Primordial non-Gaussianity, statistics of collapsed objects, and the integrated Sachs-Wolfe effect

Niayesh Afshordi[*] and Andrew J. Tolley[†]

Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario, N2L 2Y5, Canada

Any hint of non-Gaussianity in the cosmological initial conditions will provide us with a unique window into the physics of the early Universe. We show that the impact of a small local primordial non-Gaussianity (generated on superhorizon scales) on the statistics of collapsed objects (such as galaxies or clusters) can be approximated by using slightly modified, but *Gaussian*, initial conditions, which we describe through simple analytic expressions. Given that numerical simulations with Gaussian initial conditions are relatively well studied, this equivalence provides us with a simple tool to predict signatures of primordial non-Gaussianity in the statistics of collapsed objects. In particular, we describe the predictions for non-Gaussian mass function, and also confirm the recent discovery of a nonlocal bias on large scales [N. Dalal, O. Dore, D. Huterer, and A. Shirokov, Phys. Rev. D **77**, 123514 (2008).].[S. Matarrese and L. Verde, Astrophys. J. **677**, L77 (2008).], as a signature of primordial non-Gaussianity. We then study the potential of galaxy surveys to constrain non-Gaussianity using their autocorrelation and cross correlation with the cosmic microwave background (CMB) (due to the integrated Sachs-Wolfe effect), as a function of survey characteristics, and predict that they will eventually yield an accuracy of $\Delta f_{NL} \sim 0.1$ and 3, respectively, which will be better than or competitive with (but independent of) the best predicted constraints from the CMB. Interestingly, the cross correlation of the CMB and the NRAO VLA Sky Survey galaxy survey already shows a hint of a large local primordial non-Gaussianity: $f_{NL} = 236 \pm 127$.

I. INTRODUCTION

The standard cosmological model has been tremendously successful in explaining a large range of cosmological observations, ranging from cosmic microwave background (CMB) anisotropies, to the galaxy redshift surveys, supernovae Ia, baryonic acoustic oscillations and the Lyman-α forest in the quasar spectra. However, our clues into the nature of the initial conditions of the Universe, as well as its underlying physics, remain scarce.

The theory of cosmological slow-roll inflation remains the most successful and widely studied model for the cosmological initial conditions. However, despite the large inflow of cosmological observations over the past few years, it remains practically impossible to distinguish a large class of inflationary scenarios. Moreover, given the limited amount of observables accessible to today's observer, it is not clear whether any possible future observations will be able to falsify the inflationary scenario. One exception to this dilemma would be a possible detection of non-Gaussianity of cosmological initial conditions, which is predicted to be unobservably small in standard slow-roll inflationary scenarios (e.g. [1]). While a large primordial non-Gaussianity can still be accommodated in a general inflationary scenario through introducing fine-tuned secondary light fields, higher derivative interactions, or non-

adiabatic initial conditions, it could also be interpreted as a smoking gun for alternative early universe models, such as the new ekpyrotic scenario [2], which naturally predicts a large primordial non-Gaussianity [3].

The best current observational constraints on the primordial non-Gaussianity comes from the statistics of the all-sky CMB maps surveyed by the Wilkinson Microwave Anisotropy Probe (WMAP). Using the 3 yr data release of the WMAP CMB maps, Yadav and Wandelt [4] claimed a detection of an f_{NL}-type local primordial non-Gaussianity [see Eq. (1)], finding $f_{NL} = 87 \pm 30$. This is inconsistent with the expectation from single-field slow-roll inflationary models that predict a much smaller value $|f_{NL}| \lesssim 0.1$, and might be seen as evidence for a new ekpyrotic scenario which predicts $f_{NL} \sim 10$–100 [3]. However, the more recent data release of 5 yr WMAP maps shows a smaller value of $f_{NL} = 51 \pm 30$ [5], which is now consistent with the minimal inflationary expectation at the 2σ level. It is thus clear that an accurate determination of the f_{NL} parameter can have important implications for the early universe models. The next generation of all-sky CMB surveys, namely, the Planck satellite, is expected to have enough sensitivity to reduce the error on f_{NL} to ~ 3 (e.g. [6]). However, CMB secondary anisotropies (such as the Sunyaev-Zel'dovich effect) are expected to contaminate the CMB at this level, and thus prevent a more accurate measurement.

Can we expect complementary (and competitive) constraints from the study of a large-scale structure at low

_____

[*]nafshordi@perimeterinstitute.ca
[†]atolley@perimeterinstitute.ca

redshifts? This question was traditionally addressed by looking at the mass function of galaxy clusters, which was expected to be a sensitive probe of non-Gaussian initial conditions (e.g. see [7–9] and references therein). Unfortunately, the predicted constraints are not competitive with the CMB, unless f_{NL} has a strong scale dependence. Recently, Sefusatti and Komatsu [10] argued that the analysis of the bispectrum of upcoming large-volume galaxy surveys can yield competitive constraints with the CMB, but their method relies on the accuracy of a local bias model for the distribution of galaxies on small scales. Interestingly, Dalal *et al.* [7] recently predicted the presence of an anomalous large-scale power in galaxy distribution as a possible smoking gun for local primordial non-Gaussianity (also see [11,12]). This effect would only be seen on very large scales, and thus cannot be confused with any other causal astrophysical effect (although galactic contamination or large-angle survey systematics could give rise to a similar spurious power). Shortly before the submission of this paper, Slosar *et al.* [13] published a comprehensive study of this effect in the present large-scale cosmological galaxy surveys, leading to a significant constraint: $f_{NL} = 28 \pm 25$. We will briefly comment on this result in Sec. VI.

In this paper, we present a fully analytic approximation that could recast the statistics of collapsed objects (i.e. galaxies, clusters, etc.) with local primordial non-Gaussian initial conditions, in terms of the same statistics with Gaussian initial conditions but a modified power spectrum. In Sec. II, we introduce the approximation that recasts the non-Gaussian linear density field in terms of a linear combination of Gaussian fields, for collapsing objects of a given mass. Section III applies this result to the elliptical collapse framework, and compares the resulting mass function with numerical simulations. We then discuss the implications for the large-scale clustering of galaxies (due to a nonlocal bias correction) in Sec. IV, and show that it could yield competitive (or even superior) constraints ($\Delta f_{NL} \sim 0.1$), in comparison with the best upcoming CMB constraints. Finally, Sec. V shows that cross correlation of CMB with galaxy surveys, due to its sensitivity to the large-scale galaxy bias, can similarly put strong constraints on the primordial non-Gaussianity which are also competitive with the CMB ($\Delta f_{NL} \sim 3$), and possibly less sensitive to the galaxy survey systematics. Current cross correlation data between the WMAP CMB maps and the NRAO VLA Sky Survey (NVSS) radio galaxy survey already show a $\sim 2\sigma$ hint for such a non-Gaussian signal.

Unless mentioned otherwise, we use the WMAP5 best fit ΛCDM flat cosmology [5] with $\Omega_m = 0.279$, $\Omega_b = 0.0462$, $n_s = 0.96$, $h = 0.701$ throughout. For all our numerical calculations, we use the Eisenstein and Hu analytic approximation [14] for the matter transfer function that includes baryonic features, although we show the simpler Bardeen-Bond-Kaiser-Szalay (BBKS), dark matter only, analytic fit [15] in Eqs. (5) and (12), in order to illustrate

the asymptotic structure of the transfer function. Finally, all the errors on f_{NL} predicted or quoted in this paper are at 68% (1σ) confidence level.

II. NON-GAUSSIAN LINEAR PERTURBATIONS

The local primordial non-Gaussianity has the following form on superhorizon scales:

$$\Phi(\mathbf{x})|_{\text{super}} = \Phi_{pG}(\mathbf{x}) - f_{NL}[\Phi_{pG}^2(\mathbf{x}) - \langle \Phi_{pG}^2(\mathbf{x}) \rangle], \quad (1)$$

which was first introduced in [16], although similar formulations were used in [17–20]. Here Φ is the longitudinal (or conformal Newtonian) metric perturbation:

$$ds^2 = (1 + 2\Phi)dt^2 - a^2(t)(1 - 2\Phi)dx^i dx^i, \quad (2)$$

and Φ_{pG} is a statistically homogeneous random Gaussian field with a nearly scale-invariant power spectrum. On subhorizon scales, and during the matter era, the Newtonian potential Φ is suppressed by the transfer function $T(k)$, and the metric growth factor $g(t)$ (normalized to unity for $1 \ll z \ll 10^3$) [21]:

$$\Phi_{\mathbf{k}} = T(k)g(t)\Phi_{\mathbf{k}}|_{\text{super}}. \quad (3)$$

The transfer function can be well approximated by the BBKS fitting formula [15]:

$$T(k = q\Omega_m h^2 \, \text{Mpc}^{-1}) \simeq \frac{\ln[1 + 2.34q]}{2.34q}[1 + 3.89q + (16.2q)^2$$
$$+ (5.47q)^3 + (6.71q)^4]^{-1/4}, \quad (4)$$

while $g(t)$ only becomes important at the onset of dark energy domination.

