Imprints of primordial non-Gaussianities on large-scale structure: Scale-dependent bias and abundance of virialized objects

Neal Dalal,[1] Olivier Doré,[1] Dragan Huterer,[2,3] and Alexander Shirokov[1]

[1]*Canadian Institute for Theoretical Astrophysics, 60 St. George Street, University of Toronto, Toronto, ON, Canada M5S3H8*

[2]*Kavli Institute for Cosmological Physics and Department of Astronomy and Astrophysics, University of Chicago, Chicago, Illinois 60637, USA*

[3]*Department of Physics, University of Michigan, 450 Church Street, Ann Arbor, Michigan 48109, USA*

(Received 29 October 2007; published 11 June 2008)

We study the effect of primordial non-Gaussianity on large-scale structure, focusing upon the most massive virialized objects. Using analytic arguments and *N*-body simulations, we calculate the mass function and clustering of dark matter halos across a range of redshifts and levels of non-Gaussianity. We propose a simple fitting function for the mass function valid across the entire range of our simulations. We find pronounced effects of non-Gaussianity on the clustering of dark matter halos, leading to strongly scale-dependent bias. This suggests that the large-scale clustering of rare objects may provide a sensitive probe of primordial non-Gaussianity. We very roughly estimate that upcoming surveys can constrain non-Gaussianity at the level of $|f_{\rm NL}| \lesssim 10$, which is competitive with forecasted constraints from the microwave background.

DOI: 10.1103/PhysRevD.77.123514 PACS numbers: 98.65.Dx, 95.35.+d, 98.80.Cq

I. INTRODUCTION

One of the fundamental predictions of standard (single-field, slow-roll) inflationary cosmology is that the density fluctuations in the early universe that seeded large-scale structure formation were nearly Gaussian random (e.g. [1–5]). Constraining or detecting non-Gaussianity (NG) is therefore an important and basic test of the cosmological model. To the extent that it can be measured, Gaussianity has so far been confirmed; the tightest existing constraints have been obtained from observations of the cosmic microwave background [6,7]. Recently, several inflationary models have been proposed which predict a potentially observable level of non-Gaussianity; see e.g. [8–24] for a review. Improved limits on NG would rule out some of these models; conversely, a robust detection of primordial non-Gaussianity would dramatically overturn standard inflationary cosmology and provide invaluable information about the nature of physical processes in the early universe. In this regard, there has been a resurgence in studying increasingly more sophisticated methods and algorithms to constrain (or, if we are lucky, detect) non-Gaussianity [25–29].

Non-Gaussianity manifests itself not only in the cosmic microwave background [30–33], but also in the late-time evolution of large-scale structure. For example, detailed measurements of higher order correlations like the bispectrum or trispectrum of galaxy clustering could provide a handle on primordial non-Gaussianity [34–36]. The abundance of galaxy clusters, the largest virialized objects in the universe, has also long been recognized as a sensitive probe of primordial NG [35,37–42]. Because clusters are rare objects which form from the largest fluctuations on the tails of the density probability distribution, their abundance is keenly sensitive to changes in the shape of the PDF, such as those caused by non-Gaussianity. Large statistical samples of massive clusters have already been compiled from wide-area optical imaging and spectroscopic surveys such as the Sloan Digital Sky Survey (SDSS) [43,44], the Two-Degree Survey [45], as well as from the Red Sequence Survey [46] and x-ray surveys using the Chandra and XMM-Newton observatories [47,48]. Future missions, such as the Dark Energy Survey, the Supernova/Acceleration Probe, and the Large Synoptic Survey Telescope, will detect and study tens of thousands of clusters, revolutionizing our understanding of cluster physics as well as providing important constraints on cosmology [49–55].

To exploit the potential of these upcoming surveys as probes of primordial non-Gaussianity, it is important to calibrate the effects of NG on the abundance and clustering of virialized objects. While no previous work has attempted to quantify the effects of NG on halo clustering, several groups over the past decade have constructed fitting formulas for the halo mass function [56–58]. All of this work, however, was analytic and relied on the validity of the Press-Schechter [59] formalism, plus various further approximations. The resulting analytic estimates are, in general, rather cumbersome to compute and have questionable accuracy. As discussed below, the Press-Schechter model provides only a qualitative description of halo abundance, and fails to reproduce the halo mass function to within an order of magnitude over the mass and redshift range accessible to current and future cluster surveys. Therefore, analytic models for NG cluster abundance based on the Press-Schechter ansatz may not be sufficiently accurate. Given the high-quality data soon to be available, a much more precise calculation of cluster statistics will be

required. Quite recently, two groups have attempted to quantify the mass function of clusters in NG models using N-body simulations [60,61], reaching contradictory conclusions.

In this paper, we use analytic arguments and numerical simulations to estimate the effect of NG on the abundance and clustering of virialized objects. Because N-body simulations can be expensive and there is a wide NG parameter space, we also strive to make our results useful to a cosmologist who is not necessarily equipped with the machinery or patience to run simulations or evaluate difficult analytic expressions. To this end, we provide a simple, physically motivated fitting formula for the halo mass function and halo bias, which we calibrate to our N-body simulations.

Our main results are that the mass function and correlation function of massive halos can be significantly modified by primordial non-Gaussianity. We find a somewhat weaker effect of NG on the mass function than previous analytic estimates. We also show analytically and numerically that NG strongly affects the clustering of rare objects on large scales, implying that measurements of the large-scale power spectrum can place stringent bounds on NG.

The plan of the paper is as follows. In Sec. II we derive analytic expressions for the abundance and clustering of rare peaks. In Sec. III we describe our N-body simulations, followed in Sec. IV by a discussion of our measured halo mass function, and our fitting formula for the mass function. In Sec. V we present measurements of halo clustering within our simulations, and in Sec. VI we discuss cosmological implications of our findings.

II. ANALYTIC ESTIMATES

In this section, we derive analytic expressions for the abundance and clustering of dark matter halos. As mentioned above, such analytic approaches provide a useful qualitative framework for understanding gravitational collapse; however, they cannot be used to describe quantitatively either the mass function or the clustering amplitude of collapsed objects. The expressions derived here are meant solely to motivate the more precise fitting formulas described in subsequent sections.

We will focus on local NG of the form [1,58,62]

$$\Phi_{\rm NG}(\boldsymbol{x}) = \phi(\boldsymbol{x}) + f_{\rm NL}(\phi^2(\boldsymbol{x}) - \langle\phi^2\rangle). \tag{1}$$

In our notation, $\Phi = -\Psi$, where Ψ is the usual Newtonian potential. On subhorizon scales, this choice of Newtonian gauge is valid, and the potentials Φ and Ψ satisfy the Poisson equation relating them to the overdensity δ. On superhorizon scales, the Bardeen potential Φ and overdensity δ are proportional, and not related by a Poisson equation, so our analysis will be valid only on subhorizon scales. With this choice of convention, positive $f_{\rm NL}$ corresponds to positive skewness of the density probability distribution, and hence an increased number of massive objects.

For simplicity, we neglect the effect of the cold dark matter transfer functions, which modify the shape of the Φ power spectrum after non-Gaussianity is generated. Then the probability distribution for $\Phi_{\rm NG}$ is easy to write down; however, the probability distribution for the density $\delta_{\rm NG}$ cannot be expressed analytically. Nevertheless, we can make progress by assuming that the NG correction is small, and by focusing only on high peaks of the density. The Laplacian of $\Phi_{\rm NG}$ is

$$\nabla^2\Phi_{\rm NG} = \nabla^2\phi + 2f_{\rm NL}[\phi\nabla^2\phi + |\nabla\phi|^2]. \tag{2}$$

Because ϕ, $\nabla\phi$, and $\nabla^2\phi$ are all Gaussian fields whose statistics are fully specified by their power spectra, then Eq. (2) above, relating $\delta_{\rm NG} = -(3\Omega_m/2ar_H^2)\nabla^2\Phi_{\rm NG}$ to the Gaussian fields, allows us to determine fully the statistics of the non-Gaussian density $\delta_{\rm NG}$. For example, the skewness of $\delta_{\rm NG}$ becomes, to lowest order in $f_{\rm NL}$,

$$S_3 = \frac{\langle\delta_{\rm NG}^3\rangle}{\langle\delta_{\rm NG}^2\rangle^2} = 6f_{\rm NL}\frac{\langle\phi\,\delta\rangle}{\sigma_\delta^2}. \tag{3}$$

On the average, the two terms $\phi\nabla^2\phi$ and $|\nabla\phi|^2$ in Eq. (2) are of the same order; the fact that they have equal but opposite expectation values is why $\langle\delta_{\rm NG}\rangle = \langle\delta\rangle = 0$. However, we are mainly interested in high peaks, where $\delta \propto -\nabla^2\phi$ is large. Because $|\nabla\phi|^2$ is uncorrelated with $\nabla^2\phi$, and because at the peak of ϕ its derivative vanishes, we assume that $|\nabla\phi|^2$ may be neglected compared to $\phi\nabla^2\phi$ in the vicinity of rare, high peaks. Then applying the Poisson equation near the peak gives $\delta_{\rm NG} \approx \delta[1 + 2f_{\rm NL}\phi]$. This expression applies for the primordial density and potential fields at early times. At late times, $\delta_{\rm NG}$ subsequently grows according to the linear growth factor $D(a)$, while the potential decays like $g(a) \propto D(a)/a$. Therefore, rewriting this expression in terms of the late-time fields, we find

$$\delta_{\rm NG} \approx \delta[1 + 2f_{\rm NL}\phi/g(a)]. \tag{4}$$

We see that the peak height is enhanced by a factor proportional to the primordial potential $\phi_p = \phi/g(a)$, rather than the evolved potential.[1]

Equation (4) will be the basis for the rest of our discussion. We emphasize that this is only valid in the vicinity of peaks, and so we focus on peaks for the remainder of this discussion. Because the fields δ and ϕ are Gaussian distributed, we can immediately derive properties of the distribution of $\delta_{\rm NG}$. For example, consider the mean shift in peak height for a peak of Gaussian density δ:

[1]An earlier version of this paper neglected to distinguish between the primordial and late-time potential, and hence omitted the $g(a)$ factor. We are grateful to N. Afshordi for pointing this out to us.

$$\langle \delta_{\rm NG}|\delta\rangle = \delta(1 + 2f_{\rm NL}\langle\phi_p|\delta\rangle) = \delta\left(1 + 2f_{\rm NL}\frac{\langle\phi\delta\rangle}{g\sigma_\delta^2}\delta\right). \tag{5}$$

If the peak height δ and background potential ϕ were uncorrelated, then there would be no systematic shift in peak height, and hence no change in the abundance of massive halos. However, δ and ϕ are correlated, implying that rare peaks are systematically raised or lowered, depending upon the sign of $f_{\rm NL}$. Therefore, we expect changes in the mass function and the correlation function.

In the Appendix, we derive expressions for the abundance and clustering of regions above a given threshold, which then give the clustering and mass function of halos in the Press-Schechter model. However, we can derive the form of the halo correlation function using a very simple argument. The halo correlation function is usually parametrized in terms of the halo bias b, which is the rate of change of the halo abundance as the background density is varied. Writing the matter overdensity as δ and the halo overdensity as δ_h, we can define the halo bias as

$$\delta_h = b\delta. \tag{6}$$

It is normally assumed that $b \to$ const on large scales, but we will not make this assumption here. Consider a long-wavelength mode, providing a background density perturbation δ and corresponding potential fluctuation ϕ. In the absence on non-Gaussianity, this perturbation raises subthreshold peaks above threshold, and thereby enhances the abundance of superthreshold peaks by $b_L\delta$, where b_L is the usual (Gaussian) Lagrangian bias. For nonzero $f_{\rm NL}$, the long-wavelength mode also enhances the peak height by $2f_{\rm NL}\phi_p\delta_{\rm pk}$, and we will focus on peaks near threshold, such that $\delta_{\rm pk} \simeq \delta_c$. This provides an additional enhancement factor, giving a total of

$$\delta_h = b_L(\delta + 2f_{\rm NL}\phi_p\delta_c). \tag{7}$$

In Fourier space, the potential and density modes are related by $\phi = (3\Omega_m/2ar_H^2k^2)\delta$, and so we see that the non-Gaussian bias acquires a correction

$$\Delta b(k) = 2b_L f_{\rm NL}\delta_c\frac{3\Omega_m}{2ag(a)r_H^2k^2}, \tag{8}$$

where again b_L refers to the usual Lagrangian bias for halos of this mass with Gaussian fluctuations. The total Lagrangian bias is then $b_L(k) = b_L + \Delta b(k)$.

