. D'Amico *et al.*, in progress.

[57] 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/9907024].

[58] 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].

[59] A. De Simone, M. Maggiore and A. Riotto, "Excursion Set Theory for generic moving barriers and non-Gaussian initial conditions," arXiv:1007.1903 [astro-ph.CO].

[60] T. Y. Lam and R. K. Sheth, "Halo abundances in the $f_{\rm NL}$model," arXiv:0905.1702 [astro-ph.CO].

[61] D. Babich, P. Creminelli and M. Zaldarriaga, "The shape of non-Gaussianities," JCAP **0408** (2004) 009 [arXiv:astro-ph/0405356].

[62] D. 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].

[63] N. Bartolo, S. Matarrese and A. Riotto, "On non-Gaussianity in the curvaton scenario," Phys. Rev. D **69** (2004) 043503 [arXiv:hep-ph/0309033].

[64] 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].

[65] M. Alishahiha, E. Silverstein and D. Tong, "DBI in the sky," Phys. Rev. D **70**, 123505 (2004) [arXiv:hep-th/0404084].

[66] N. Arkani-Hamed, P. Creminelli, S. Mukohyama and M. Zaldarriaga, "Ghost Inflation," JCAP **0404**, 001 (2004) [arXiv:hep-th/0312100].

[67] P. Creminelli, "On non-Gaussianities in single-field inflation," JCAP **0310**, 003 (2003) [arXiv:astro-ph/0306122].

[68] 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].

[69] 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).

[70] I. S. Gradshteyn and I. M. Ryzhik, "Tables of Integrals, Series and Products", 7th ed., Amsterdam : Elsevier (2007) 1171pp.