Let us *define* δ_m via the Poisson equation

$$\delta_{m,\mathbf{k}} \equiv -\frac{k^2 \Phi_{\mathbf{k}}}{4\pi G a^2 \rho_m} = -A^{-1}(k; t)\Phi_{\mathbf{k}}, \quad (5)$$

where

$$A(k; t) = \frac{3}{2}\Omega_m H_0^2 a(t)^{-1} k^{-2}. \quad (6)$$

While δ_m reduces to the matter overdensity ($= \delta\rho_m/\rho_m$) in the Newtonian limit ($k \gg aH$), the correspondence in the $k \lesssim aH$ limit will be gauge dependent, and generally does not hold. However, the determining factor in the dynamics of a spherically collapsing region is its density in units of the local critical density Ω_{tot}. This can be calculated in longitudinal gauge and within the matter era:

$$\Omega_{\text{tot}} - 1 = \frac{10\nabla^2\Phi}{9a^2H^2}. \quad (7)$$

This justifies our definition of δ_m in Eq. (5), i.e. only the $\nabla^2\Phi$ term (and not the full matter overdensity) is relevant for the statistics of collapsed objects.

Combining the above formulas, we find the following expression for the non-Gaussian δ_m perturbations:

$$\delta_{m,\mathbf{k}} = \delta_{mG,\mathbf{k}} + \frac{f_{NL}}{g(t)} \int \frac{d^3 k'}{(2\pi)^3}$$

$$\times \frac{A(|\mathbf{k} - \mathbf{k}'|; t)A(k'; t)T(k)}{T(|\mathbf{k} - \mathbf{k}'|)T(k')A(k; t)} \delta_{mG,\mathbf{k}-\mathbf{k}'}\delta_{mG,\mathbf{k}'}, \quad (8)$$

where

$$\delta_{mG,\mathbf{k}} = -A^{-1}(k; t)T(k)g(t)\Phi_{pG,\mathbf{k}} \qquad (9)$$

is the Gaussian density perturbation (if $f_{NL} = 0$). In order to find a local form for δ_m in real space, let us consider the approximation:

$$\frac{A(|\mathbf{k} - \mathbf{k}'|; t)A(k'; t)T(k)}{T(|\mathbf{k} - \mathbf{k}'|)T(k')A(k; t)} \simeq \frac{A(k'; t)}{T(k')} + \frac{A(|\mathbf{k} - \mathbf{k}'|; t)}{T(|\mathbf{k} - \mathbf{k}'|)}$$
$$- \frac{A(k; t)}{T(k)}. \qquad (10)$$

Assuming k, $k' \gg k_{\mathrm{eq}}$ (associated with collapsed objects today), notice that for large k's $A(k; t)/T(k)$ only logarithmically depends on k. Therefore, since Eq. (10) holds for constant A/T, it must be a good approximation in the large k limit. The only exception to this limit is when $k' \ll k_{\mathrm{eq}}$ ($|\mathbf{k} - \mathbf{k}'| \ll k_{\mathrm{eq}}$), for which the first (second) term on the right-hand side dominates ($A/T \propto k^{-2}$ for small k's). However, since the two other terms are still slowly varying, Eq. (10) remains a good approximation. In fact, one can numerically confirm that Eq. (10) remains accurate at the percent level for $k' \lesssim k$.

Plugging Eq. (10) into Eq. (9), we arrive at the following local approximation for the linear non-Gaussian overdensity in the real space:

$$\delta_m(\mathbf{x}; t, M) \simeq \delta_{mG}(\mathbf{x}; t, M)$$
$$- f_{NL}\Big\{2\Phi_{pG}(\mathbf{x}; M)\delta_{mG}(\mathbf{x}; t, M) + \frac{\epsilon(M)}{ag}$$
$$\times [\delta_{mG}(\mathbf{x}; t, M)]^2\Big\},$$

$$(11)$$

where, for collapsed haloes of mass M, we assumed $k/\pi \sim R^{-1}(M) \equiv [4\pi\bar{\rho}_m/(3M)]^{1/3}$, and thus

$$\epsilon(M) \equiv \frac{a(t)A(k; t)}{T(k)}$$

$$\simeq \frac{1.9 \times 10^{-5}}{(\Omega_m h^2/0.137)\ln(1 + 2.34q)}[1 + 0.081q^{-1}$$
$$+ 0.129q^{-2} + 0.00192q^{-3} + 0.00049q^{-4}]^{1/4},$$

$$\text{with} \quad q = \left(\frac{M}{1.9 \times 10^{15}M_\odot}\right)^{-1/3}\left(\frac{\Omega_m h^2}{0.137}\right)^{-2/3}. \qquad (12)$$

A second approximation that we made in writing Eq. (11) was to ignore the contribution of Gaussian modes with a wavelength smaller than the mass filter ($k > \pi/R(M)$) to the quadratic term. This is justified as the power *per mode*

drops as $\sim k^{-3}$ on small scales, and thus the constructive interference of short wavelength modes in the quadratic term cannot significantly change the density averaged over the mass filter. In other words, the contribution of the Fourier modes to the total power in the non-Gaussian term (averaged over the mass filter) is sharply cut off beyond $k \sim \pi/R(M)$, which justifies using the smoothed Gaussian fields on the right-hand side of Eq. (11).

III. STATISTICS OF COLLAPSED OBJECTS

In order to estimate the impact of non-Gaussianity on the statistics of collapsed objects, we will use the ellipsoidal collapse framework [22], where the collapse criterion is that $\delta_m(\mathbf{x}; t, M)$ (smoothed with a top-hat filter of mass M) exceeds the mass-dependent collapse threshold $\delta_{\mathrm{ec}}(M)$:

$$\delta_{\mathrm{ec}}(M) \simeq q^{1/2}\delta_{\mathrm{sc}}[1 + \beta(q\nu)^{-\gamma}], \qquad \nu = \left[\frac{\delta_{\mathrm{sc}}}{\sigma(M)}\right]^2, \qquad (13)$$

where $(q, \beta, \gamma) = (0.707, 0.47, 0.615)$ are fitted to simulations [23], and $\delta_{\mathrm{sc}} \simeq 1.68$ is the spherical collapse threshold.

For Gaussian initial conditions, numerical simulations indicate that, to a good approximation, the mass function (or number density) of collapsed objects of mass M is simply a function of $\sigma(M)/\delta_{\mathrm{ec}}(M)$, where $\sigma^2(M) = \langle \delta_{mG}^2(\mathbf{x}; t, M) \rangle$.

On the other hand, Eq. (11) shows that, for the non-Gaussian initial conditions, the local overdensity depends on the local values of both δ_{mG} and Φ_{pG}. To linear order in f_{NL}, the elliptical collapse condition reduces to

$$\delta_m > \delta_{\mathrm{ec}} \Rightarrow \delta_{mG} - 2f_{NL}\delta_{\mathrm{ec}}\Phi_{pG} \gtrsim \delta_{\mathrm{ec}} + \frac{\epsilon f_{NL}}{ag}\delta_{\mathrm{ec}}^2. \qquad (14)$$

We can now define $\tilde{\delta}_{mG}$ and $\tilde{\delta}_{\mathrm{ec}}$ as

$$\tilde{\delta}_{mG} \equiv \delta_{mG} - 2f_{NL}\delta_{\mathrm{ec}}\Phi_{pG}, \qquad \tilde{\delta}_{\mathrm{ec}} \equiv \delta_{\mathrm{ec}} + \frac{\epsilon f_{NL}}{ag}\delta_{\mathrm{ec}}^2. \qquad (15)$$

We thus find that, as $\tilde{\delta}_{mG}$ is a Gaussian random field, we could still use the old Gaussian framework, provided we replace δ_{ec} by a slightly modified collapse threshold $\tilde{\delta}_{\mathrm{ec}}$. Moreover, as the mass function only depends on the ratio $\tilde{\sigma}(M)/\tilde{\delta}_{\mathrm{ec}}$, we have

$$\frac{\tilde{\sigma}(M)}{\tilde{\delta}_{\mathrm{ec}}} = \frac{\langle\tilde{\delta}_{mG}\rangle^{1/2}}{\tilde{\delta}_{\mathrm{ec}}} = \frac{\bar{\sigma}(M)}{\delta_{\mathrm{ec}}}, \qquad (16)$$

where

$$\bar{\sigma}(M) \equiv \tilde{\sigma}(M) \frac{\delta_{ec}}{\tilde{\delta}_{ec}}$$

$$\simeq [\sigma^2(M) - 4f_{NL}\delta_{ec}\langle\Phi_{pG}\delta_{mG}\rangle]^{1/2}\left(1 - \frac{\epsilon f_{NL}}{ag}\delta_{ec}\right)$$

$$\simeq \sigma(M)\left[1 - \frac{f_{NL}\delta_{ec}}{ag}\left(\frac{2ag\langle\Phi_{pG}\delta_{mG}\rangle}{\sigma^2(M)} + \epsilon\right)\right].$$
(17)

We also note that

$$\sigma^2(M) = \int \frac{d^3k}{(2\pi)^3}P(k)W^2(kR), \tag{18}$$

$$ag\langle\Phi_{pG}\delta_{mG}\rangle = -a\int \frac{d^3k}{(2\pi)^3}\frac{A(k)}{T(k)}P(k)W^2(kR), \tag{19}$$

where, $W(kR)$ is the Fourier transform of the spherical top-hat filter of radius R:

$$W(x) = \frac{3(\sin x - x\cos x)}{x^3}, \tag{20}$$

and $4\pi R^3\bar{\rho}_m/3 = M$, while ϵ was defined in Eq. (12).