Since we have been working with the clustering of peaks in the initial density distribution, the above expression for the bias applies only to the early-time, Lagrangian bias. Translating these results to late-time, Eulerian bias is straightforward, however. The bias of Eulerian halos is simply $b = 1 + b_L$: the excess of halos in some Eulerian volume with overdensity δ is $b\delta = b_L\delta + \delta$. The first term corresponds to the excess of peaks in the initial Lagrangian volume, which are advected into the Eulerian volume. The second term arises because a Eulerian volume with overdensity δ has δ times more mass than an average volume, and therefore δ times more peaks.

In summary, local NG generates a scale-dependent correction to the bias of galaxies and halos, of the form

$$\Delta b(k) = 2(b-1)f_{\rm NL}\delta_c\frac{3\Omega_m}{2ag(a)r_H^2k^2} \tag{9}$$

where b here now refers to the Eulerian bias of the tracer population. In subsequent sections, we show that this simple expression, despite the underlying assumptions and approximations in its derivation, matches surprisingly well the halo clustering measured in our numerical simulations.

III. NUMERICAL SIMULATIONS

We numerically simulate the growth of structure in non-Gaussian cosmologies using the adaptive P^{3}M parallel N-body code GRACOS[2] [63,64]. Non-Gaussian initial conditions were generated using the following procedure. First, we generated a Gaussian random potential field $\phi(\mathbf{x})$ using a power-law power spectrum with a scalar (density) index $n_s = 0.96$, and normalized so that $\sigma_8 = 0.76$ [6] when multiplied by the matter transfer function. Following Refs. [1,58,62], we then computed the non-Gaussian potential Φ by adding a quadratic correction in configuration space,

$$\Phi(\boldsymbol{x}) = \phi(\boldsymbol{x}) + f_{\rm NL}(\phi^2 - \langle\phi^2\rangle). \tag{10}$$

We then multiplied Φ by matter transfer functions in Fourier space for $\Omega_m = 0.24$, $\Omega_\Lambda = 0.76$, and computed particle displacements and velocities using the Zeldovich approximation [65].

One immediate drawback to this approach is that, due to the strong Fourier mode coupling generated by the $f_{\rm NL}$ term, our results may be affected by the absence of modes below the fundamental frequency or above the Nyquist frequency of our simulation volume. All N-body simulations can cover only a finite dynamic range, and therefore have zero power outside of their k-space volumes. For Gaussian simulations, this is believed not to be a serious defect, because mode coupling is unimportant on linear scales, and on nonlinear scales the mode coupling generally transfers power to small scales. In our case, however, the $f_{\rm NL}$ term couples all the modes sampled in our simulation to all the modes absent in our simulation. We have performed rudimentary estimates of the magnitude of this effect, by running simulations in which we high-pass or low-pass filter the $f_{\rm NL}$ correction, and do not observe significant changes in the overall behavior. Strictly speaking, however, it must be borne in mind that our results

[2]http://www.gracos.org

apply only for power spectra that are nonvanishing only over the finite range covered by our simulation volume.

We have performed several simulations using both Gaussian and non-Gaussian initial conditions. For each Gaussian realization, we construct non-Gaussian realizations using the same Fourier phases, with various $f_{\rm NL}$, e.g. $f_{\rm NL} = \pm 500$, ± 50, and ± 5. We ran simulations from a starting expansion factor $a = 0.02$ until the present time, $a = 1$, using 512^3 particles in a box of side length $L = 800h^{-1}$ Mpc. For these parameters, each particle has a mass $m_p = 2.52 \times 10^{11} h^{-1} M_\odot$, so that clusters with masses exceeding $M > 10^{14} h^{-1} M_\odot$ are resolved with $N \gtrsim 400$ particles. Since we are interested mainly in the masses and positions of cluster-sized halos, and not their internal structure, we have not used high force resolution: we employ a Plummer softening length l of 0.2 times the mean interparticle spacing. We have checked that using higher force resolution (l half as large) does not appreciably change the mass function. All simulations were performed at the Sunnyvale cluster at CITA; depending upon the value of $f_{\rm NL}$, the simulations completed in 2–3 hours each on typically 8–10 nodes. As a consistency check, we have also run a small number of 1024^3 particle simulations with the same particle mass and force softening as above, but with twice the box size. These larger runs typically completed in 18–20 hours on 64 nodes. In Fig. 1, we plot slices through our simulation volume at redshift $z = 0$, and the effects of varying $f_{\rm NL}$ are readily apparent. Large positive $f_{\rm NL}$ accelerates the evolution of overdense regions and retards the evolution of underdense regions, while large negative $f_{\rm NL}$ has precisely the opposite effect.

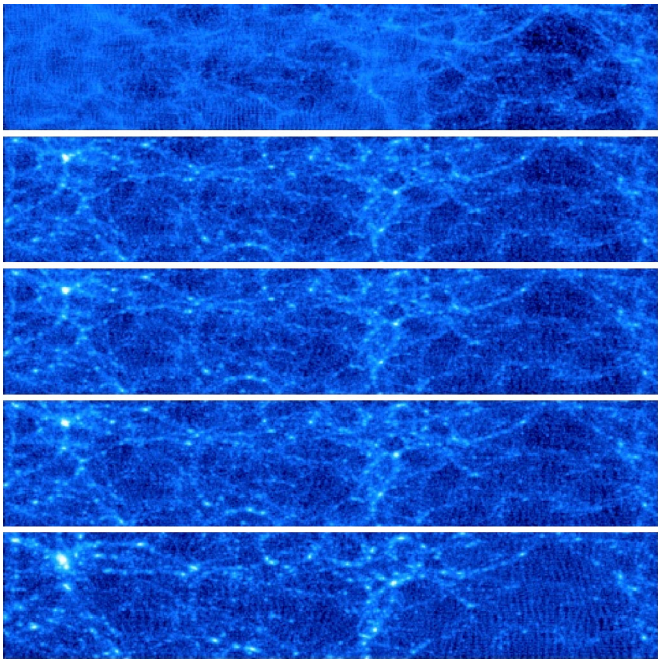

FIG. 1 (color online). Slice through simulation outputs at $z = 0$ generated with the same Fourier phases but with $f_{\rm NL} = -5000$, -500, 0, $+500$, $+5000$, respectively, from top to bottom. Each slice is $375h^{-1}$ Mpc wide, and $80h^{-1}$ Mpc high and deep. We can easily match by eye much of the large-scale structure; for example, an overdense region sits on the left, while an underdense region (void) falls on the right, in all panels. Note that for positive $f_{\rm NL}$, overdense regions are more evolved and produce more clusters than their Gaussian counterparts, while underdense regions are less evolved (e.g. grid lines are still visible). For negative $f_{\rm NL}$, underdense regions are more evolved, producing deeper voids, while overdense regions are less evolved, as illustrated by the grid lines apparent in the left of the top panel.

IV. THE HALO MASS FUNCTION

We constructed late-time halo catalogues at redshifts $z = 1$, 0.5, and 0 using the friends-of-friends group finder [66], with linking length $b = 0.2$. For Gaussian simulations, the halo mass function constructed this way has been extensively calibrated [67,68]. Resulting mass functions are plotted in Fig. 2.

A. A new fitting formula

Having measured the halo mass function, we next would like to construct a fitting function along the lines of those used for Gaussian simulations [67,68]. As mentioned above, previous techniques for estimating the non-Gaussian mass function have been based upon the Press-Schechter [59] ansatz. Given that the Press-Schechter mass function fails to match the halo mass function to within an order of magnitude over the mass and redshift ranges of interest to us [67], and given the lack of any physical basis to the Press-Schechter ansatz [69,70], we have instead adopted an alternative approach which we describe next.

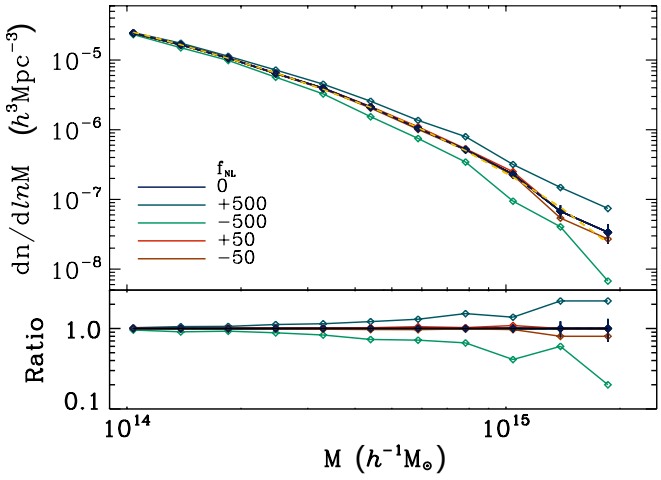

FIG. 2 (color online). Mass functions measured from simulations with various $f_{\rm NL}$ and identical phases (3 sets of initial conditions were used for each $f_{\rm NL}$). The top panel shows the mass function as well as the Gaussian fitting formula (dashed yellow line) from Warren *et al.* [67]. The bottom panel shows the ratio between the measured $f_{\rm NL} = 0$ Gaussian mass functions and the respective non-Gaussian ones.

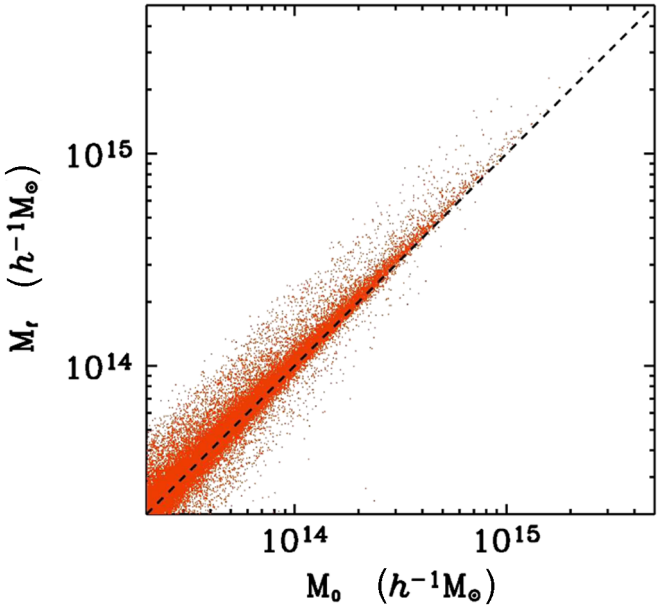

FIG. 3 (color online). Distribution of M_f as a function of M_0 for one $f_{\rm NL} = +500$ simulation. The average shift towards higher masses is clearly visible.

We start by noting that the halo mass function dn/dM has been precisely calibrated for Gaussian cosmologies. Consider a Gaussian realization of the density field, which at late times evolves to produce halos with mass function dn/dM_0. As we slowly vary $f_{\rm NL}$ away from zero, the structures forming at late times also slowly vary (cf. Fig. 1), producing a different mass spectrum dn/dM_f. If we vary $f_{\rm NL}$ slowly enough, we can track the change in mass and position for individual halos: i.e., for each halo of mass M_0 for $f_{\rm NL} = 0$, we can uniquely identify a corresponding halo of mass M_f for $f_{\rm NL} \neq 0$, as long as $|f_{\rm NL}|$ is sufficiently small. Since we know precisely the number of halos as a function of M_0, if we can determine the mapping $M_0 \rightarrow M_f$, we will then have an estimate of the non-Gaussian mass function dn/dM_f via

$$\frac{dn}{dM_f} = \int dM_0 \frac{dn}{dM_0}\frac{dP}{dM_f}(M_0), \tag{11}$$

where $dP/dM_f(M_0)$ is the probability distribution that a Gaussian halo of mass M_0 maps to a non-Gaussian halo of mass M_f. Note that the probability distribution function dP/dM_f need not integrate to unity, $\int dM_f dP/dM_f \neq 1$ in general, since the total number of halos is not conserved: halos can merge or split as $f_{\rm NL}$ is varied.

The next step is to determine the probability distribution $dP/dM_f(M_0)$, by matching halos between Gaussian and non-Gaussian simulations. We match halos by requiring that matching pairs have significantly overlapping Lagrangian volumes; i.e. by requiring that halos have many particles in common, where particles are labeled by their Lagrangian coordinates in the initial conditions. For each halo M_f in a non-Gaussian run, we loop over the halo's particles and identify which Gaussian halos own those particles in the run with $f_{\rm NL} = 0$. The Gaussian halo owning the largest fraction (exceeding 1/3) of the particles is then identified as the match for the non-Gaussian halo M_f. Each Gaussian halo M_0 can have one, several, or zero matching non-Gaussian halos, depending on $f_{\rm NL}$. By stacking Gaussian halos of similar mass M_0, we can determine $dP/dM_f(M_0)$.

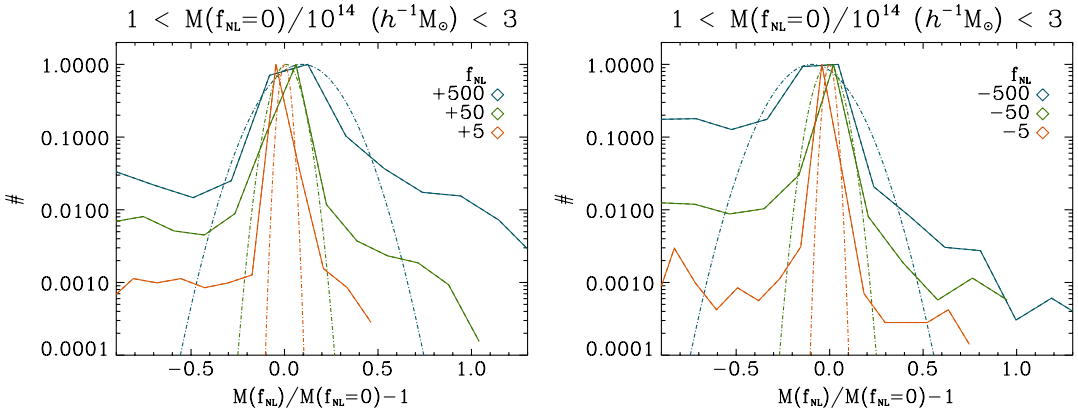

FIG. 4 (color online). The probability distribution that a Gaussian halo of mass M_0 maps to a non-Gaussian halo of mass M_f, i.e. $dP/dM_f(M_0)$. This plot can be understood as a (binned) slice through Fig. 3. Here we show the measured $dP/dM_f(M_0)$ (solid lines) and Gaussian fit (dashed line) for various $f_{\rm NL}$ in the mass bin $1 < M_0/10^{14}M_\odot < 3$. The left panel corresponds to $f_{\rm NL} > 0$ and the right panel corresponds to $f_{\rm NL} < 0$. Note that both the width and mean value of the PDF vary with $f_{\rm NL}$. The probability distribution is clearly poorly fit by a Gaussian; however, as discussed in the text, it provides an adequate fit given the precision with which we can determine the halo mass function from N-body simulations. Whereas the high mass tail for $f_{\rm NL} > 0$ (left panel) indicates that many $f_{\rm NL} = 0$ halos will merge into more massive ones, the low mass tail for $f_{\rm NL} < 0$ (right panel) accounts for the disruption of $f_{\rm NL} = 0$ halos into lighter ones.

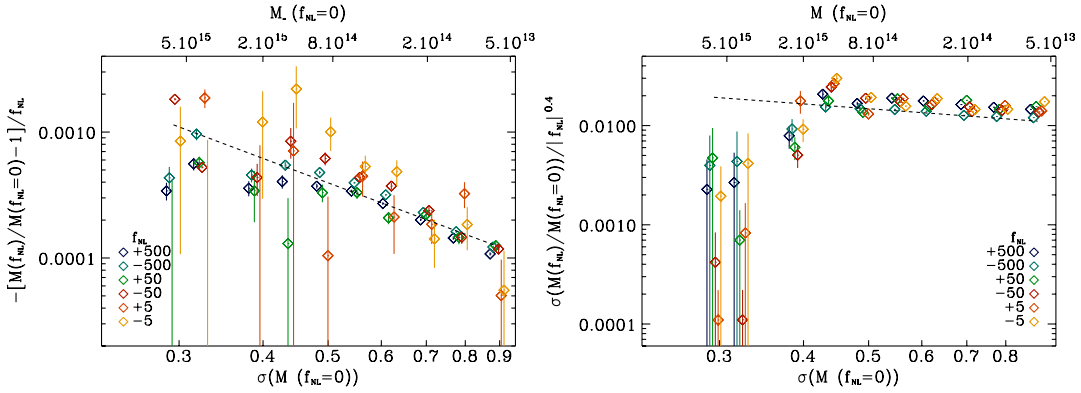

FIG. 5 (color online). Measured mean (left panel) and rms dispersion (right panel) of the mass shift, $M_f/M_0 - 1$, as a function of $\sigma(M_0, z)$. Note that measurements at various redshift outputs ($z = 0, 0.5, 1$) have been combined in this plot, and that the $f_{\rm NL}$ scaling has been divided out. The dashed lines show our fits to these moments; cf. Eqs. (12) and (13).

Examples of the probability distribution are shown in Fig. 4, and the mean and variance of the PDF are plotted in Fig. 5. The behavior of the mean $\langle M_f \rangle$ and variance are quite regular, and appear consistent with simple power laws:

$$\left\langle \frac{M_f}{M_0} \right\rangle - 1 = 1.3 \times 10^{-4} f_{\rm NL} \sigma_8 \sigma(M_0, z)^{-2}, \tag{12}$$

$$\mathrm{var}\left(\frac{M_f}{M_0}\right) = 1.4 \times 10^{-4} (f_{\rm NL}\sigma_8)^{0.8} \sigma(M_0, z)^{-1}, \tag{13}$$

where the rms overdensity dispersion $\sigma(M, z)$ is defined as usual by

$$\sigma^2 = \int \frac{k^3}{2\pi^2} P(k) W^2(kR) \frac{dk}{k}, \tag{14}$$

where we use a top-hat window $W(x) = 3j_1(x)/x$ for $R = (3M/4\pi\bar{\rho}_m)^{1/3}$, and $P(k)$ and $\bar{\rho}_m$ are the matter power spectrum and energy density, respectively.

Because we desire a simple fitting formula, we assume that we can approximate the PDF as a normalized Gaussian whose mean and variance are given above, even though the PDF shape is quite clearly non-Gaussian (cf. Fig. 4). As we show below, however, even this crude approximation is sufficient to achieve the ~10% precision in the halo mass function provided by standard fitting formulas for Gaussian simulations [71]. Then Eq. (11), together with $dP/dM_f(M_0)$ which is assumed to be a Gaussian with the mean and variance given in Eqs. (12) and (13), fully specifies our fitting function. Essentially, we have written the NG mass function as a convolution of the Gaussian mass function with a Gaussian kernel.

B. Review of previous fitting formulas

Before showing a comparison of the simulated mass function to our proposed fitting formula, we first describe alternative fitting formulas previously suggested in the literature: the extended Press-Schechter (EPS) formalism [72], and the model of Matarrese *et al.* [58], hereafter MVJ.

1. Extended Press-Schechter

The EPS formalism generalizes the widely used Press-Schechter [59] model, which posits that the fraction of mass in collapsed objects is equal to *twice* the fraction of the volume occupied by density peaks exceeding some critical overdensity δ_c. (The factor of 2 arises from the so-called "cloud-in-cloud" problem [69].) Therefore the collapsed fraction becomes

$$F(>M) \equiv 2 \int_{\delta_c}^{\infty} P(\delta; M) d\delta \tag{15}$$

$$= 2 \int_{\delta_c/\sigma(M)}^{\infty} P_{\rm G}(\nu) d\nu \tag{16}$$

where the probability distribution $P(\delta; M)$ that the density smoothed on mass scale M equals δ is simply $P_{\rm G}(\nu)$, the Gaussian distribution with zero mean and unit variance for $\nu = \delta_c/\sigma$, with $\sigma(M)$ given by Eq. (14). Note that $F(0) = 1$ for hierarchical cosmologies where $\sigma(M)$ diverges as $M \to 0$; that is, *all* matter is assumed to be in virialized objects of some mass.

The differential mass function may then be readily computed:

$$M \frac{dn}{dM} = \bar{\rho}_m \left| \frac{dF}{dM} \right|. \tag{17}$$

For a Gaussian PDF, the mass M enters the right-hand side of this expression only via the lower bound of the integral, $\delta_c/\sigma(M)$, so we immediately obtain

$$\left(\frac{dn}{d\ln M}\right)_{\rm PS} = 2 \frac{\bar{\rho}_m}{M} \frac{\delta_c}{\sigma} \left| \frac{d\ln\sigma}{d\ln M} \right| P_{\rm G}(\delta/\sigma). \tag{18}$$

This gives the well-known Press-Schechter mass function.

A class of fitting formulas based on this approach, and loosely called "extended Press-Schechter" (though appar-

ently unrelated to the work of Refs. [73,74]), attempt to generalize this argument using non-Gaussian PDFs. The most obvious way of making this generalization, i.e. inserting the non-Gaussian PDF $P(\delta; M)$ into Eq. (17), faces several immediate difficulties, however. First, the Press-Schechter factor of 2 is no longer valid; rather, the cloud-in-cloud correction will depend upon the specific form of non-Gaussianity. Second, the shape of the PDF $P(\delta; M)$ now depends upon M, and so we cannot simply replace the derivative of the integral in Eq. (17) by the integrand. Lastly, and prosaically, the Press-Schechter mass function does not, in fact, fit the halo mass function in N-body simulations well, and so starting from Press-Schechter is guaranteed to fail in fitting the non-Gaussian mass function.

The approach adopted by many previous works (e.g. [72]) has instead been to assume that, although Press-Schechter cannot be used to derive the Gaussian mass function, it may be used to compute the departure of the mass function from its Gaussian value, i.e.

$$\frac{n_{\rm NG}(M,z)}{n_{\rm G}(M,z)} = \frac{\frac{d}{dM}F_{\rm NG}(>M)}{\frac{d}{dM}F_{\rm G}(>M)} \tag{19}$$

where F is given by Eq. (15). In this approach, the non-Gaussian mass function is computed by multiplying the Gaussian mass function (not Press-Schechter, but Jenkins *et al.* [68] or Warren *et al.* [67]) by the above ratio.

The EPS prediction for the halo mass function is therefore given by the derivatives of the PDF tails given in Eq. (19) above. To implement this prescription, we compute the PDF tails directly from the initial conditions of our simulations: at each redshift we are interested in, we integrate the linearly evolved PDF to compute $F_{\rm NG}(M)$ at masses ranging from about $10^{12}M_\odot$ to $10^{16}M_\odot$. Then we compute the mean value of $F_{\rm NG}(M)$ averaged over 10 independent N-body simulations. Finally, we fit a cubic spline through the (computed mean of) $F_{\rm NG}(\ln M)$ and differentiate with respect to mass. Evaluation of this formula becomes extremely difficult at high masses, simply because the statistics of peaks at these high masses becomes too noisy.

2. MVJ

The MVJ [58] mass function is a further approximation to the EPS model described above. Instead of numerically computing the PDF and its tails for each $f_{\rm NL}$, it is assumed that the ratio in Eq. (19) may be determined from the skewness of the PDF. The expression for the mass function becomes [61]

$$\left(\frac{dn}{d\ln M}\right)_{\rm MVJ} = 2\frac{\bar\rho_M}{M}P_{\rm G}\left(\frac{\delta_*}{\sigma_M}\right) \times\left[\frac{1}{6}\frac{\delta_*^3}{\delta_c}\left|\frac{dS_{3,M}}{d\ln M}\right| + \delta_*\left|\frac{d\sigma_M}{d\ln M}\right|\right] \tag{20}$$

where

$$\delta_* \equiv \frac{\delta_c}{\sqrt{1 - S_{3,M}\delta_c/3}}, \tag{21}$$

$$S_{3,M} \equiv \frac{\langle\delta^3\rangle_M}{\langle\delta^2\rangle_M^2}, \tag{22}$$

where $\langle\delta^n\rangle_M$ is the nth moment of the density field evaluated on the characteristic mass scale M, and $S_{3,M}$ is the skewness on that mass scale.