Figure 1 shows the non-Gaussian correction to $\bar{\sigma}(M)$ for $f_{NL} = 100$ as described in Eq. (17), for three different redshifts. We first notice the increasing trend with mass, i.e. the effective non-Gaussian amplification of the power spectrum is more significant for larger haloes/clusters. Second, we notice the increasing trend with redshift, which simply results from the inverse scaling with ag in Eq. (17).
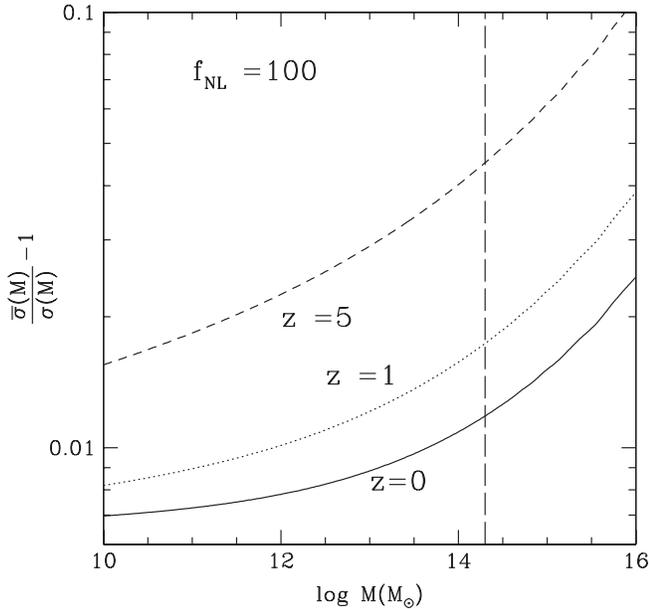

FIG. 1. The non-Gaussian correction to $\bar{\sigma}(M)$, the *effective* amplitude of density fluctuations for collapsed objects of mass M. The solid, dotted, and dashed lines are for $z = 0$, 1, 5, respectively, while we assumed $f_{NL} = 100$. The vertical line shows M_8, the mass within comoving $8h^{-1}$ Mpc.

The change in the shape of the correction at different redshifts, however, is a result of the mass dependence of the elliptical collapse threshold [Eq. (13)], which kicks in below the (redshift-dependent) nonlinear mass scale.

The vertical line shows M_8, the mass within comoving $8h^{-1}$ Mpc, which is often used to normalize the amplitude of linear density fluctuations: $\sigma_8 = \sigma(M_8)$. For our assumed cosmology, we find approximately

$$\frac{\Delta\sigma_8}{\sigma_8} \simeq 1.21 \times 10^{-2}D(z)^{-0.8}\left(\frac{f_{NL}}{100}\right)\left(\frac{\Omega_m h^2}{0.137}\right)^{-1}, \tag{21}$$

$$\Delta n_8 \simeq 7.2 \times 10^{-3}D(z)^{-1.2}\left(\frac{f_{NL}}{100}\right)\left(\frac{\Omega_m h^2}{0.137}\right)^{-1}, \tag{22}$$

where $\Delta\sigma_8$ and Δn_8 are corrections to the effective normalization and tilt of the matter power spectrum at $8h^{-1}$ Mpc, *if* it is deduced from the statistics of collapsed objects at redshift z, using Gaussian predictions. Here, $D(z) \propto ag$ is the linear density growth factor, which is normalized to unity at $z = 0$, and we used the definition

$$\frac{n_{\text{eff}} + 3}{6} = -\frac{d\ln\bar{\sigma}(M)}{d\ln M}. \tag{23}$$

One implication of this result is that, assuming constant $f_{NL} < 100$, one needs to measure σ_8 to better than 1%, in order to be able to distinguish Gaussian and non-Gaussian models from cluster number counts. However, assuming that this is the case, the degeneracy with σ_8 can be broken through the redshift evolution of the mass function (or the effective power spectrum).

We should emphasize that this calculation is *not* linear in δ_m. However, in order to model the nonlinear evolution/collapse, we use the semianalytic approximation of elliptical collapse, which assumes that when the linear δ_m is larger than the collapse threshold, the region collapses into a halo. To the extent that this approximation holds, the halo statistics is simply fixed by the statistics of linear δ_m, which is given by Eq. (14) [and (15)], to 1st order in f_{NL}.

Figure 2 compares the predictions of our analytic framework for the cluster mass function with the simulated mass functions with non-Gaussian initial conditions. In order to find the mass function, we plug the effective $\sigma(M)$ [Eqs. (17) or (21) and (22)] into the Warren *et al.* analytic formula [24] for the halo mass function, which was fitted to ΛCDM N-body simulations with Gaussian initial conditions. This result is compared with the mass functions directly obtained in Dalal *et al.* [7] using non-Gaussian initial conditions. The analytic prediction uses the same cosmology assumed in the non-Gaussian simulations.

It appears that our analytic framework can reproduce the numerical results in the large mass regime (when the corrections are >10%), but overpredicts the mass function by up to ~50% at lower masses, although the discrepancy seems less pronounced at higher redshifts. One possible reason for this discrepancy is that the elliptical collapse
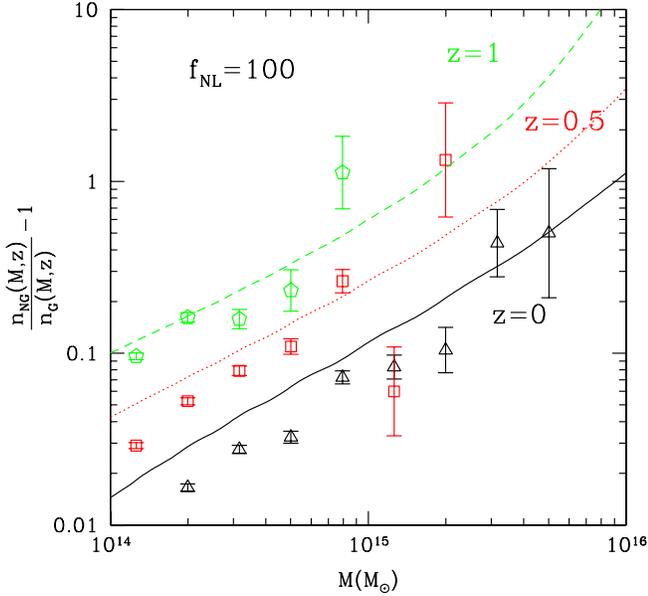
FIG. 2 (color online). The relative change in the cluster mass function for $f_{NL} = 100$. The solid (black), dotted (red), and dashed (green) lines correspond to our analytic prediction based on the $\sigma(M)$ correction [Eq. (21) and (22)] for $z = 0$, 0.5, and 1, respectively. The triangles, squares, and pentagons are simulations of Dalal *et al.* [7] in the same order. The error bars show simple Poisson errors in simulated cluster numbers.

threshold of Eq. (13) [22] is based on an analytic fit to random Gaussian fields. Using non-Gaussian initial conditions will most likely introduce corrections to the elliptical collapse threshold, which could become significant at small masses (although for large masses one still expects the spherical collapse threshold $\delta_{sc} \simeq 1.68$).

Another possibility is the presence of an unknown systematic error (or numerical artifact) in the non-Gaussian simulations. For example, it is clear from the simulation data points in Fig. 2 that the Poisson errors are only an underestimate for the true numerical error. Moreover, it appears that Dalal *et al.* simulations [7] predict a lower non-Gaussian correction to the mass function, in comparison with independent simulations of Grossi *et al.* [25].

Alternatively, one may argue that such a discrepancy at low masses, even if real, may be of little consequence as non-Gaussian corrections are only large (and thus observable) for massive clusters. Nevertheless, more thorough numerical studies (including a suite of non-Gaussian simulations to account for sample variance), as well as an extension of the elliptical collapse framework to non-Gaussian initial conditions should elucidate this issue.