The advantage of the MVJ formula is that it does not require specification of the PDF of the density field—however, it does require knowledge of the skewness $S_{3,M}$. In this work, we compute the moments of the density field directly from our simulations at the starting epoch $a = 0.02$, then scale them with the linear growth function to the desired epoch. We can then evaluate the MVJ expression in Eq. (21). Unlike the EPS approach described above, the formula does not become intractable at high masses. On the other hand, MVJ becomes onerously expensive to calculate at low levels of non-Gaussianity (e.g. $|f_{\rm NL}| \lesssim 100$), simply because the scatter in the measured skewness from run to run becomes comparable to that generated by primordial non-Gaussianity. To see this, note that the scatter in $\langle\delta^3\rangle_M$ is roughly $\sigma_3 \simeq \sigma_M^3\sqrt{15/N}$ for $N = M_{\rm box}/M$ samples. Approximating $\langle\delta^3\rangle_M \sim 6f_{\rm NL}\sigma_M^3\sigma_\phi$, then for a 512^3 grid and mass scale of 200 cells, and taking $\sigma_\phi = 4\times10^{-5}$, requiring $\langle\delta^3\rangle_M/\sigma_3 > 5$ translates into $|f_{\rm NL}| \gtrsim 100$.

C. Results and comparison to previous work

Figure 6 shows the ratios $n_{\rm NG}/n_{\rm G}$ for $f_{\rm NL} = 500$ (left panel) and -500 (right panel). Simulation values are denoted with error bars, colored black ($z = 0$), blue ($z = 0.5$), and red ($z = 1$). To compute the error bars in the ratios, and taking account of the fact that $n_{\rm NG}$ and $n_{\rm G}$ measurements are correlated, we have adopted the larger error of the two (rather than adding them in quadrature), which is the error in $n_{\rm G}$ ($n_{\rm NG}$) for $f_{\rm NL} > 0$ ($f_{\rm NL} < 0$). The solid lines denote our fits explained in Sec. III. Dashed lines refer to the EPS results, while the dotted lines represent the MVJ fitting function. The results clearly indicate that, while the EPS and the MVJ functions mutually agree,[3], they both overestimate the effects of non-Gaussianity as found by our simulations, at a level typically $\lesssim 100\%$, although dependent upon mass and redshift.

This result appears to disagree with the work of Kang *et al.* [60], who find a large discrepancy between EPS/MVJ and their simulations' mass function, in the sense that their

[3]The agreement between the EPS and MVJ is even better when an alternative expression is used for $f_{\rm NL} > 0$, as pointed out by Grossi *et al.* [61]; see their Eq. (4). We have not used this correction in our Fig. 6.

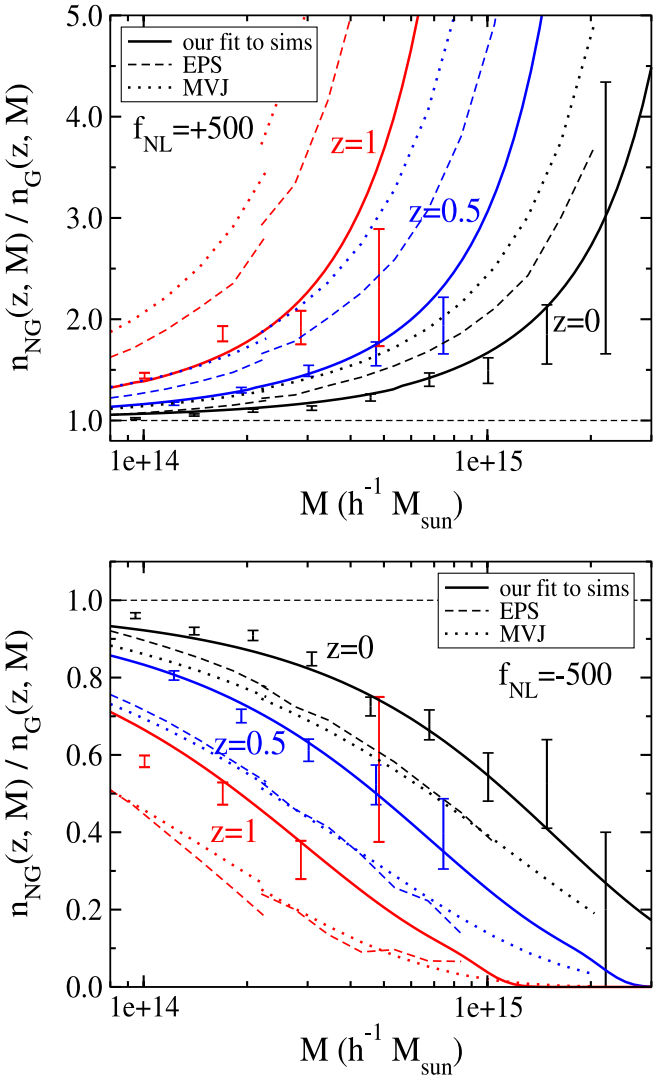

FIG. 6 (color online). Ratios of the NG to Gaussian mass functions as a function of mass and at redshifts $z = 0$ (black), $z = 0.5$ (blue), and $z = 1$ (red). Points with error bars denote results from our simulations. Solid lines represent our fitting formula. Dashed and dotted lines denote the EPS and MVJ fitting functions, respectively. Note that the EPS and MVJ agree mutually, but both significantly overestimate the effects of non-Gaussianity. (The discontinuity of EPS and MVJ fitting functions at $M \sim 2 \times 10^{14} M_\odot$ is due to transition from a smaller simulation box to the larger one.)

simulations show a much *larger* effect of non-Gaussianity than predicted by the EPS-type formalism. However, as noted by these authors, their simulations used a rather small number of particles ($\sim 128^3$) in a volume nearly 20 times smaller than ours, so it is unclear how well they probe the statistics of the rare objects of interest to us. In contrast, Grossi *et al.* [61] have found very good agreement between the MVJ formula and their simulations' results. While our fitting function is in mild disagreement with the MVJ fitting formula, it is unclear whether our simulations are in disagreement with the simulations of Grossi *et al.* [61]. Their simulations used a somewhat different cosmological model (higher σ_8) than ours; they have plotted cumulative rather than differential mass functions, and of course the error bars in both their plots and ours are considerable.

In summary, we conclude that our simple fitting function appears consistent with the measured mass function from our simulations to within $\sim 10\%$ over the entire range of masses and redshifts that we consider. Since this is the level of precision that various N-body codes agree with each other in the mass function [71], we have not attempted to achieve better agreement. EPS-like fitting formulas, such as the model of MVJ [58], appear to overestimate the effects of non-Gaussianity. The level of discrepancy increases with increasing mass and redshift.

V. HALO CLUSTERING

Beyond one-point statistics like the halo mass function, N-body simulations also allow us to compute higher order statistics like the correlation function or its Fourier transform, the power spectrum. As shown in Sec. II, we expect non-Gaussianity to produce pronounced effects on the halo power spectrum, specifically in the form of scale-dependent halo bias on large scales. This may seem somewhat surprising, due to very general arguments previously given in the literature that galaxy bias is expected to be independent of scale in the linear regime [75–77]. We can summarize the argument as follows. Suppose that the halo overdensity is some deterministic function of the local matter overdensity, $\delta_h = F(\delta)$. On large scales, where $|\delta| \ll 1$, we can Taylor expand this function, $\delta_h = a + b\delta + \ldots$. Keeping only the lowest order terms and requiring that $\langle \delta_h \rangle = 0$ then gives $\delta_h = b\delta$, which is linear deterministic bias. The key assumption in this argument was locality; i.e. that the halo abundance is determined entirely by the local matter density. N-body simulations with Gaussian initial conditions have confirmed that halo bias tends to a constant on large scales well in the linear regime.

Once we allow for primordial non-Gaussianity, however, the above argument need not hold. For example, in this paper we have considered NG of the form $f_{\rm NL}\Phi^2$, and note that the gravitational potential is a nonlocal quantity. Hence the locality-based argument above does not apply for this form of non-Gaussianity, and our derived scale dependence of the bias is not surprising. The specific form we have derived is particular to the quadratic, local form of NG that we have assumed; however, we expect that any NG that couples density modes with potential modes will, in general, lead to scale-dependent bias. On the other hand, non-Gaussianity of the form $f_{\rm NL}\delta^2$ does not lead to scale-dependent bias.

In order to test our prediction for the scale dependence of bias, we have computed halo bias in our N-body simulations by taking the ratio of the matter power spectrum

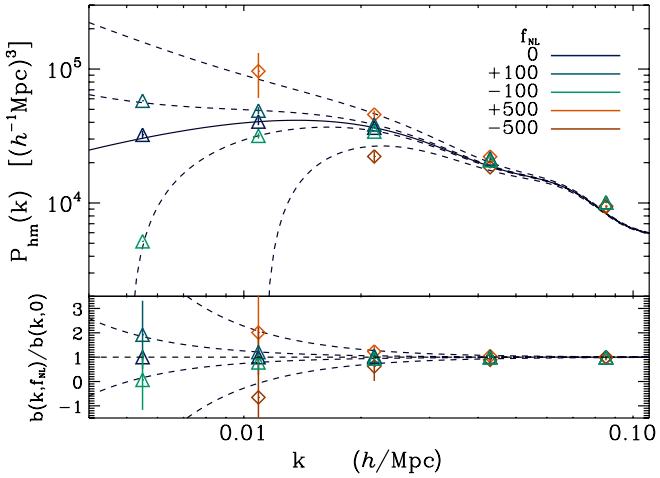

FIG. 7 (color online). Cross-power spectra for various $f_{\rm NL}$. The upper panel displays $P_{h\delta}(k)$, measured in our simulations at $z = 1$ for halos of mass $1.6 \times 10^{13} M_\odot < M < 3.2 \times 10^{13} M_\odot$. The solid line corresponds to the theoretical prediction for $P_{\delta\delta}$ with a fitted bias $b_0 = 3.25$. We see a strongly scale-dependent correction to the bias for $f_{\rm NL} \neq 0$, increasing towards small k (large scales). The bottom panel displays the ratio $b(k, f_{\rm NL})/b(k, f_{\rm NL} = 0)$. The errors are computed from the scatter amongst our simulations and within the bins. Triangles correspond to our large $(1024^3$ particle) simulations whereas diamonds correspond to our smaller $(512^3$ particle) simulations. The dotted lines correspond to our expression for the bias dependence on $f_{\rm NL}$ defined in Eq. (9).

$P_{\delta\delta}$ and the halo-matter cross spectrum $P_{h\delta} = \langle \delta_h^* \delta \rangle$. We have used the cross spectrum rather than the halo auto spectrum because the former should be less sensitive to shot noise from the small number of halos compared to dark matter particles. We have checked, however, that

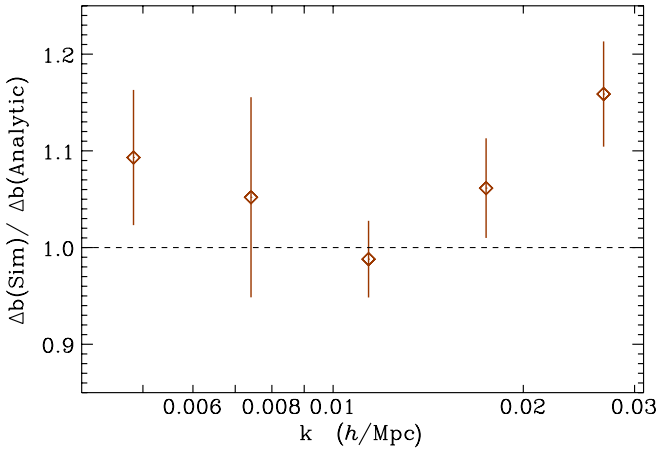

FIG. 8 (color online). Ratio of the bias shift Δb measured from our simulations to that predicted by Eq. (9), using $\delta_c = 1.686$. Biases were computed from cross spectra measured on 28 simulations with 5 various $f_{\rm NL}$ $(-500, -100, 100, 500)$, 3 various redshifts $(z = 0, 0.5, 1)$, and 5 halo mass bins. Note that at higher k, nonlinear evolution also generates scale dependence in the bias [80].

using the halo auto spectra to compute bias gives consistent results as the cross spectra; i.e. we find no evidence for stochasticity. Examples of the various power spectra and resulting bias factors are plotted in Fig. 7.

As can be seen, we numerically confirm the form of the predicted scale dependence. Because we focus on the statistics of rare objects, the errors on bias from individual simulations plotted in Fig. 8 are large. We therefore attempt to improve the statistics on the comparison by combining the bias measurements from multiple simulations. Figure 8 plots the average ratio between the bias measured in our simulations and our analytic prediction, Eq. (9), using $\delta_c = 1.686$ as predicted from the spherical collapse model [78]. In computing the average plotted in this figure, we used a uniform weighting across the different simulations, redshifts, and mass bins. Alternative weightings can shift the results by $\sim$10%, so we conservatively estimate the systematic error in our comparison to be 20%. The agreement between our numerical simulation results and our predicted bias scale dependence, Eq. (9), is excellent and perhaps surprising. Naively, we might expect a somewhat larger collapse threshold δ_c to apply, considering the ellipsoidal rather than spherical nature of the collapse of halos in this mass range [70].

VI. COSMOLOGICAL CONSEQUENCES

Having derived fitting formulas for the abundance and clustering of halos in NG models, we now investigate how well upcoming surveys may constrain $f_{\rm NL}$, and whether NG could possibly affect the constraints derived on other cosmological parameters. We focus on galaxy cluster surveys and redshift surveys. Cluster surveys aim to constrain cosmological parameters, in particular dark energy parameters, by exploiting the exponential sensitivity of the galaxy cluster abundance on cosmology. Similarly, a major goal for upcoming redshift surveys is to constrain dark energy by localizing baryonic acoustic oscillation (BAO) features in the galaxy power spectrum at multiple redshifts. Examples of upcoming surveys include the Atacama Cosmology Telescope,[4] South Pole Telescope,[5] Dark Energy Survey,[6] WiggleZ,[7] Planck,[8] SuperNova/Acceleration Probe,[9] and the Large Synoptic Survey Telescope.[10]

Because primordial non-Gaussianity affects both the abundance and power spectra of massive halos, both of these types of surveys will be well suited for constraining NG. On the other hand, potential NG could, in principle,

[4] http://wwwphy.princeton.edu/act/
[5] http://spt.uchicago.edu
[6] http://www.darkenergysurvey.org
[7] http://astronomy.swin.edu.au/wigglez/WiggleZ/Welcome.html
[8] http://www.rssd.esa.int/Planck
[9] http://snap.lbl.gov
[10] http://www.lsst.org

degrade the expected constraints on dark energy parameters, due to possible degeneracies. We use the Fisher matrix formalism to extract errors on seven cosmological parameters as well as $f_{\rm NL}$. Our estimates are only illustrative; accurate forecasts for specific surveys will require a more sophisticated analysis.

A. Constraints from $P(k)$: Galaxy surveys, BAO, and ISW

We can (crudely) estimate constraints on parameters $\{p_i\}$ derived from measurements of the power spectrum by assuming that band powers are measured with errors $\delta P = (P + n^{-1})/\sqrt{m}$, where P is the power in a band of width dk centered at wave number k, n is the number density of galaxies, and m is the number of independent Fourier modes sampled by the survey, roughly given by $m = (2\pi^2)^{-1} V k^2 dk$ [79]. Then the Fisher matrix can be written as

$$F_{ij} = \int_{k_{\rm min}}^{k_{\rm max}} \frac{\partial P}{\partial p_i} \frac{\partial P}{\partial p_j} \left(P + \frac{1}{n}\right)^{-2} \frac{V k^3}{2\pi^2} d\ln k. \qquad (23)$$

For simplicity, in Eq. (23) we use the linear theory power spectrum and number density corresponding to $z = 0.5$, and assume an all-sky, volume-limited survey extending to $z = 0.7$. We integrate over wave numbers between $k_{\rm min} = 10^{-3} h/{\rm Mpc}$ and $k_{\rm max} = 0.1 h/{\rm Mpc}$. The results are insensitive to $k_{\rm max}$ but depend strongly on $k_{\rm min}$. We believe the $k_{\rm min}$ used here is optimistic but reasonable. At high k (small scales), late-time nonlinear evolution can also generate scale-dependent bias [80]; however, the redshift and scale dependence of this effect is quite distinctive from NG and we ignore it here.

We assume that the target galaxies have properties similar to luminous red galaxies (LRGs) [81], with comoving number density $n = 4 \times 10^{-4} (h^{-1}\,{\rm Mpc})^{-3}$ and bias $b_0 = 2$. Equation (23) then gives estimated errors on $f_{\rm NL}$ of $\sigma(f_{\rm NL}) \approx 7$, which compares well with forecasted constraints on non-Gaussianity for Planck.

Unsurprisingly, we find little degeneracy between $f_{\rm NL}$ and other cosmological parameters, given its distinctive effect on the shape of the power spectrum. Accordingly, there is little reason to believe that BAO determinations of dark energy parameters will be biased by non-Gaussianity, especially since the scale dependence is small over the wave numbers of interest for the BAO wiggles. To quantify this effect, we determine the acoustic peak positions by locating extrema of the *ratio* of the power spectrum (including baryonic effects) to the power spectrum with zero baryons [82]. When multiplying the matter power spectrum with baryons by our scale-dependant bias, we find that $f_{\rm NL} = 100$ would shift the first BAO peak at $k \simeq 0.07 h/{\rm Mpc}$ by 0.4% at $z = 1$, and has a considerably smaller effect at the higher BAO peaks. The magnitude of this effect is comparable to the effect of nonlinear corrections to the power spectrum [83,84], although the NG effect is primarily important on large scales while nonlinearities are most important on small scales. In principle, NG and nonlinearities could conspire to lead to a $\sim$1%–2% bias in the dark energy equation of state parameter w inferred from BAO observations [83], and so a careful joint analysis allowing both for NG and nonlinear corrections will be required, which should not be difficult.

Another probe of $P_{h\delta}$ on large scales is the cross correlation between cosmic microwave background (CMB) temperature anisotropies and large-scale structure, due to the integrated Sachs-Wolfe (ISW) effect [85–87]. First detections of the ISW effect from cross correlations of the Wilkinson Microwave Anisotropy Probe (WMAP) with various large-scale surveys have been obtained with reported detections at the 2σ–4σ level [88–95]. A combined analysis yields a $\simeq 5\sigma$ detection [96], whereas a cosmic variance limited measurement would allow a $\simeq 7.5\sigma$ detection for the currently favored ΛCDM cosmology, and a somewhat more significant detection if the dark energy equation of state parameter is smaller [97,98]. The cross correlation between large-scale structure and CMB is directly proportional to a weighted projection of the scale-dependant bias. Since the z and k dependence of our bias is very specific, we do not expect it to be severely degenerate with other parameters affecting the amplitude of the ISW effect (mostly w and Ω_m for a flat universe). We can thus translate the ISW detection level into constraints on $f_{\rm NL}$. For the sake of simplicity, we assume that the ISW signal comes from $z \simeq 1$ and is dominated by the angular multipole $\ell \simeq 20$, corresponding to a wave number $k \simeq 6.66 \times 10^{-3} h/{\rm Mpc}$ at $z = 1$. [97]. According to Eq. (9) the current 3σ (5σ) detections of ISW translate into upper limits on $|f_{\rm NL}|$ of 123 (61) (1σ) assuming a bias $b_0 = 2$, as appropriate for LRGs [93]. A prospective 7.5σ detection would translate into $|f_{\rm NL}| \lesssim 38$ (1σ). These estimates are clearly very crude, but are likely correct at the order-of-magnitude level. In comparison, the current limits from CMB bispectrum measurements from WMAP give $-54 < f_{\rm NL} < 114$ (95% CL) [6] whereas Planck is expected to constrain $|f_{\rm NL}| < 10$ (1σ) [99].

In summary, the large-scale galaxy power spectrum appears capable of constraining local NG quite stringently for surveys reaching $\sim$Gpc scales: $|f_{\rm NL}| \lesssim 10$. ISW or BAO observations could provide somewhat weaker bounds on $f_{\rm NL}$, though of course any constraints they can provide would be independent of the CMB bispectrum and therefore worthwhile. Our estimates of forecasted bounds on $f_{\rm NL}$ were rather crude, but given the encouraging results, a more sophisticated treatment for specific survey parameters appears warranted.

B. Constraints from cluster counts

We next consider how well upcoming cluster surveys can constrain $f_{\rm NL}$ by measurements of the cluster mass function dn/dM. Other forms of non-Gaussianity may also be constrained by these surveys [100], but we focus on the $f_{\rm NL}$ form. For our fiducial survey parameters, we consider a fixed, redshift-independent lower mass limit of $M_{\rm lim} = 2 \times 10^{14} M_\odot$ and assume redshift bins of width $\Delta z = 0.1$ uniformly distributed between $z = 0.1$ and $z = z_{\rm max} \leq 2.0$. We simultaneously vary seven cosmological parameters besides $f_{\rm NL}$: A, the normalization of the primordial power spectrum at $k_{\rm fid} = 0.002h\,{\rm Mpc}^{-1}$; physical matter and baryon densities $\Omega_m h^2$ and $\Omega_b h^2$; the spectral index n_s; the sum of the neutrino masses m_ν; the matter energy density today relative to critical Ω_m; and the equation of state parameter of dark energy w. We assume no mass information (which would improve our parameter constraints) but also no systematic errors (which would degrade the constraints). We further assume 5000 square degrees on the sky, roughly consistent with expectations for the Dark Energy Survey or the South Pole Telescope. The fiducial survey has about 7000 clusters for $\sigma_8 = 0.76$ cosmology, and about 23 000 for $\sigma_8 = 0.9$. We use WMAP3 [6] cosmological parameters in determining error forecasts. The mass power spectrum $\Delta^2(k, a) \equiv k^3 P(k, a)/(2\pi^2)$ is written as

$$\Delta^2(k, a) = \frac{4A}{25\Omega_M^2}\left(\frac{k}{k_{\rm fid}}\right)^{n_s - 1}\left(\frac{k}{H_0}\right)^4 g^2(a) T^2(k) \qquad (24)$$

where $T(k)$ is the transfer function adopted from Eisenstein and Hu [82], and the growth function $g(a)$ is computed exactly by integrating the well-known second order differential equation for growth [e.g. Eq. (1) in [101]].

Following the results of Sec. IV A, we assume that the mass function may be written as

$$\frac{dn}{dM}(z, M) = \left(\frac{dn}{dM}\right)_{\rm Jenk}(z, M)\left[\frac{n_{\rm NG}(z, M)}{n_{\rm G}(z, M)}\right] \qquad (25)$$

where the non-Gaussian correction is computed using either our fitting formula or EPS for comparison. For a given mass function, the total number of objects in a redshift interval of width Δz and centered at z is

$$N(z, \Delta z) = \Omega_{\rm survey} \int_{z - \Delta z/2}^{z + \Delta z/2} n(z, M_{\rm min}) \frac{dV(z)}{d\Omega dz} dz \qquad (26)$$

where $\Omega_{\rm survey}$ is the total solid angle covered by the survey, $n(z, M_{\rm min})$ is the comoving density of clusters more massive than $M_{\rm min}$, and $dV/d\Omega dz$ is the comoving volume element. We assume $\Omega_{\rm survey} = 5000$ square degrees, roughly consistent with expectations for the Dark Energy Survey or the South Pole Telescope. The fiducial survey has about 7000 clusters (for $\sigma_8 = 0.76$ cosmology) and about 23 000 (for $\sigma_8 = 0.9$).

Assuming Poisson statistics, the Fisher information matrix reads [102–104]

$$F_{ij}^{\rm clus} = \sum_k \frac{1}{N_k(z_k, \Delta z)} \frac{\partial N_k}{\partial p_i} \frac{\partial N_k}{\partial p_j} \qquad (27)$$

where p_i are the 8 cosmological parameters including $f_{\rm NL}$, N_k is the number of clusters in the kth redshift bin, and the sum runs over the redshift bins extending to maximal redshift $z_{\rm max}$.[11] We assume no mass information (which would improve our parameter constraints) but also no systematic errors (which would degrade the constraints).