IV. CLUSTERING ON LARGE SCALES

The clustering of collapsed objects on large scales is often described in terms of the bias parameter, which quantifies the ratio of the overdensity of collapsed objects to the total matter overdensity:

$$b(k, M) \equiv \frac{\langle \delta_{m,\mathbf{k}} \delta_{M,-\mathbf{k}} \rangle}{\langle \delta_{m,\mathbf{k}} \delta_{m,-\mathbf{k}} \rangle}, \qquad (24)$$

where

$$\delta_{M,\mathbf{k}} = \langle n_{\mathrm{halo}}(M) \rangle^{-1} \int d^3\mathbf{x} e^{-i\mathbf{k}\cdot\mathbf{x}} \delta n_{\mathrm{halo}}(\mathbf{x}, M), \qquad (25)$$

is the Fourier transform of the overdensity of haloes of mass M.

In the last section, we showed that the statistics of collapsed objects of mass M, resulting from (slightly) non-Gaussian initial conditions, can be described in terms of the statistics of a new Gaussian field $\tilde{\delta}_{mG}$, with a modified elliptical collapse threshold $\tilde{\delta}_{ec}$ [Eq. (15)]. In particular, the fraction of the mass in collapsed objects larger than mass M is given by

$$f(>M) = F(x); \qquad x \equiv \frac{\tilde{\delta}_{ec}(M)}{\tilde{\sigma}(M)} = \frac{\delta_{ec}(M)}{\bar{\sigma}(M)}, \qquad (26)$$

and $F(x)$ is a universal function that can be fitted from simulations, and roughly has the Press-Schecter form:

$$F(x) \simeq F_{\mathrm{PS}}(x) = \sqrt{\frac{2}{\pi}} \int_x^\infty e^{-y^2/2} dy. \qquad (27)$$

Similarly, the fraction of mass in the halo mass interval $(M, M + dM)$ is given by

$$f(M, M + dM) = -F'(x)\left(\frac{\tilde{\delta}_{ec}(M + dM)}{\tilde{\sigma}(M + dM)} - \frac{\tilde{\delta}_{ec}(M)}{\tilde{\sigma}(M)}\right). \qquad (28)$$

Now, adding a positive background of $\tilde{\delta}_0$ ($= \delta_{mG,0} - 2f_{NL}\delta_{ec}\Phi_{pG,0}$) to $\tilde{\delta}_{mG}$ in a large patch of the Universe effectively decreases the collapse threshold to $\tilde{\delta}_{ec} - \tilde{\delta}_0$ for small-scale fluctuations in that region. Therefore, the collapsed mass fraction in that region is slightly modified:

$$\delta f(M, M + dM) = \tilde{\delta}_0 \left\{ \frac{F''(x)}{\tilde{\sigma}(M)} d\left[\frac{\tilde{\delta}_{ec}(M)}{\tilde{\sigma}(M)}\right] \right.$$
$$\left. + F'(x) d\tilde{\sigma}^{-1}(M) \right\}. \qquad (29)$$

Combining Eqs. (28) and (29), and after simple manipulations, we find

$$\frac{\delta f}{f}(M, M + dM) = \frac{\tilde{\delta}_0}{\tilde{\delta}_{ec}(M)} \left\{ -\frac{d \ln F'}{d \ln x} \right.$$
$$\left. + \left[\frac{d \ln \tilde{\delta}_{ec}(M)}{d \ln \tilde{\sigma}(M)} - 1\right]^{-1} \right\}, \qquad (30)$$

which is also known as *Lagranigan* overdensity. Since the mean halo number density is given by $\bar{\rho}_m f(M, M + dM)$, the (Eulerian) halo overdensity with the $\tilde{\delta}_0$ background takes the form

$$\delta_{\text{halo}} = \delta_{mG0} + \frac{\tilde{\delta}_0}{\tilde{\delta}_{\text{ec}}(M)}\left\{-\frac{d\ln F'}{d\ln x}\right.$$
$$\left. + \left[\frac{d\ln\tilde{\delta}_{\text{ec}}(M)}{d\ln\tilde{\sigma}(M)} - 1\right]^{-1}\right\}. \qquad (31)$$

Recalling the definition of $\tilde{\delta}$, this yields the following expression for the (Eulerian) bias:

$$b(k, M) = 1 + \tilde{\delta}_{\text{ec}}(M)^{-1}\left[1 + \frac{2f_{NL}\delta_{\text{ec}}(M)A(k, t)}{g(t)T(k)}\right]$$
$$\times\left\{-\frac{d\ln F'}{d\ln x} + \left[\frac{d\ln\tilde{\delta}_{\text{ec}}(M)}{d\ln\tilde{\sigma}(M)} - 1\right]^{-1}\right\}, \qquad (32)$$

or

$$b(k, M) = b_\infty(M) + \frac{2(b_\infty - 1)f_{NL}\delta_{\text{ec}}(M)A(k, t)}{g(t)T(k)}$$
$$= b_\infty(M) + \frac{3(b_\infty - 1)f_{NL}\delta_{\text{ec}}(M)\Omega_m H_0^2}{k^2 a(t)g(t)T(k)}. \qquad (33)$$

Here, $b_\infty(M) = b(M; f_{NL} = 0) + \mathcal{O}(\frac{f_{NL}\epsilon}{ag})$ is the $k \to \infty$ limit of the *linear* bias. For reasonable values of (constant) $f_{NL} < 100$, similar to $\bar{\sigma}(M)$, this would only lead to percent level corrections to the Gaussian bias, which we can safely ignore and assume $b_\infty(M) \simeq b(M; f_{NL} = 0)$. With this assumption, the ratio of the elliptical to spherical collapse threshold, $\delta_{\text{ec}}(M)/\delta_{\text{sc}}$, (again ignoring f_{NL}) can also be described as a function of b_∞ by combining Eqs. (13), (27), and (32). This is plotted in Fig. 3, and can be

used to estimate the non-Gaussian scale-dependent bias [second term in Eq. (33)] as a function of the small-scale (or Gaussian) linear bias.

Figure 4 shows the scale-dependent part of linear bias [the second term in Eq. (33)] as a function of k for $b_\infty = 1.5, 3$ and $f_{NL} = 100$, which is compared to the Dalal *et al.* derivation [7]. For small k's ($k < k_{\text{eq}}$), both biases scale as k^{-2}, although our prediction is slightly higher by the elliptical to spherical collapse ratio (Fig. 3). However, for $k \gtrsim k_{\text{eq}}$ the non-Gaussian scale-dependent bias factor is further enhanced by the inverse of the transfer function, which was missing in [7] (and was also recently pointed out by [11]).

The unique aspect of this non-Gaussian correction to the galaxy bias is that, rather than going to a constant on large scales, it blows up as $k \to 0$. This is due to the fact that, on large scales, $\tilde{\delta}_{mG}$ is dominated by the Φ_{pG} term [see Eq. (15)], and thus, the galaxy distribution follows the primordial curvature perturbations rather than the matter density. An easy way to understand this phenomenon is illustrated in Fig. 5, which shows how modulation of the amplitude of small-scale density perturbations by large-scale modes (caused by primordial non-Gaussianity) can lead to an additional non-Gaussian bias.

The peculiar scale dependence of the non-Gaussian bias correction differentiates it from any causal astrophysical effect, where bias should go to a constant on large scales (beyond the sound horizon $\sim$ few Mpc's). Therefore, this effect can provide us with a smoking gun for primordial non-Gaussianity, through observations of large-scale clustering in galaxy surveys [7,11]. In fact, Dalal *et al.* [7]
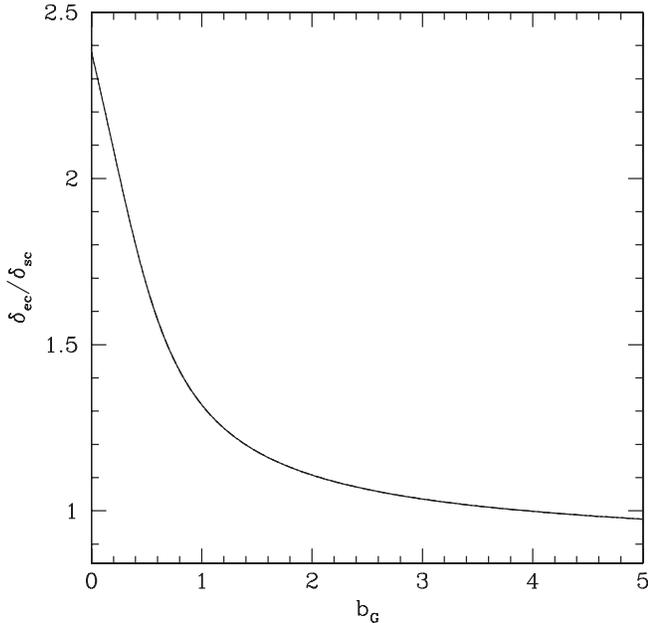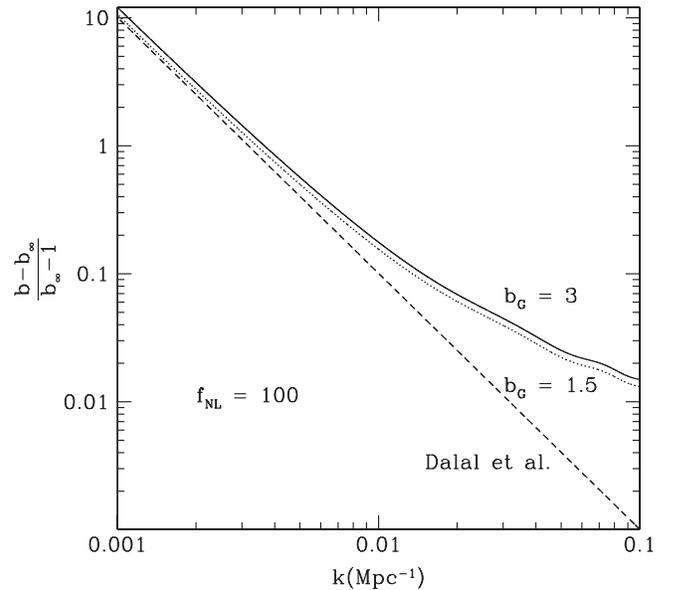

FIG. 3. The ratio of the elliptical to spherical collapse thresholds, as a function of the scale-independent Gaussian linear bias. The scale-dependent non-Gaussian bias is enhanced by this factor [Eq. (33)].