Lastly, we add a Planck prior on the parameter set, neglecting forecasted constraints on $f_{\rm NL}$ expected from future CMB bispectrum measurements (since we are interested in the sensitivity to NG of cluster counts alone). The full Fisher matrix is given by

$$F = F^{\rm clus} + F^{\rm CMB}. \qquad (28)$$

Figure 9 shows the result of our Fisher matrix estimate, i.e. the forecasted errors on $f_{\rm NL}$ as a function of the maximum extent of the cluster survey, $z_{\rm max}$. Black solid and dashed lines show the marginalized and unmarginalized errors using our fitting formulas for $n_{\rm NG}/n_{\rm G}$, while the red lines show the errors using the EPS formalism. The former errors are clearly well behaved, and asymptote at

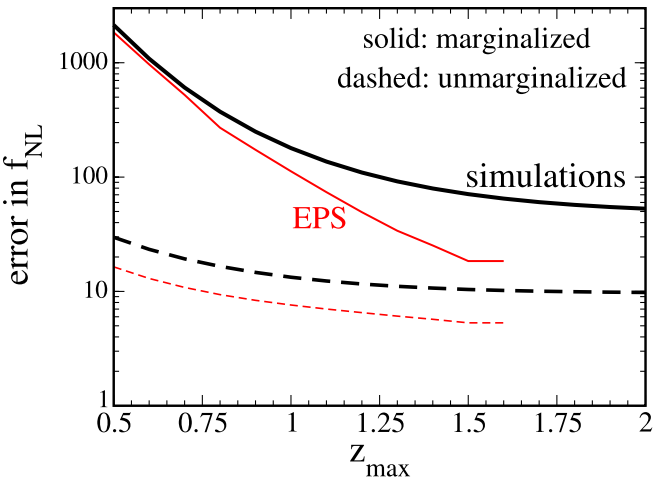

FIG. 9 (color online). Forecasted errors on $f_{\rm NL}$ from measurements of the cluster mass function, as a function of the maximum extent of the cluster survey, $z_{\rm max}$. Black solid and dashed lines show the marginalized and unmarginalized error using our fitting formulas for $n_{\rm NG}/n_{\rm G}$, while the red lines show the errors using the EPS formalism. Even though we use the average of the non-Gaussian PDF tail over 10 simulations, the EPS result becomes noisy at $z \gtrsim 1.5$ due to poor sampling of the tails. Furthermore, it is clear that the EPS errors on $f_{\rm NL}$ underestimate those based on our simulations by up to a factor of 3.

high $z_{\rm max}$ as expected since cluster abundance rapidly vanishes. On the other hand, EPS-produced $f_{\rm NL}$ errors disagree with the simulations by up to a factor of 3. Moreover, even though we use the average of the non-Gaussian PDF tail over 10 simulations, the EPS result becomes noisy at $z \gtrsim 1.5$ due to poor sampling of the tails. The magnitude of the discrepancy between EPS estimates and our simulations appears similar even for the higher $\sigma_8 = 0.9$ model. As with our estimates from power spectrum constraints, we do not find significant degeneracies between $f_{\rm NL}$ and other cosmological parameters (correlation coefficient $\lesssim 0.5$).

Since we find a weaker effect on cluster abundance than previous formulas like EPS or MVJ, this implies that constraints found by Sefusatti *et al.* [72], who performed a similar Fisher matrix estimate but with somewhat different assumptions, will be weaker once the NG sensitivity is calibrated off simulations. Direct quantitative comparison to two other relevant papers, Kang *et al.* [60] and Grossi *et al.* [61], is however difficult since these authors do not compute cosmological parameter error estimates.

VII. DISCUSSION

We have quantified the effects of primordial non-Gaussianity on the abundance and power spectra of massive halos. Our two principal results are as follows.

First, we have provided a new fitting formula for the halo mass function. The formula is based on matching halos in Gaussian and non-Gaussian simulations: for $f_{\rm NL} > 0$ the corresponding halos are more massive than in the Gaussian case, and vice versa. The formula is consistent with the measured mass function from our simulations to within $\sim 10\%$ over the entire range of masses and redshifts that we consider. Being essentially a convolution of the Gaussian mass function and a Gaussian kernel [Eqs. (11)–(13)], the formula is also easy to use and does not require estimating the extreme tails of the non-Gaussian PDF of the density field. Our results also indicate that previous work based on extended Press-Schechter–type formulas overestimated the effects of non-Gaussianity on the abundance of halos by a factor of ~ 2 over the relevant mass scales.

Second, we showed both analytically and numerically that non-Gaussianity (in the $f_{\rm NL}$ model) leads to strong scale dependence of the bias of dark matter halos. We find remarkably good agreement between our analytic expression and our numerical results. Measurement of the power spectrum of biased objects therefore provides a new avenue to detect and measure non-Gaussianity. While cluster counts can constrain NG at a level comparable to existing CMB constraints, $|f_{\rm NL}| \lesssim 100$, we found that future large-scale redshift surveys can potentially do much better, roughly $|f_{\rm NL}| \lesssim 10$. We do not find significant degeneracies between $f_{\rm NL}$ and dark energy parameters in our Fisher matrix calculations, either for mass function measurements or power spectrum measurements. More precise estimates will require considerably more sophisticated treatments than we have attempted in our illustrative examples above.

We close this paper by considering, in light of our findings, the optimal methods for constraining NG of the $f_{\rm NL}$ form. Measurements of the power spectrum would appear the most promising; observations of high redshift, highly clustered objects on large scales would allow the strongest constraints on the scale-dependent bias signature of $f_{\rm NL}$. Fortunately, upcoming BAO surveys will likely provide the necessary observations of, e.g. LRGs. Photometric surveys may also be useful in this regard. Since the effects of NG are most pronounced on large scales, rather than small scales, precise spectroscopic redshifts may not be necessary. Photometric redshifts with errors of order $\Delta z \approx 0.03$ have already been achieved for LRGs and for optically selected groups and clusters with prominent red sequences [46,105,106]. At $z = 0.5$, this corresponds to roughly $100h^{-1}$ Mpc comoving, fairly small compared to the $\sim$Gpc scales where NG becomes most important. Since photometric surveys can cover wider areas more deeply than spectroscopic surveys, they may turn out to provide tighter bounds.

Besides their abundance and clustering, the internal properties of massive halos may also be sensitive to non-Gaussianity. For instance, the concentrations and substructure content of massive halos have been found to depend upon primordial NG [107]. Our simulations lacked sufficient force resolution to explore this in detail, but we note in passing that multiple groups find a tension between observations of massive lensing clusters and theoretical predictions for Gaussian perturbations [108–111].

Another intriguing possibility for probing primordial NG is to use statistics of the largest voids in the universe. Just as the abundance and clustering of high density peaks are affected by non-Gaussianity, so are the same properties for deep voids (albeit with an opposite sign, cf. Fig. 1). In a sense, because voids are not as nonlinear as overdense regions, their properties are more easily related to the initial Lagrangian underdensities whose statistics are straightforward to compute. Voids may be detected at high redshift as a deficit of Lyman-α forest absorption features in quasi-stellar object (QSO) spectra. The SDSS has already measured spectra for high redshift QSO's over a roughly $\sim 8000\,{\rm deg}^2$ area, corresponding to a volume of $\gtrsim 30({\rm Gpc}/h)^3$ [112]. Each QSO spectrum typically probes $\sim 400h^{-1}$ Mpc, and the typical transverse separation between QSO sight lines in SDSS is $\sim 100h^{-1}$ Mpc (P. McDonald, private communication), so measurements of the clustering of ~ 10 Mpc-sized voids on $\sim$Gpc scales may already be feasible.

Finally, we note that our conclusions are based on simulations implementing a very specific type of local primordial non-Gaussianity quantified by the $f_{\rm NL}$ parameter. The validity of our conclusions in the context of other types of primordial non-Gaussianity is the subject of ongoing studies.

ACKNOWLEDGMENTS

We are grateful to the organizers and participants of the "Life Beyond the Gaussian" workshop held at KICP (Chicago, June 2007) where a preliminary version of this work was presented and discussed. We thank Xuelei Chen and the National Astronomical Observatories of the Chinese Academy of Sciences in Beijing where part of this work was completed. Finally we thank Niayesh Afshordi, Neil Barnaby, Wayne Hu, Lam Hui, Nikhil Padmanabhan, Uros Seljak, Robert Smith, and Emiliano Sefusatti for useful conversations, Wayne Hu for providing the CMB Fisher matrix, and Pat McDonald for discussions of the Lyman-α forest. All simulations were performed on CITA's Sunnyvale cluster, funded by the Canada Foundation for Innovation and the Ontario Research Fund for Research Infrastructure. This work was supported by the Canadian Institute for Theoretical Astrophysics (CITA) and the Natural Sciences and Engineering Research Council of Canada (NSERC).

APPENDIX A: THE ABUNDANCE AND CLUSTERING OF HIGH PEAKS

In this appendix we derive analytic expressions for the abundance and clustering of regions above the spherical collapse threshold δ_c. We first review previous results derived for Gaussian statistics, and then show how they are modified by $f_{\rm NL}$ non-Gaussianity.

1. Review of Gaussian results

We begin by identifying massive halos at late times with high peaks in the initial density distribution $\delta(\boldsymbol{x}) \equiv \delta\rho(\boldsymbol{x})/\bar{\rho}$. Note that we work entirely in early-time, Lagrangian coordinates $\boldsymbol{x}$ in this section rather than late-time, Eulerian coordinates. Earlier work [59,69,113,114] has shown that the abundance of peaks above threshold $\delta_c \approx 1.686$ reasonably describes (at the order-of-magnitude level) the statistics of halos forming at subsequent times. We briefly review some of these previous results, as the methods will be used in our analysis.

Following the Press and Schechter [59] ansatz, we smooth the density field and assume that density peaks with $\delta > \delta_c$ produce halos. The density smoothed on scale R is given by

$$\delta_R(\boldsymbol{x}) = (2\pi)^{-3} \int d^3\boldsymbol{k}\, \delta_{\boldsymbol{k}} W(kR) e^{i\boldsymbol{k}\cdot\boldsymbol{x}}, \tag{A1}$$

where $W(x)$ is some smoothing window, e.g. top-hat or Gaussian. Assuming that the Fourier modes $\delta_{\boldsymbol{k}}$ are Gaussian distributed with power spectrum $P(k) = \langle|\delta_k|^2\rangle$, then $\delta_R(\boldsymbol{x})$ also has a Gaussian distribution, with variance

$$\sigma_\delta^2(R) = \langle\delta_R^2\rangle = \int d\ln k \frac{k^3 P(k)}{2\pi^2} W^2(kR). \tag{A2}$$

The probability P_1 for a given randomly selected region to exceed the threshold δ_c is then simply the integral of the Gaussian probability distribution,

$$P_1 = \int_{\delta_c}^\infty d\delta \frac{dP}{d\delta} = \int_{\nu_c}^\infty d\nu (2\pi)^{-1/2} e^{-\nu^2/2} = \frac{1}{2}\mathrm{erfc}\left(\frac{\nu_c}{\sqrt{2}}\right), \tag{A3}$$

where $\nu = \delta/\sigma_\delta$, and similarly $\nu_c = \delta_c/\sigma_\delta$.

The same power spectrum $P(k)$ describing the density variance $\sigma_\delta^2(R)$ also gives the matter correlation function $\xi(r_{12})$, and, following an elegant argument by Kaiser [113], can also be used to determine the correlation function of rare peaks. Let us compute the probability P_2 that two randomly selected regions separated by distance $r_{12} \gg R$ are both above threshold. Again, this is simply an integral over the (joint) Gaussian distribution:

$$P_2 = \int_{\delta_c}^\infty d\delta_1 \int_{\delta_c}^\infty d\delta_2 \frac{\exp(-\frac{1}{2}\boldsymbol{\delta}\cdot\boldsymbol{\Sigma}^{-1}\cdot\boldsymbol{\delta})}{2\pi|\boldsymbol{\Sigma}|^{1/2}} \tag{A4}$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2)$, the covariance matrix is given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_\delta^2 & \xi \\ \xi & \sigma_\delta^2 \end{pmatrix}, \tag{A5}$$

and the matter correlation function $\xi(r_{12})$ is given by

$$\xi(r_{12}) = \int d\ln k \frac{k^3 P(k)}{2\pi^2} W^2(kR) j_0(kr_{12}). \tag{A6}$$

Rescaling the δ's by their variance, this becomes

$$P_2 = \int_{\nu_c}^\infty d\nu_1 \int_{\nu_c}^\infty d\nu_2 \frac{\exp(-\frac{1}{2}\boldsymbol{\nu}\cdot\mathbf{S}^{-1}\cdot\boldsymbol{\nu})}{2\pi|\mathbf{S}|^{1/2}}, \tag{A7}$$

where

$$\mathbf{S} = \begin{pmatrix} 1 & \psi \\ \psi & 1 \end{pmatrix}, \tag{A8}$$

and we follow the notation of BBKS [114] in writing the normalized correlation function as $\psi(r_{12}) = \xi(r_{12})/\sigma_\delta^2$; note that $\psi < 1$.