FIG. 4. The scale-dependent part of bias $b - b_\infty$ divided by $b_\infty - 1$, for $f_{NL} = 100$ and $z = 0$. The scale-independent part of bias, $b_\infty = 1.5$ and 3 for solid and dotted curves. The dashed line is Dalal *et al.*'s derivation [7].
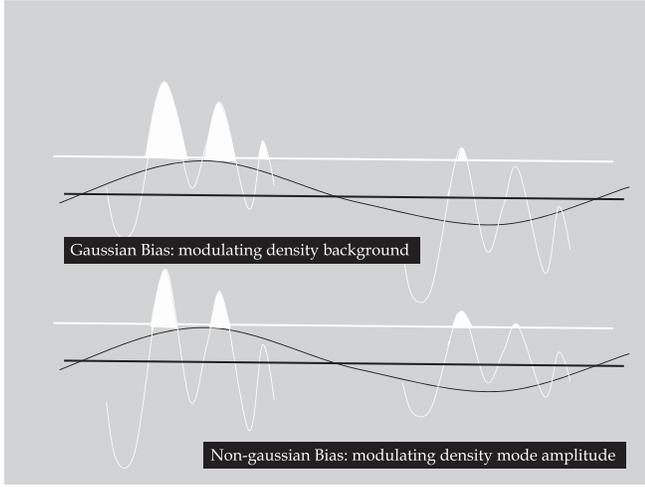
FIG. 5. This figure illustrates the contrast between the Gaussian and non-Gaussian halo/galaxy bias. The two plots show cartoon versions of linear density vs spatial position. The white shaded areas indicate collapse regions. For Gaussian initial conditions, different Fourier modes are uncorrelated, and so long wavelength modes (thin black curve) only change the background local mean value of small-scale modes (thin white curves), which in turn changes the number of density peaks that cross the collapse threshold (thick white line) and form haloes. However, for a non-Gaussian initial condition, the long wavelength modes can also modulate the amplitude of small-scale modes, which causes an additional modulation of collapsed halo density.

claim that next generation of large-scale galaxy surveys can yield $\Delta f_{NL} \lesssim 10$. In Fig. 6, we illustrate this potential for a galaxy redshift survey with a given number of galaxies N_{gal}, fraction of sky coverage f_{sky}, and maximum redshift z_{max}. Here, for simplicity, we assume a uniform comoving galaxy number density, and $b_{\mathrm{gal}} = 2$ for the Gaussian bias. In order to compute the sensitivity of the observation of galaxy power spectrum to an f_{NL}-type primordial non-Gaussianity, we consider the Gaussian χ^2:

$$\chi^2 = \int dV(z) \int_{k_{\mathrm{min}}}^{k_{\mathrm{max}}} \frac{d^3k}{(2\pi)^3} \frac{|\delta_{\mathrm{gal}}(\mathbf{k})|^2}{P_{\mathrm{gal}}(k, z)} + \ln[P_{\mathrm{gal}}(k, z)],$$

(34)

where $P_{\mathrm{gal}}(k, z)$ is the power spectrum of galaxy distribution, which, in the linear regime, is related to the matter power spectrum $P(k, z)$, through bias and the comoving galaxy number density:

$$P_{\mathrm{gal}}(k, z) = b_{\mathrm{gal}}^2(k, z)P(k, z) + n_{\mathrm{gal}}^{-1}.$$

(35)

$b_{\mathrm{gal}}(k)$ is given by Eq. (33), which is parameterized by f_{NL} and b_∞. Assuming that these parameters remain constant across the survey, they can be constrained observationally through minimizing the χ^2 in Eq. (34). Marginalizing over b_∞ leads to the 1σ error predictions (Δf_{NL}'s) shown in Fig. 6.
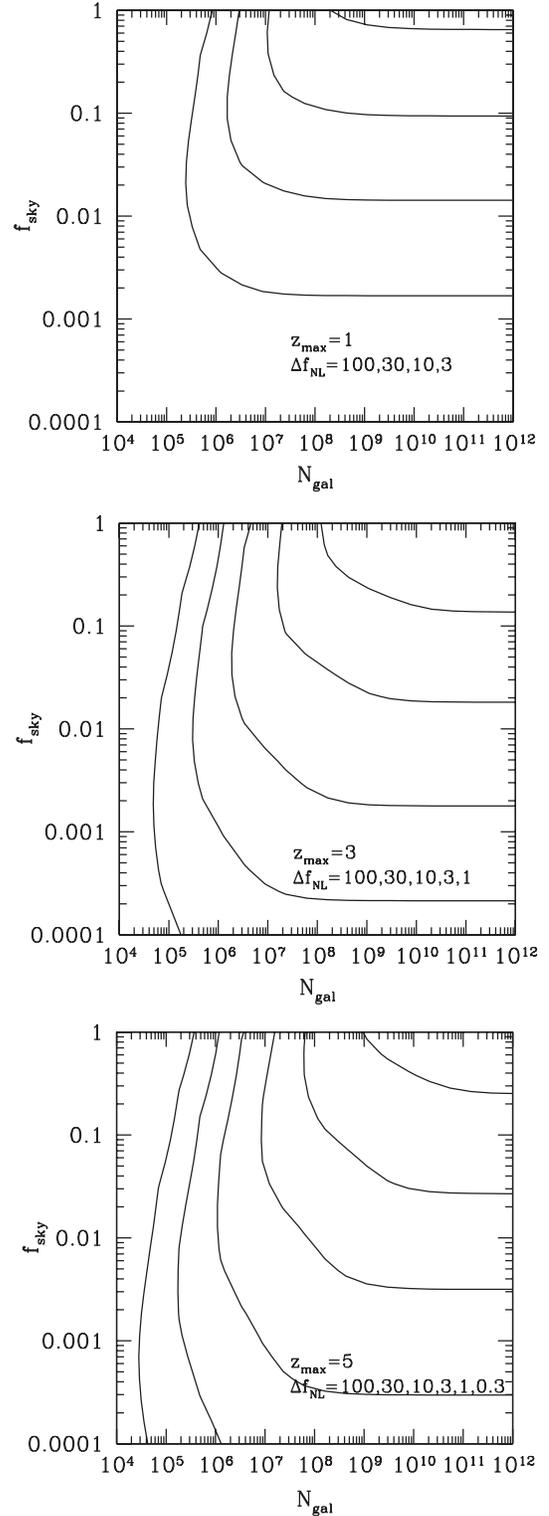

FIG. 6. The expected precision for the measurement of local primordial non-Gaussianity, using the autopower spectrum of a galaxy survey (in lieu of any survey systematics). The contours show the expected precision as a function of the survey sky coverage, f_{sky}, and the number of galaxies in the survey, N_{gal} (assuming uniform comoving galaxy number density and $b_{\mathrm{gal}} = 2$), for $z_{\mathrm{max}} = 1, 3, 5$.

We should note that as the non-Gaussian bias correction grows as k^{-2} on large scales, almost all the statistical power for constraints on f_{NL} comes from the largest wavelength modes in the survey, i.e. roughly $\Delta f_{NL} \propto (k_{\min}/V)^{1/2}$, where V is the volume of the survey. While, for our predictions, we use the rough redshift-dependent estimate:

$$k_{\min}(z) \simeq \frac{2\pi}{[4\pi f_{\mathrm{sky}} r(z)^2]^{1/2}}, \qquad (36)$$

the actual constraints will depend on the large-scale geometry of the survey. We also assume $k_{\max}$ to be the rough scale above which the linear bias framework cannot be used reliably:

$$\frac{k_{\max}^3 P(k_{\max}, z)}{2\pi^2} \simeq 1. \qquad (37)$$

This choice also becomes important for marginalization over b_∞.

Finally, for completeness, we should point out how the results of Fig. 6 can be translated for different values of (small-scale Gaussian) galaxy bias. It is easy to see that if we define

$$\tilde{f}_{NL} \equiv (1 - b_\infty^{-1}) f_{NL}, \qquad \tilde{N}_{\mathrm{gal}} \equiv b_\infty^2 N_{\mathrm{gal}}, \qquad (38)$$

the observational constraints on $\tilde{f}_{NL}$ will only depend on $\tilde{N}_{\mathrm{gal}}$ and the survey geometrical properties, but will be independent of the actual value of b_∞ [26]. As an example, if the mean bias of the galaxy sample is 1.5 rather than 2, the predicted errors will be 50% larger than those in Fig. 6 for the same survey geometry. However, to achieve this, the survey should also have more galaxies by a factor of $(4/3)^2 = 16/9$.

To summarize this section, we have shown that clustering of galaxies on large scales could provide a very sensitive smoking gun for (local) primordial non-Gaussianity. For a (futuristic) deep enough galaxy survey, and in lieu of survey/atmospheric systematics, the constraint could be as good as $\Delta f_{NL} \sim 0.1$ which is an order of magnitude better than the best predicted CMB constraints (and is consistent with recent prediction of [12]). We will discuss more immediate future prospects in Sec. VI.

V. NON-GAUSSIANITY AND THE INTEGRATED SACHS-WOLFE EFFECT

In the last section, we showed that observations of clustering of galaxies on large scales could potentially lead to very sensitive constraints on a local primordial non-Gaussianity. However, in order to be competitive with upcoming CMB constraints, the selection strategy of the survey needs to be uniform to 1 part in 10^5 on the largest angles (which are most sensitive to the non-Gaussian bias). Otherwise, the correlation function will

be dominated by the survey systematics, which can lead to a false detection of positive non-Gaussianity.

While such survey calibration seems daunting by today's standards, an alternative method that is not prone to such systematic errors is to use cross correlation of the galaxy survey with an already well-calibrated large-scale structure survey, i.e. the CMB sky. Although most of the CMB anisotropies on large angles were originated at the last scattering surface, a small fraction of them were generated after the onset of dark energy domination through the integrated Sachs-Wolfe (ISW) effect [27]. It was proposed that the ISW effect could be detected through cross correlating CMB with the galaxy surveys [28–30], and its subsequent detection in cross correlation of WMAP maps with various galaxy surveys has been interpreted as independent evidence for the presence of dark energy [31–41]. However, this analysis should also be sensitive to galaxy bias on large scales, and therefore (as also pointed out in [7]) could be a sensitive probe of primordial non-Gaussianity through the non-Gaussian bias correction of the previous section. In fact, *both the non-Gaussian contribution to the galaxy distribution and the ISW effect follow the Newtonian potential on large scales*, and therefore should be well correlated. In this section, we predict the potential of this method for constraining primordial non-Gaussianity.

In the linear regime, the perturbations in the number density of galaxies of a redshift survey projected onto the sky can be written as a line of sight integral over the total matter overdensity:

$$\delta_{\mathrm{gal}}(\hat{\mathbf{n}}) = N^{-1} \int dr r^2 n_{\mathrm{gal}}(r) b_{\mathrm{gal}}(r) \delta_m(r\hat{\mathbf{n}}; t), \qquad (39)$$

where $n_{\mathrm{gal}}(r)$ is the comoving mean number density of galaxies, $b_{\mathrm{gal}}(r)$ is the average bias for the galaxy distribution, and N is the mean number of galaxies per steradian. The secondary anisotropy induced in the CMB due to the ISW effect is given by

$$\delta_{\mathrm{ISW}}(\hat{\mathbf{n}}) = -2 \int dr \frac{d \ln g}{dr} \Phi(r\hat{\mathbf{n}}; t), \qquad (40)$$

where $g(t)$ is again the growth factor for the Newtonian potential. The angular cross-power spectrum of the projected galaxy distribution and the ISW effect is well described by the Limber approximation [35]:

$$C(\ell)_{\mathrm{gal, ISW}} = -2N^{-1} \int dr \frac{d \ln g}{dr} n_{\mathrm{gal}}(r) b_{\mathrm{gal}}(r)$$
$$\times P_{\Phi, m}\left(\frac{\ell + 1/2}{r}; t\right), \qquad (41)$$

where $P_{\Phi, m}(k; t)$ is the 3D cross-power spectrum of the Newtonian potential with the comoving matter overdensity.

We should point out that the non-Gaussian (quadratic) term in the (linear) ISW effect [$f_{NL}\Phi_{pG}^2$ in Φ of Eq. (40)]

can only lead to terms proportional to the three-point function of Gaussian fields in $C(\ell)_{\rm gal,ISW}$ [Eq. (41)], which vanish. The first nonvanishing term would be the product of two non-Gaussian (f_{NL}) terms, which is quadratic in f_{NL} and thus neglected here. There could still be non-vanishing contributions due to secondary non-Gaussianities (caused by nonlinear evolution) in $\delta_{\rm halo}$, but this is expected to be negligible on large scales where ISW cross correlation is observed, since density contrast is small.

As we argued in the previous section, a local primordial non-Gaussianity results in a scale-dependent bias factor, $b_{\rm gal}(k, r)$, where $k \simeq \frac{\ell+1/2}{r}$ in the Limber approximation. Since the non-Gaussian bias correction [Eq. (33)] grows as k^{-2} on large scales, it can dominate the Gaussian bias for $f_{NL} \gtrsim 1$ on scales smaller than the horizon. Therefore, the cross correlation of the CMB and galaxy distribution can be a sensitive probe of f_{NL} on large angles. As an example, Fig. 7 compares the expected cross-power spectra between the NVSS radio galaxy survey and the CMB sky. The data points are the recent measurements from Ho *et al.* [40]. The dashed line shows the expectation from the WMAP5 concordance cosmological model and the redshift distribution provided in [40], assuming a constant bias. The solid line uses the best (unmarginalized) fit *non-Gaussian* bias to the data, which corresponds to $f_{NL} = 236 \pm 127$.
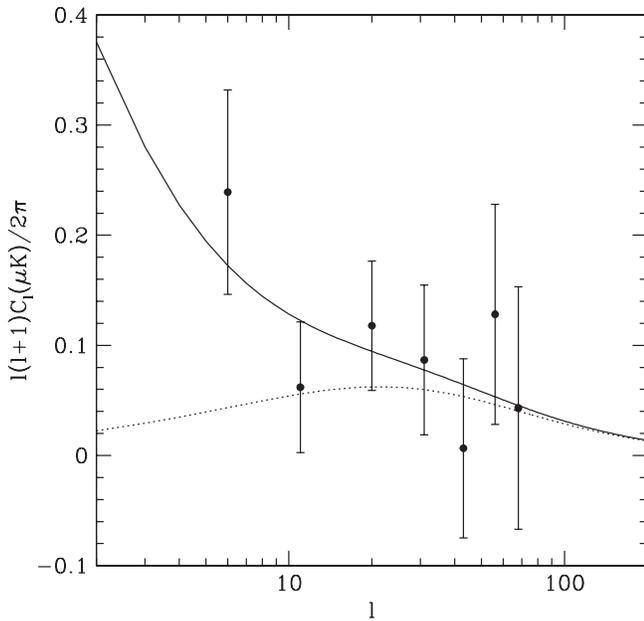

FIG. 7. Cross-power spectrum between NVSS radio galaxy survey and the CMB sky. The data points are from Ho *et al.* [40]. The dashed line shows the expectation from WMAP5 concordance cosmological model assuming a constant bias, while the solid line uses the best (unmarginalized) fit *non-Gaussian* bias to the data, which corresponds to $f_{NL} = 236 \pm 127$.

While this result can only be cautiously interpreted as a 2σ hint for a local primordial non-Gaussianity, we notice that it has the same sign (and comparable magnitude) as the current $2-3\sigma$ evidence for primordial non-Gaussianity from WMAP CMB maps (see Sec. I), even though they correspond to different scales, eras, and measurement methods.

Finally, we can predict how well one can measure the local non-Gaussianity from cross correlation of CMB with future galaxy surveys. To do this, we follow the method outlined in [42], where we divide up the galaxy survey into separate spherical shells and add up the signal-to-noise for the detection of a given f_{NL} in each shell in quadratures. The error for this measurement is proportional to the clustering of galaxies, and so the correlation between measurements in different shells can be safely ignored on large scales (at least for small f_{NL}). Within this approximation, the optimum signal-to-noise for a (full-sky) detection of a given f_{NL} is given by

$$(S/N)^2 \simeq \sum_\ell \frac{4(2\ell+1)}{C_{TT}(\ell)} \int \frac{dr(z)}{r(z)^2} \left(\frac{d\ln g}{dr}\right)^2 \left(\frac{\Delta b_{\rm gal}(k, z)}{b_{\rm gal}(k, z)}\right)^2$$
$$\times \frac{P^2_{\Phi,m}(k, z)}{P(k, z)}, \tag{42}$$

where $C_{TT}(\ell)$ is the angular autopower spectrum of the CMB, and $P(k, z)$ is the linear matter power spectrum. Moreover, $b_{\rm gal}$ and $\Delta b_{\rm gal}$ ($\propto f_{NL}$) are the total bias and its non-Gaussian correction [obtained in Eq. (33)], respectively. As in the Limber approximation, $k \simeq \frac{\ell+1/2}{r}$ in Eq. (43).