To evaluate the integral in Eq. (A7), we change coordinates to variables that are uncorrelated. Using the Cholesky decomposition of the covariance matrix $\mathbf{S} = \mathbf{C}^\mathsf{T}\mathbf{C}$, we write $\boldsymbol{\nu} = \mathbf{C}^\mathsf{T}\boldsymbol{y}$ for new variables $\boldsymbol{y}$. Next, we rotate coordinates $\boldsymbol{y} = \mathbf{R}\cdot\boldsymbol{x}$ to bring the point $(\nu_1 = \nu_c, \nu_2 = \nu_c)$ along the x_1 axis. Then the integral becomes

$$P_2 = \int d^2\boldsymbol{x} \frac{\exp(-\frac{1}{2}|\boldsymbol{x}|^2)}{2\pi} \Theta[(1,0)\cdot\mathbf{C}^\mathsf{T}\cdot\mathbf{R}\cdot\boldsymbol{x} - \nu_c] \times \Theta[(0,1)\cdot\mathbf{C}^\mathsf{T}\cdot\mathbf{R}\cdot\boldsymbol{x} - \nu_c], \tag{A9}$$

where the Heaviside function Θ accounts for the two integration bounds. Since $\psi < 1$, we can write this as

$$P_2 = \frac{1}{2\pi}\int_{x_c}^{\infty} dx_1 e^{-x_1^2/2}\int_{-c(x_1-x_c)}^{c(x_1-x_c)} dx_2 e^{-x_2^2/2} = \frac{1}{2\pi}\int_{x_c}^{\infty} dx_1 e^{-x_1^2/2} f(x_1), \quad \text{(A10)}$$

where $c = \sqrt{(1+\psi)/(1-\psi)}$ and $x_c = \nu_c\sqrt{2/(1+\psi)}$. We could easily evaluate the integral for $f(x_1)$ in terms of the error function, but the resulting integral over x_1 would then not be analytic. However, we can derive an approximate solution in the limit $\nu_c \gg 1$. For integrals of the form

$$I = \int_{x_0}^{\infty} dx e^{-x^2/2} f(x), \quad \text{(A11)}$$

we can construct an asymptotic series by repeated partial integrations:

$$I \approx e^{-x_0^2/2}\left[\frac{f(x_0)}{x_0}\left(1 - \frac{1}{x_0^2} + \ldots\right) + \frac{f'(x_0)}{x_0^2} + \ldots\right]. \quad \text{(A12)}$$

In our case, $f(x_c) = 0$ and $f'(x_c) = 2c$. Therefore, in the limit $\nu_c \gg 1$, we obtain

$$P_2 \approx \frac{1}{2\pi} e^{-x_c^2/2}\frac{2c}{x_c^2} = \frac{1}{2\pi} e^{-\nu_c^2/(1+\psi)}\frac{(1+\psi)^{3/2}}{(1-\psi)^{1/2}}\nu_c^{-2}. \quad \text{(A13)}$$

Comparing this expression to the probability for a single peak to be above threshold then gives the peak-peak correlation function $\xi_{\rm pk}$, which in the limit $\nu_c \gg 1$, $\psi \ll 1$ becomes

$$1 + \xi_{\rm pk} = P_2/P_1^2 \approx e^{\nu_c^2(1-1/(1+\psi))} \approx 1 + \nu_c^2\psi = 1 + \frac{\nu_c^2}{\sigma_\delta^2}\xi, \quad \text{(A14)}$$

and therefore the (Lagrangian) bias $b_L^2 = \xi_{\rm pk}/\xi$ becomes

$$b_L \approx \nu_c^2/\delta_c. \quad \text{(A15)}$$

2. Non-Gaussianity

Our discussion so far has merely reviewed previous results for Gaussian fluctuations; we now turn to non-Gaussian fluctuations. As noted above, we focus on NG of the form

$$\Phi_{\rm NG} = \phi + f_{\rm NL}(\phi^2 - \langle\phi^2\rangle). \quad \text{(A16)}$$

We adopt the approximation of Sec. II that the heights of rare peaks are modified by NG as $\delta_{\rm NG} \approx \delta[1 + 2f_{\rm NL}\phi]$, where in this appendix we adopt the notation that ϕ refers to the primordial potential. At late times, ϕ decays as the growth suppression factor $g(a)$.

Let us first consider the one-point distribution of peaks above threshold, $\delta_{\rm NG} > \delta_c$. We express this as an integral over Gaussian variables ϕ and δ; the integration bound then becomes $\delta_{\rm NG} = \delta(1 + 2f_{\rm NL}\phi) > \delta_c$, and using the fact that typically $f_{\rm NL}|\phi| \ll 1$, we have $\delta > \delta_c(1 - 2f_{\rm NL}\phi)$. The probability for $\delta_{\rm NG}$ to exceed threshold then is

$$P_1 = \int d\phi \int_{\delta_c(1-2f_{\rm NL}\phi)}^{\infty} d\delta \frac{\exp(-\frac{1}{2}(\phi,\delta)\cdot\Sigma^{-1}\cdot(\phi,\delta))}{2\pi|\Sigma|^{1/2}} = \int d\mu \int_{\nu_c-\eta\mu}^{\infty} d\nu \frac{\exp(-\frac{1}{2}(\mu,\nu)\cdot\mathbf{S}^{-1}\cdot(\mu,\nu))}{2\pi|\mathbf{S}|^{1/2}}, \quad \text{(A17)}$$

where $\mu = \phi/\sigma_\phi$, $\nu = \delta/\sigma_\delta$, $\nu_c = \delta_c/\sigma_\delta$, and $\eta = 2f_{\rm NL}\sigma_\phi\nu_c$. We write the off-diagonal part of the normalized covariance matrix $\mathbf{S}$ as $\langle\mu\nu\rangle = r$, where the cross-correlation coefficient r is not to be confused with the peak-peak separation r_{12} appearing above and below. By a coordinate transformation, we can orient the integration bound along a single axis. Changing coordinates from (μ, ν) to $(\mu, \upsilon = \nu + \eta\mu)$, and noting that the variance of υ is $\langle\upsilon^2\rangle = 1 + 2\eta r + \eta^2$, we can integrate out μ to obtain

$$P_1 = \frac{1}{\sqrt{2\pi}}\int_{x_c}^{\infty} dx e^{-x^2/2} = \frac{1}{2}\mathrm{erfc}\left(\frac{x_c}{\sqrt{2}}\right), \quad \text{(A18)}$$

where $x_c = \nu_c/\sqrt{1 + 2\eta r + \eta^2}$. The effect of NG on the peak abundance is therefore simply to rescale the threshold density by a mass- and redshift-dependent factor.

Next, we turn to the peak-peak correlation function. As in the Gaussian case, we write the probability for two points both to be above threshold as

$$P_2 = \int d^4\boldsymbol{u}\frac{\exp(-\frac{1}{2}\boldsymbol{u}\cdot\boldsymbol{\Sigma}\cdot\boldsymbol{u})}{(2\pi)^2|\boldsymbol{\Sigma}|^{1/2}}\Theta(\nu_1 + \eta\mu_1 - \nu_c) \times \Theta(\nu_2 + \eta\mu_2 - \nu_c), \quad \text{(A19)}$$

where $\boldsymbol{u} = (\mu_1, \mu_2, \nu_1, \nu_2)$, and the notation is otherwise the same as above. We write the off-diagonal parts of the normalized covariance matrix as $\langle\nu_1\nu_2\rangle = \psi$, $\langle\mu_1\mu_2\rangle = \gamma$, $\langle\nu_1\mu_1\rangle = r$, and $\langle\nu_1\mu_2\rangle = \beta$. As above, we change variables from ν_i to $\upsilon_i = \nu_i + \eta\mu_i$ to align the integration bounds along the density coordinate axes. This allows us to integrate out the two potential variables, leaving behind a 2-D integral. Rescaling the remaining two variables by their (identical) variance, and again writing $x_c = \nu_c/\sqrt{1 + 2\eta r + \eta^2}$, brings the integral to the form

$$\int_{x_c}^{\infty} dx_1 \int_{x_c}^{\infty} dx_2 \frac{\exp(-\frac{1}{2}\boldsymbol{x}\cdot\mathbf{S}\cdot\boldsymbol{x})}{2\pi|\mathbf{S}|^{1/2}} \quad \text{(A20)}$$

where the off-diagonal component of $\mathbf{S}$ is $\langle x_1 x_2\rangle = \chi$, given by

$$\chi = \frac{\psi + 2\eta\beta + \eta^2\gamma}{1 + 2\eta r + \eta^2}. \quad \text{(A21)}$$

The form of Eq. (A20) is identical to Eq. (A7), with $\nu_c \to x_c$ and $\psi \to \chi$. So we can immediately write down the

approximate solution,

$$P_2 \approx \frac{1}{2\pi} e^{-x_c^2/(1+\chi)} x_c^{-2} \frac{(1+\chi)^{3/2}}{(1-\chi)^{1/2}}. \quad \text{(A22)}$$

Comparing with the single-peak probability, we obtain the peak-peak correlation function, in the limit $\nu_c \gg 1$, $\chi \ll 1$:

$$1 + \xi_{\rm pk} = P_2/P_1^2 \approx 1 + x_c^2 \chi, \quad \text{(A23)}$$

which to lowest order in η becomes

$$\xi_{\rm pk} \approx \nu_c^2[\psi + 2\eta(\beta - 2r\psi)] \approx \nu_c^2(\psi + 2\eta\beta) = b_L^2(\xi_{\delta\delta} + 4f_{\rm NL}\delta_c\xi_{\phi\delta}), \quad \text{(A24)}$$

where b_L is the Lagrangian bias obtained for Gaussian peaks; cf. Eq. (A15). Note that in the first line of Eq. (A24), we neglect $r\psi$ relative to β since $\xi_{\delta\delta}(r_{12})/\sigma_{\delta\delta}$ is smaller than $\xi_{\phi\delta}(r_{12})/\sigma_{\phi\delta}$ by a factor scaling like $(R/r_{12})^2$, where R is the smoothing scale of the peak and r_{12} is the peak-peak separation.

The peak-peak correlation function is now no longer simply proportional to the matter correlation function, implying that the peak bias is not independent of scale. Fourier transforming this expression gives the peak power spectrum,

$$P_{\rm pk} = b_L^2(P_{\delta\delta} + 4f_{\rm NL}\delta_c P_{\phi\delta}), \quad \text{(A25)}$$

which gives a scale-dependent change in the bias due to NG of

$$\Delta b(k) = 2b_L f_{\rm NL}\delta_c \frac{P_{\phi\delta}}{P_{\delta\delta}} = 2b_L f_{\rm NL}\delta_c \frac{3\Omega_m}{2agr_H^2k^2}, \quad \text{(A26)}$$

where we have used the relation between the potential-density cross spectrum and the matter power spectrum $P_{\phi\delta} = (3\Omega_m/2agr_H^2k^2)P_{\delta\delta}$, arising from the Poisson equation. The total Lagrangian bias is then $b_L(k) = b_L + \Delta b(k)$.

[1] J. Maldacena, J. High Energy Phys. 5 (2003) 13.
[2] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, Nucl. Phys. **B667**, 119 (2003).
[3] P. Creminelli, J. Cosmol. Astropart. Phys. 10 (2003) 003.
[4] D. H. Lyth and Y. Rodriguez, Phys. Rev. Lett. **95**, 121302 (2005).
[5] D. Seery and J. E. Lidsey, J. Cosmol. Astropart. Phys. 06 (2005) 003.
[6] D. N. Spergel, R. Bean, O. Doré, M. R. Nolta, C. L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page *et al.*, Astrophys. J. Suppl. Ser. **170**, 377 (2007).
[7] P. Creminelli, L. Senatore, M. Zaldarriaga, and M. Tegmark, J. Cosmol. Astropart. Phys. 03 (2007) 005.
[8] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 04 (2004) 001.
[9] N. Bartolo, S. Matarrese, and A. Riotto, Phys. Rev. D **69**, 043503 (2004).
[10] D. H. Lyth and Y. Rodriguez, Phys. Rev. D **71**, 123508 (2005).
[11] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. van Tent, Phys. Rev. D **73**, 083522 (2006).
[12] L. E. Allen, S. Gupta, and D. Wands, J. Cosmol. Astropart. Phys. 01 (2006) 006.
[13] X. Chen, Phys. Rev. D **72**, 123518 (2005).
[14] N. Barnaby and J. M. Cline, Phys. Rev. D **73**, 106012 (2006).
[15] N. Barnaby and J. M. Cline, Phys. Rev. D **75**, 086004 (2007).
[16] N. Barnaby and J. M. Cline, J. Cosmol. Astropart. Phys. 07 (2007) 017.
[17] M. Sasaki, J. Valiviita, and D. Wands, Phys. Rev. D **74**, 103003 (2006).
[18] X. Chen, M.-x. Huang, S. Kachru, and G. Shiu, J. Cosmol. Astropart. Phys. 01 (2007) 002.
[19] X. Chen, R. Easther, and E. A. Lim, J. Cosmol. Astropart. Phys. 06 (2007) 023.
[20] T. Battefeld and R. Easther, J. Cosmol. Astropart. Phys. 03 (2007) 020.
[21] H. Assadullahi, J. Valiviita, and D. Wands, Phys. Rev. D **76**, 103003 (2007).
[22] D. Battefeld and T. Battefeld, J. Cosmol. Astropart. Phys. 05 (2007) 012.
[23] R. Bean, S. E. Shandera, S. H. Henry Tye, and J. Xu, J. Cosmol. Astropart. Phys. 05 (2007) 004.
[24] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, Phys. Rep. **402**, 103 (2004).
[25] D. Babich, Phys. Rev. D **72**, 043003 (2005).
[26] D. Babich, P. Creminelli, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 08 (2004) 009.
[27] P. Creminelli, L. Senatore, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 03 (2007) 019.
[28] K. M. Smith and M. Zaldarriaga, arXiv:astro-ph/0612571.
[29] J. R. Fergusson and E. P. S. Shellard, Phys. Rev. D **76**, 083523 (2007).
[30] T. Falk, R. Rangarajan, and M. Srednicki, Astrophys. J. **403**, L1 (1993).
[31] X.-c. Luo and D. N. Schramm, Phys. Rev. Lett. **71**, 1124 (1993).
[32] A. Gangui, F. Lucchin, S. Matarrese, and S. Mollerach, Astrophys. J. **430**, 447 (1994).
[33] L.-M. Wang and M. Kamionkowski, Phys. Rev. D **61**, 063504 (2000).

[34] L. Verde, R. Jimenez, M. Kamionkowski, and S. Matarrese, Mon. Not. R. Astron. Soc. **325**, 412 (2001).
[35] R. Scoccimarro, E. Sefusatti, and M. Zaldarriaga, Phys. Rev. D **69**, 103513 (2004).
[36] E. Sefusatti and E. Komatsu, Phys. Rev. D **76**, 083004 (2007).
[37] F. Lucchin and S. Matarrese, Astrophys. J. **330**, 535 (1988).
[38] J. Robinson and J. E. Baker, arXiv:astro-ph/9905098.
[39] A. J. Benson, C. Reichardt, and M. Kamionkowski, Mon. Not. R. Astron. Soc. **331**, 71 (2002).
[40] S. Matarrese, L. Verde, and R. Jimenez, Astrophys. J. **541**, 10 (2000).
[41] E. Komatsu *et al.* (WMAP), Astrophys. J. Suppl. Ser. **148**, 119 (2003).
[42] L. Verde, M. Kamionkowski, J. J. Mohr, and A. J. Benson, Mon. Not. R. Astron. Soc. **321**, L7 (2001).
[43] E. Rozo, R. H. Wechsler, B. P. Koester, T. A. McKay, A. E. Evrard, D. Johnston, E. S. Sheldon, J. Annis, and J. A. Frieman, arXiv:astro-ph/0703571.
[44] B. Koester *et al.* (SDSS), Astrophys. J. **660**, 239 (2007).
[45] V. R. Eke *et al.* (2dFGRS Team), Mon. Not. R. Astron. Soc. **348**, 866 (2004).
[46] H. K. C. Yee *et al.* (RCS-2), arXiv:astro-ph/0701839.
[47] J. P. Willis *et al.*, Mon. Not. R. Astron. Soc. **675**, 1369 (2005).
[48] I. Valtchanov *et al.*, Astron. Astrophys. **423**, 75 (2004).
[49] Z. Haiman, J. J. Mohr, and G. P. Holder, Astrophys. J. **553**, 545 (2001).
[50] S. Majumdar and J. J. Mohr, Astrophys. J. **585**, 603 (2003).
[51] S. Wang, J. Khoury, Z. Haiman, and M. May, Phys. Rev. D **70**, 123008 (2004).
[52] R. A. Battye and J. Weller, Mon. Not. R. Astron. Soc. **362**, 171 (2005).
[53] M. Lima and W. Hu, Phys. Rev. D **72**, 043006 (2005).
[54] L. Marian and G. M. Bernstein, Phys. Rev. D **73**, 123525 (2006).
[55] M. Takada and S. Bridle, New J. Phys. **9**, 446 (2007).
[56] J. Robinson and J. E. Baker, Mon. Not. R. Astron. Soc. **311**, 781 (2000).
[57] J. Robinson, E. Gawiser, and J. Silk, Astrophys. J. **532**, 1 (2000).
[58] S. Matarrese, L. Verde, and R. Jimenez, Astrophys. J. **541**, 10 (2000).
[59] W. H. Press and P. Schechter, Astrophys. J. **187**, 425 (1974).
[60] X. Kang, P. Norberg, and J. Silk, Mon. Not. R. Astron. Soc. **376**, 343 (2007).
[61] M. Grossi, K. Dolag, E. Branchini, S. Matarrese, and L. Moscardini, arXiv:0707.2516.
[62] E. Komatsu and D. N. Spergel, Phys. Rev. D **63**, 063002 (2001).
[63] A. Shirokov and E. Bertschinger, arXiv:astro-ph/0505087.
[64] A. V. Shirokov, Ph.D. thesis, Massachusetts Institute of Technology, 2005.
[65] T. Padmanabhan, *Structure Formation in the Universe* (Cambridge University Press, Cambridge, New York, 1993).
[66] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, Astrophys. J. **292**, 371 (1985).
[67] M. S. Warren, K. Abazajian, D. E. Holz, and L. Teodoro, Astrophys. J. **646**, 881 (2006).
[68] A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida, Mon. Not. R. Astron. Soc. **321**, 372 (2001).
[69] J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, Astrophys. J. **379**, 440 (1991).
[70] J. R. Bond and S. T. Myers, Astrophys. J. Suppl. Ser. **103**, 1 (1996).
[71] Z. Lukic, K. Heitmann, S. Habib, S. Bashinsky, and P. M. Ricker, arXiv:astro-ph/0702360.
[72] E. Sefusatti, C. Vale, K. Kadota, and J. Frieman, Astrophys. J. **658**, 669 (2007).
[73] C. Lacey and S. Cole, Mon. Not. R. Astron. Soc. **262**, 627 (1993).
[74] R. G. Bower, Mon. Not. R. Astron. Soc. **248**, 332 (1991).
[75] P. Coles, Mon. Not. R. Astron. Soc. **262**, 1065 (1993).
[76] J. N. Fry and E. Gaztanaga, Astrophys. J. **413**, 447 (1993).
[77] R. J. Scherrer and D. H. Weinberg, Astrophys. J. **504**, 607 (1998).
[78] J. E. Gunn and J. R. I. Gott, Astrophys. J. **176**, 1 (1972).
[79] C. Blake, D. Parkinson, B. Bassett, K. Glazebrook, M. Kunz, and R. C. Nichol, Mon. Not. R. Astron. Soc. **365**, 255 (2006).
[80] R. E. Smith, R. Scoccimarro, and R. K. Sheth, Phys. Rev. D **75**, 063512 (2007).
[81] N. Padmanabhan, D. J. Schlegel, U. Seljak, A. Makarov, N. A. Bahcall, M. R. Blanton, J. Brinkmann, D. J. Eisenstein, D. P. Finkbeiner, J. E. Gunn *et al.*, Mon. Not. R. Astron. Soc. **378**, 852 (2007).
[82] D. J. Eisenstein and W. Hu, Astrophys. J. **511**, 5 (1999).
[83] M. Crocce and R. Scoccimarro, Phys. Rev. D **77**, 023533 (2008).
[84] R. E. Smith, R. Scoccimarro, and R. K. Sheth, Phys. Rev. D **77**, 043525 (2008).
[85] R. K. Sachs and A. M. Wolfe, Astrophys. J. **147**, 73 (1967).
[86] R. G. Crittenden and N. Turok, Phys. Rev. Lett. **76**, 575 (1996).
[87] R. Bean and O. Dore, Phys. Rev. D **69**, 083503 (2004).
[88] S. Boughn and R. Crittenden, Nature (London) **427**, 45 (2004).
[89] M. R. Nolta *et al.*, Astrophys. J. **608**, 10 (2004).
[90] P. Fosalba *et al.*, Astrophys. J. Lett. **597**, L89 (2003).
[91] R. Scranton *et al.*, arXiv:astro-ph/0307335.
[92] P. Fosalba and E. Gaztañaga, Mon. Not. R. Astron. Soc. **350**, L37 (2004).
[93] N. Padmanabhan *et al.*, Phys. Rev. D **72**, 043525 (2005).
[94] N. Afshordi, Y.-S. Loh, and M. A. Strauss, Phys. Rev. D **69**, 083524 (2004).
[95] A. Cabre *et al.*, Mon. Not. R. Astron. Soc. **372**, L23 (2006).
[96] R. Scranton *et al.* (unpublished).
[97] N. Afshordi, Phys. Rev. D **70**, 083536 (2004).
[98] W. Hu and R. Scranton, Phys. Rev. D **70**, 123002 (2004).
[99] K. M. Smith and M. Zaldarriaga, arXiv:astro-ph/0612571.
[100] S. Sadeh, Y. Rephaeli, and J. Silk, Mon. Not. R. Astron. Soc. **380**, 637 (2007).

[101] A. Cooray, D. Huterer, and D. Baumann, Phys. Rev. D **69**, 027301 (2004).
[102] G. Holder, Z. Haiman, and J. Mohr, Astrophys. J. **560**, L111 (2001).
[103] D. Huterer, A. Kim, L. M. Krauss, and T. Broderick, Astrophys. J. **615**, 595 (2004).
[104] M. Lima and W. Hu, Phys. Rev. D **70**, 043504 (2004).
[105] N. Padmanabhan *et al.* (SDSS), Mon. Not. R. Astron. Soc. **359**, 237 (2005).
[106] O. Ilbert *et al.*, arXiv:astro-ph/0603217.
[107] V. Avila-Reese, P. Colín, G. Piccinelli, and C. Firmani, Astrophys. J. **598**, 36 (2003).
[108] M. D. Gladders, H. Hoekstra, H. K. C. Yee, P. B. Hall, and L. F. Barrientos, Astrophys. J. **593**, 48 (2003).
[109] N. Dalal, G. Holder, and J. F. Hennawi, Astrophys. J. **609**, 50 (2004).
[110] J. F. Hennawi, N. Dalal, P. Bode, and J. P. Ostriker, Astrophys. J. **654**, 714 (2007).
[111] T. Broadhurst and R. Barkana, arXiv:0801.1875.
[112] P. McDonald, U. Seljak, S. Burles, D. J. Schlegel, D. H. Weinberg, R. Cen, D. Shih, J. Schaye, D. P. Schneider, N. A. Bahcall *et al.*, Astrophys. J. Suppl. Ser. **163**, 80 (2006).
[113] N. Kaiser, Astrophys. J. Lett. **284**, L9 (1984).
[114] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).