Figure 8 shows the predicted signal-to-noise [from Eq. (43)] for the detection of primordial non-Gaussianity through cross correlation with the CMB for a full-sky redshift survey that extends out to $z_{\rm max} = 0.5, 1, 2,$ and 5, where we assumed the small-scale Gaussian bias $b_{\rm gal} \simeq 2$ for concreteness. Therefore, we predict that cross correlation of CMB and large-scale structure surveys can eventually be competitive with the CMB surveys in providing independent constraints on local primordial non-Gaussianity at the level of $\Delta f_{NL} \sim 3$.

In this calculation, we have ignored the Poisson noise (due to a finite number of galaxies) or other possible survey systematics. However, we should note that since these errors do not contribute to the ISW signal [43], they cannot cause a systematic bias in f_{NL} determinations from ISW-galaxy cross correlation analyses. Nevertheless, Poisson noise/survey systematics can increase the random error in cross correlation, which is manifested through increasing the apparent autocorrelation of the galaxy survey. Since we have already calculated the sensitivity of the galaxy auto-correlation function to f_{NL} in the last section (Fig. 6), we can say that assuming $f_{NL} \gtrsim 3$, for $N_{\rm gal} \gtrsim 10^{7-8}$ the galaxy autocorrelation will become roughly insensitive to the Poisson noise on large scales (where the non-Gaussian
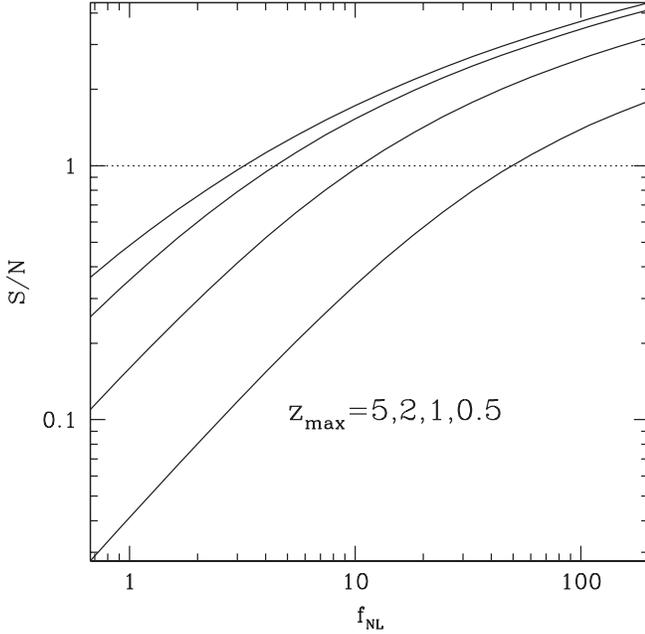
FIG. 8.   Optimum full-sky signal-to-noise for the detection of a possible f_{NL} through cross correlation of a full-sky redshift survey (assuming $b_{gal} \simeq 2$) with the CMB sky. The four curves correspond to the maximum redshift of the survey, $z_{max} = 0.5, 1, 2$, and 5.

signal dominates). Therefore, the determination of f_{NL} from ISW-galaxy cross correlation will not significantly suffer from Poisson noise, as long as the number of galaxies in a (full-sky) survey is larger than 10^{7-8}. Assuming that survey systematics also look similar to Poisson noise on large angles, this also implies that the calibration of the galaxy survey selection function must be better than $1/\sqrt{10^8} = 10^{-4}$, in order not to affect the sensitivity of non-Gaussianity measurements.

VI. CONCLUSIONS AND FUTURE PROSPECTS

In this paper, we have studied the impact of a local primordial non-Gaussianity (expected from superhorizon evolution in multifield inflationary models or the new ekpyrotic scenario) on the statistics of collapsed objects in the late Universe. We developed an approximation that describes the linear non-Gaussian density field as a local bilinear combination of Gaussian fields. Applying this to the elliptical collapse framework enabled us to find an analogue Gaussian linear density field, which should produce the same predictions as the non-Gaussian system for collapsed objects of a given mass, assuming small primordial non-Gaussianity. For halo mass function, this prescription leads to an effective analytic correction to $\sigma(M)$, the amplitude of linear fluctuations on mass scale M, which can be readily applied to Press-Schechter-like analytic approximations for the mass function. We found this to give roughly consistent results with direct simulations with

non-Gaussian initial conditions. We then focused on the impact of non-Gaussianity on the clustering of galaxies, and confirmed the recent discovery of a nonlocal bias correction associated with local non-Gaussian initial conditions [7,11]. We found that autocorrelation of galaxy surveys can ultimately lead to a precision $\Delta f_{NL} \sim 0.1$ which is an order of magnitude better than the best anticipated CMB constraints. Cross correlation of CMB and galaxy surveys (through the ISW effect), even though suffering from a low signal-to-noise due to primary CMB contamination, is less prone to survey systematics, and can also yield $\Delta f_{NL} \sim 3$, which is again competitive with future CMB constraints. Interestingly, the cross correlation of CMB and the NVSS radio galaxy survey already provides a 2σ hint for a large primordial non-Gaussianity $f_{NL} = 236 \pm 127$.

Finally, we should comment on the future developments of the studies of primordial non-Gaussianity While five more years of WMAP observations can only marginally improve the current constraints to $\Delta f_{NL} \sim 15$–20, the most anticipated observational improvement is the first data release of Planck satellite, expected in 2012. After careful accounting for CMB secondary anisotropies, as well as the 2nd order corrections to the CMB Boltzmann solvers, Planck is expected to achieve $\Delta f_{NL} \sim 3$. However, as we demonstrated in this paper, competitive or even superior constraints can come from the study of clustering of large-scale galaxy surveys. For example, Large Synoptic Survey Telescope (LSST) is expected to survey half of the sky down to magnitude $r \sim 27.5$ finding $\sim 10^{10}$ galaxies out to $z_{max} \sim 2$. Using Fig. 6, LSST can yield a precision of $\Delta f_{NL} \sim 2$–3 through its galaxy autocorrelation function. The cross correlation of LSST with the CMB sky could independently yield $\Delta f_{NL} \sim 7$. Similar proposed or underway deep, large area surveys such as Pan-STARRS, ADEPT, CIP, etc. could potentially lead to comparable constraints. However, one important point to note is that *precise spectroscopic redshifts are not required for non-Gaussianity determination, as most of the non-Gaussian signal comes from the largest scales of the survey.*

Shortly before the submission of this paper, Slosar *et al.* [13] published a comprehensive study of this effect in the present large-scale cosmological galaxy surveys, leading to a significant constraint: $f_{NL} = 28 \pm 25$. Nearly all the statistical weight of this constraint is based on the lack of anomalous power in the lowest l-bin of the autopower spectrum of a high redshift sample of $\sim 10^5$ Sloan Digital Sky Survey photometric quasi-stellar objects, which cover 15% of sky and have an effective mean bias of ~ 2.75. Plugging these numbers into the error estimates of Sec. IV (Fig. 6), we would predict an error of about a factor of 2 worse than what was found in [13]. As we pointed out in Sec. IV, the actual error on f_{NL} will be sensitive to the survey geometry, and so this disagreement between our rough error prediction and Slosar *et al.*'s findings may not

be an immediate cause for alarm. Nevertheless, it is curious that despite the lack of any redshift information (which was used in our error prediction), as well as possible contaminations/survey systematics, Slosar *et al.* could achieve an accuracy in f_{NL} comparable or better than an optimistic error prediction. A further concern is that the correlated Poisson noise (typically expected from nonlinear clustering, e.g. [45]) is not marginalized over in their analysis.

On a different note, Slosar *et al.* point out that a degeneracy between f_{NL} and the redshift and bias distribution of NVSS (which are both poorly known), hinders any accurate determination of f_{NL} from NVSS data, or its correlation with the CMB. While, at face value, this might seem inconsistent with our $\sim 2\sigma$ evidence for non-Gaussianity from the correlation of NVSS with WMAP, we should note that both redshift and bias distributions may eventually be fixed, either from theoretical modeling, or direct observations, which can then realize the full statistical power of cross correlation constraints.

On the theoretical front, we are still far from a good theoretical understanding of nonlinear structures from non-Gaussian initial conditions. While the present work provides a bridge between this problem, and the huge body of work on nonlinear/collapsed structures from Gaussian initial conditions (such as the semianalytic frameworks, the halo model, as well the N-body/hydro numerical simulations), it does suffer from some simplifying assumptions. The first is the ellipsoidal collapse scenario [22,23]. Although for large masses and $f_{NL}\epsilon(M) \ll 1$ (i.e. near-Gaussian initial conditions), the collapse should be well described as spherical (leading to a universal spherical collapse threshold), for smaller masses, the threshold depends on the ellipticity of the proto-halo, which is a random number with a distribution set by the statistics of the initial conditions. The current analytic form [Eq. (13)] is a fit to random realizations of Gaussian fields with ΛCDM power spectrum, and thus there is no reason to exclude corrections linear in f_{NL}, if the linear density field obeys non-Gaussian statistics. Therefore, one clear direction for improvement is to investigate/extend the ellipsoidal collapse framework for non-Gaussian initial conditions.

The second simplification (and probably the most nontrivial part of this work) is the local approximation of Eq. (11), which was instrumental in all the subsequent analysis/predictions. While the consistency of our results for mass function and clustering with other numerical and analytic studies [7,11] verifies the accuracy of this approximation, Eq. (11) could also be viewed as a *convenient parametrization for the primordial non-Gaussianity.* To see this, note that for an f_{NL}-type non-Gaussianity [Eq. (1)], the bispectrum of Φ, $B(k_1, k_2, k_3)$, is written in terms of the power spectrum of the Gaussian field Φ_{pG}, $P_{pG}(k)$:

$$B(k_1, k_2, k_3) = -2 f_{NL} P_{pG}(k_1) P_{pG}(k_2) P_{pG}(k_3)$$
$$\times [P_{pG}^{-1}(k_1) + P_{pG}^{-1}(k_2) + P_{pG}^{-1}(k_3)], \quad (43)$$

where $P_{pG}(k) \propto k^{-3}$ for scale-invariant spectra. While our local approximation does not exactly preserve this form, different observables are mostly sensitive to a limited region in the k_1, k_2, k_3 space, where Eq. (11) may provide a simple parametrization. For example, it is reasonable to think that measurements of mass function of haloes of mass M would only be sensitive to non-Gaussianity around $k_1 \sim k_2 \sim k_3 \sim \pi/R(M)$ [see the line preceding Eq. (12)]. Therefore, constraints on f_{NL} using mass function could be roughly translated into constraints on the bispectrum for *near-equilateral triangles* of the side $\sim \pi/R(M)$.

A more interesting case is the clustering (or the power spectrum) of collapsed haloes (e.g. galaxies) that we studied in Secs. IV and V. In this case, we are dealing with a hierarchy of scales: $k_3 \ll k_1$, k_2, where $k_1 \ll 0.01$ Mpc^{-1} is the scale over which we measure the galaxy power spectrum, and as we showed predominantly corresponds to the modes of the size of the survey. On the other hand, k_1, $k_2 \sim \pi/R(M) \gtrsim 0.1$ Mpc^{-1} correspond to the comoving size of the collapsing regions. Therefore, observations of the galaxy power spectrum provides an exquisite measure of the bispectrum for *squeezed triangles.* In particular, the non-Gaussian bias correction $\Delta b_{NG}(k_3) \propto k_3^{-2}$ found in Sec. IV (as well as in [7,11]) are specific to the local form of non-Gaussianity ([Eq. (43)] which is only produced through superhorizon evolution, where $B(k_1, k_2, k_3) \propto k_3^{-3}$ as $k_3 \to 0$. In contrast, for non-Gaussianities produced through subhorizon evolution (such as in slow-roll single-field inflation, DBI, ghost inflation and k-inflation), $B(k_1, k_2, k_3) \to$ const or 0 in this limit [46]. Therefore, by way of comparison, the scale-dependent bias $\Delta b_{NG}(k_3) \to 0$ (and so is unobservable) in the long wavelength limit, for models that produce equilateral primordial non-Gaussianity.

We thus close this paper by the exciting (although, by now, familiar) premise of the cosmological microscope as a probe of the fundamental physics of the early Universe. In particular, that an anomalously large clustering of galaxies on large scales will be a unique smoking gun for nonlinear superhorizon evolution (and thus large primordial isocurvature modes) during the early stages of cosmogenesis.

ACKNOWLEDGMENTS

[1] J. M. Maldacena, J. High Energy Phys. 05 (2003) 013.
[2] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, Phys. Rev. D **76**, 123503 (2007).
[3] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, Phys. Rev. Lett. **100**, 171302 (2008).
[4] A. P. S. Yadav and B. D. Wandelt, Phys. Rev. Lett. **100**, 181301 (2008).
[5] E. Komatsu *et al.*, arXiv:0803.0547.
[6] P. Serra and A. Cooray, Phys. Rev. D **77**, 107305 (2008).
[7] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, Phys. Rev. D **77**, 123514 (2008).
[8] J. A. Willick, arXiv:astro-ph/9904367.
[9] M. LoVerde, A. Miller, S. Shandera, and L. Verde, J. Cosmol. Astropart. Phys. 04 (2008) 014.
[10] E. Sefusatti and E. Komatsu, Phys. Rev. D **76**, 083004 (2007).
[11] S. Matarrese and L. Verde, Astrophys. J. **677**, L77 (2008).
[12] P. McDonald, arXiv:0806.1061.
[13] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, J. Cosmol. Astropart. Phys. 08 (2008) 031.
[14] D. J. Eisenstein and W. Hu, Astrophys. J. **496**, 605 (1998).
[15] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).
[16] E. Komatsu and D. N. Spergel, Phys. Rev. D **63**, 063002 (2001).
[17] L. Kofman, Phys. Scr. **T36**, 108 (1991).
[18] L. Moscardini, S. Matarrese, F. Lucchin, and A. Messina, Mon. Not. R. Astron. Soc. Lett. **248**, 424 (1991).
[19] R. Scherrer, Astrophys. J. **390**, 330 (1992).
[20] A. Gangui, F. Lucchin, S. Matarrese, and S. Mollerach, Astrophys. J. **430**, 447 (1994).
[21] $g(t)$ is the growth factor of the metric or Newtonian potential, which is constant and normalized to unity during the matter era, and only starts to decay at the onset of dark energy domination. This should not be confused with $D(z)$, the overdensity growth factor, which grows as a in the matter era, and is often normalized to unity at $z = 0$.
[22] R. K. Sheth, H. J. Mo, and G. Tormen, Mon. Not. R. Astron. Soc. **323**, 1 (2001).
[23] R. K. Sheth and G. Tormen, Mon. Not. R. Astron. Soc. **329**, 61 (2002).
[24] M. S. Warren, K. Abazajian, D. E. Holz, and L. Teodoro,

Astrophys. J. **646**, 881 (2006).
[25] M. Grossi, K. Dolag, E. Branchini, S. Matarrese, and L. Moscardini, Mon. Not. R. Astron. Soc. **382**, 1261 (2007).
[26] Here, we have ignored the bias dependence of the elliptical collapse threshold Fig. 3, as it could only change the error predictions by 10%–20%.
[27] R. K. Sachs and A. M. Wolfe, Astrophys. J. **147**, 73 (1967).
[28] R. G. Crittenden and N. Turok, Phys. Rev. Lett. **76**, 575 (1996).
[29] H. V. Peiris and D. N. Spergel, Astrophys. J. **540**, 605 (2000).
[30] A. Cooray, Phys. Rev. D **65**, 103510 (2002).
[31] S. Boughn and R. Crittenden, Nature (London) **427**, 45 (2004).
[32] P. Fosalba and E. Gaztanaga, Mon. Not. R. Astron. Soc. **350**, L37 (2004).
[33] R. Scranton *et al.*, arXiv:astro-ph/0307335.
[34] M. R. Nolta *et al.*, Astrophys. J. **608**, 10 (2004).
[35] N. Afshordi, Y.-S. Loh, and M. A. Strauss, Phys. Rev. D **69**, 083524 (2004).
[36] N. Padmanabhan *et al.*, Phys. Rev. D **72**, 043525 (2005).
[37] A. Cabre, E. Gaztanaga, M. Manera, P. Fosalba, and F. Castander, Mon. Not. R. Astron. Soc. Lett. **372**, L23 (2006).
[38] T. Giannantonio *et al.*, Phys. Rev. D **74**, 063520 (2006).
[39] A. Rassat, K. Land, O. Lahav, and F. B. Abdalla, Mon. Not. R. Astron. Soc. **377**, 1085 (2007).
[40] S. Ho, C. M. Hirata, N. Padmanabhan, U. Seljak, and N. Bahcall, Phys. Rev. D **78**, 043519 (2008).
[41] T. Giannantonio *et al.*, Phys. Rev. D **77**, 123520 (2008).
[42] N. Afshordi, Phys. Rev. D **70**, 083536 (2004).
[43] A notable exception is galactic foregrounds in the CMB which could be correlated with galactic contaminations of galaxy surveys. However, similar to the analysis of primary CMB anisotropies [44], the frequency dependence of these foregrounds can provide a way to clean the cross correlation signal from galactic contaminations.
[44] B. Gold *et al.*, arXiv:0803.0715.
[45] P. McDonald, Phys. Rev. D **74**, 103512 (2006).
[46] D. Babich, P. Creminelli, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 08 (2004) 009.